ΦSX

Theoretical Physics IV
Statistical Physics

Peter E. Blöchl

Caution! This version is not completed and may contain errors

Institute of Theoretical Physics; Clausthal University of Technology;
D-38678 Clausthal Zellerfeld; Germany;
`http://www.pt.tu-clausthal.de/atp/`

[1]

---

[1] To the title page: What is the meaning of ΦSX? Firstly, it reads like "Physics". Secondly the symbols stand for the three main pillars of theoretical physics: "X" is the symbol for the coordinate of a particle and represents Classical Mechanics. "Φ" is the symbol for the wave function and represents Quantum Mechanics and "S" is the symbol for the Entropy and represents Statistical Physics.

Foreword and Outlook

The goal of statistical physics is to learn about the behavior of systems with *very* many particles, such as atoms, electrons. etc. For such systems the behavior of an individual particle is on the one hand too complicated to follow, on the other hand the interest is more on the effective behavior of the entire system. For example we need not measure the trajectories of all atoms in a hot plate to find out that it is hot. Statistical mechanics tries to describe the average behavior of the parts of a large system instead of following each particle individually. Furthermore it derives this average behavior of a large system from the underlying microscopic equations. Thus, statistical mechanics links the microscopic and the macroscopic world, and forms the foundation of thermodynamics.

The topic of statistical physics are systems with a large number degrees of freedom and systems of which our knowledge is incomplete. Since our knowledge of almost any real system is incomplete, statistical mechanics is a vital ingredient in almost any discipline of physics, and beyond.

Typical questions are:

- How does a combustion engine work?
- What will the weather be tomorrow?
- What is the maximum efficiency of a power plant?
- How does a refrigerator work?
- By how much does a gas expand when it is heated?
- How are molecules arranged in a liquid?
- Why does the magnetization disappear at high temperature?
- What is the reason for the microstructure in a metal?
- What happens when crystals grow?
- What is turbulence?
- What is the dopant concentration profile in a transistor?
- Why does iron corrode and why does Aluminum not corrode?
- What is the rate of a chemical reaction?
- How does a catalytic reaction occur at a surface?
- How does a Tiger get his stripes?

Not all of these topics will be discussed in this book. However, the idea is to provide the underlying concepts that will allow the student to tackle some of these problems and to enter the advanced subjects efficiently as well as to follow the actual developments in the field.

The book corresponds to a course taught in the Bachelor program in a 56-academic hour course spread over 14 weeks, augmented by 28 hour exercise classes with homework.

Here some comments on books

- Kerson Huang,*Introduction to Statistical Physics*, (Taylor and Francis, London 2001) Based on a one-semester undergraduate course at M.I.T. A rather recent book from 2001.
- Dilip Kondepudi and Ilya Prigogine,*Modern Thermodynamics: From Heat Engines to Dissipative Structures* (Wiley, 1998) Ilya Prigogine received the Nobel price in chemistry for his contributions to non-equilibrium physics and especially thermodynamics far from equilibrium.
- Franz Schwabl, *Statistische Mechanik* (Springer, 2000)
- Jürgen Schnakenberg *Thermodynamik und Statistische Physik* (Wiley-VCH, 2002)
- H.B. Callen, *Thermodynamics*. This book deals with thermodynamics on a macroscopic scale. It concentrates first on the formal structure and then shows a wide range of applications. Finally irreversible thermodynamics is discussed. I enjoyed this book a lot.
- P. Atkins *Physical Chemistry*. This book covers the chemical point of view very well. It starts very simple but later also goes into the statistical foundations. Many practical applications to chemistry.
- B.K. Agarwal and M. Eisner, *Statistical mechanics*, Second edition (New age international Publishers, New Dehli, 1988, 1998) ISBN 81-224-1157-6. I bought this book once in Bangalore, India and I am not sure if one can obtain this book also in the first world. While one needs to get accustomed to the notation, it is very thorough and in particular discusses a number of very interesting examples. The selected problems are also widely useful in practice of the non-academic world.
- Roger Balian *From Microphysics to macrophysics* Vol 1 and 2, (Springer Verlag, 1992)

In this course I will try to explicitly derive all the fundamental concepts and all equations used. This should provide the student with a thorough understanding of the meaning of the equations and provide insight into the way theories are built. The capability of extending or building theories is of fundamental importance of a physicist, whose real value plays out when new situations are encountered and need to be analyzed.

An attempt is made to link all equations. I noticed that the beginner does not yet memorize all the equation, respectively does not immediately recognize which of the earlier equations enter a derivation. Thus a lot of time is spent searching back and forth. Since this searching is not always efficient to also repeat the material, as it should, I have decided to make these connections explicit. The danger for the student is that he may not see the need to memorize the main equations. A wise student will use the time, he saves, in a concentrated study of the material.

Contents

3 Equilibrium 67

4 Thermodynamic Processes 105

Chapter 1

Entropy and Information

1.1 History of probability theory

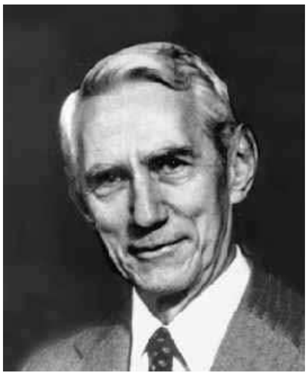


Fig. 1.1: Claude Shannon, 1916-2001

The roots of the theory of probability lie in the theory of games of chance.[1] In the French Society of the 1650's gambling was a popular and fashionable habit. A passionate gambler, chevalier De Méré, consulted the mathematician Blaise Pascal[1] in Paris. In the correspondence between him and his colleagues such as Pierre Fermat, the origin of probability theory has been laid. At this time the games were of the kind like rolling dice[2] with a finite number of equally probable outcomes. The questions were related to finding the chances of winning for complex combinations of such simple events.

At about 1700 the Swiss James Bernoulli introduced the first result of general importance in probability theory, namely the Bernoulli Theorem, which was published in his work "Ars Conjectandi". In the early 18'th century, De Moivre, a French Huguenot, who had to leave France because of his religion, developed in England the so-called multiplication rule for conditional probabilities in his work "The Doctrine of Chances", and provided first indications for the normal distribution. At that time, the theory was applied to problems of estimating risks of insurances. The first systematic account of the mathematical theory of games of chance has been given in 1812 by Laplace in his treatise "Théorie analytique des probabilités". Later, Gauss and Laplace discussed the error estimates in physical and astronomical measurements. Quetelet applied the theory to social science. Maxwell, Boltzmann and Gibbs laid the foundation of statistical mechanics.

1.2 Probability

PROBABILITY

Probability is a measure of the likelihood of an event.

The probability is a cognitive concept.[3] *A person uses the probabilities in his/her mind to form*

---

[1] Blaise Pascal. French Scientist 1623-1662.

[2] Singular: die; plural: dice; German: Würfel.

[3] There are different schools of interpreting probabilities. We follow here the subjective interpretation. In the objective interpretation the probability is not a property of the observer, but of the system and is directly measurable. The distinction is a philosophical one and the mathematical formalisms of the two schools are identical.

expectations about possible events. The person has to know the possible events and its properties. The person will form probabilities about the events based on previous experience. The person is then asked to predict the value of a property of an event that will occur in future. The person will infer the most likely value of that property from his/her probabilities.

In the following, we will formulate a mathematical concept that reflects this understanding of probability. We start out with an experiment, which can lead to one out of a number of possible events. The possible events are numbered and characterized by an index i. If we throw a die, there are six possible events

1. Event 1: "face with one point points up"
2. Event 2: "face with two points points up"
3. etc..

We will call each event a **state** of the respective system. The die itself is the **system**. It is important to include *all* possible events. For example, if it is possible that the die rolls off the table and gets lost somehow, we would need to include this as one of the possible events.

We define now the **probability** P_i. P_i^X is the probability that the system X is in its i-th state. The definition of this concept rests on four axioms:

PROBABILITIES

1. The probability for each event is scalar. It must be scalar because the purpose of the probability is to compare the expected likelihood of events.
2. The probability can take values between 0 and 1, that is

$$0 \leq P_i \leq 1\,. \tag{1.1}$$

The expectation "Event i is not likely at all" is attributed a probability $P_i = 0$. The expectation "Event i will occur with certainty" is attributed a probability $P_i = 1$

3. The probability that one out of a subset of events takes place is the sum of the probabilities of the events in the subset.[a]

$$P[i \in \{i_1, \ldots, i_M\}] = \sum_{j=1}^{M} P_{i_j}\,. \tag{1.2}$$

A consequence of this and the previous axiom, Eqs. 1.1,1.2 is the **normalization condition**

$$\sum_{i=1}^{N} P_i = 1\,. \tag{1.3}$$

4. The probability for a joint event (i, j) in two statistically uncorrelated systems[b] X and Y is equal to the product of the probabilities that event i occurs in system X and that event j takes place in system Y. That is

$$P_{i,j}^{XY} = P_i^X P_j^Y \qquad \text{for statistically uncorrelated systems.}$$

(See figure 1.2 for a demonstration.)

---

[a] The same axiom can also be written as follows. For two independent subsets of events, named A and B, with a vanishing intersection, i.e. $A \cap B = \varnothing$, the probability $P(A \cup B)$ for the union of A and B is the sum of the probabilities for A and B separately, i.e. $A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B)$. $\varnothing$ is the empty set.

[b] This requirement can also been seen as a definition of statistical uncorrelated systems.

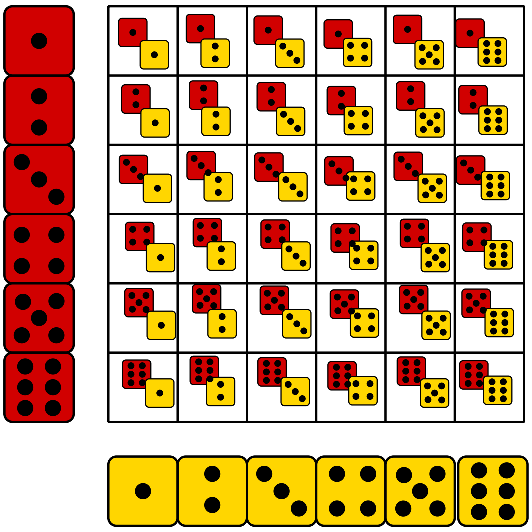


Fig. 1.2: Demonstration that probabilities for uncorrelated events are multiplicative. An example: The probability that the red die shows a 3 is $P_X(3)=\frac{1}{6}$. Similarly, the probability that the yellow die shows a 5 is $P_Y(5)=\frac{1}{6}$. The probability for the joint event, namely that the red die shows a 3 **and** that the yellow die shows a 5, is $P_{XY}(3,5)=P_X(3)P_Y(5)=\frac{1}{36}$, because there are 36 events with equal probabilities.

How can we obtain probabilities?

- Certainty: If we are certain that the outcome is that the system is in state i_0, we will assign probabilities

  $$P_i=\delta_{i,i_0}=\begin{cases}1 & \text{for} \quad i=i_0\\ 0 & \text{for} \quad i\neq i_0\end{cases}$$

  that is we attribute probability one for the state i_0 and probability zero for all others.

- Uncertainty: If we do not know anything about the system other than its N possible states, the first guess would be to attribute equal probabilities to each state,

  $$P_i=\frac{1}{N}\,.$$

  There is no reason to assume that one event will occur more likely than the other. Hence we assign the same probability to each state. The normalization condition determines its value.

- Frequencies: If we have reason to believe that a die is asymmetric, we would try to reproduce the experiment as closely as possible in a "laboratory experiment" and measure the relative frequencies of the events. Let us say that we performed M times an experiment with N possible outcomes. We list the number of times M_i a certain event i occurred. (Note that $\sum_{i=1}^{N} M_i=M$). The **relative frequencies** [4] are defined as

  $$f_i=\frac{M_i}{\sum_{i=1}^{N} M_i} \tag{1.4}$$

  Note that frequencies have all the mathematical properties of probabilities as required by the axioms.

---

[4] The german word for "frequency" is "Häufigkeit". It is not to be mixed up with "Frequenz". The word "frequency" is derived from "frequent".

A more frequent event is also more likely to occur. Thus a rational decision would be to use the frequencies from this experiment as values for our set of probabilities.

$$P_i = f_i$$

Now we can apply these probabilities to estimate the outcome of the (non-laboratory) experiment. If we found that the die is not symmetric, we may bet on the outcome with the largest probability. **Note the subtle distinction between frequencies and probabilities:** Frequencies are an outcome of a series of experiments. They can be considered as a property of the system, on which a repeatable experiment can be performed. The probability composed from these frequencies is a property of the observer, and not of the system. It describes the expectations of the observer.

- Let us assume that we have the probabilities P_i^X and P_j^Y for two *independent* systems, X and Y individually. For example we could throw a red and a green die simultaneously and ask for the probability that red die shows the number i and the green die shows the number j. The probability for this event (i,j) is the product of the probabilities of the constituent events.

$$P_{i,j}^{XY} = P_i^X P_j^Y$$

- If the *a-priori* probabilities of elementary events are defined, we can calculate the probabilities of more complex events using the postulates for probabilities. The values of the *a-priori* probabilities are a postulate of a theory. By saying that the six faces of a die occur on-top we specify the *a-priori* probabilities. With these values I can calculate, for example, that the probability of showing a face with value greater or equal to 4 is 0.5.

1.2.1 Example:

In order to demonstrate the power of probability theory and how it may be related to real physics, let us investigate a simple toy model, that resembles closely the nature of real world problems in statistical physics. We can use this model later to demonstrate a number of principles of statistical physics.

Let us consider a toy model for surface adsorption. We would like to know the coverage of the surface with molecules. To simplify the problem, we consider a two-dimensional world with discrete states in the gas and on the surface. A molecule can be in each of the boxes shown in Fig. 1.3

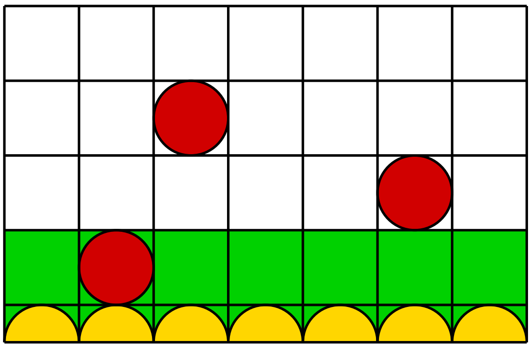


Fig. 1.3: Toy model for adsorption of molecules, indicated as red balls, on a surface. The available adsorption sites are indicated by the green boxes. The white boxes represent the available states in the gas.

Each of the three molecules, shown as red spheres in Fig. 1.3, may be in one of two configurations:

- it may be adsorbed on the surface, that is within one of the 7 green boxes, or
- it may be in the gas, that is within one of the 21 white boxes.

We would like to determine the mean coverage θ of surface sites. The coverage θ is defined as the ratio of adsorbed atoms N_A and the total number adsorption sites M_A, that is

$$\theta \overset{\text{def}}{=} \frac{N_A}{M_A} \tag{1.5}$$

In our toy model the coverage can have one of four possible values, namely $\theta \in \{0, \frac{1}{7}, \frac{2}{7}, \frac{3}{7}\}$, because $M_A = 7$ and $N_A \in \{0, 1, 2, 3\}$.

In order to estimate the probability $P(\theta)$ that the system has a certain coverage, we need to estimate the number of possible states $q(\theta)$ for each coverage. The probability is then obtained as the number of possible arrangements for a given coverage divided by the total number of possible arrangements.

$$P(\theta_i) = \frac{q(\theta_i)}{\sum_i q(\theta_i)} \tag{1.6}$$

As will be explained in the following, $q(\theta)$ can be expressed using some basic **combinatorics** by **factorials**[5] or the **binomial coefficients**[6] as follows:

$$q(\theta\frac{N_A}{M_A}) \overset{Eqs.\ 1.8,1.9}{=} \underbrace{\frac{M_A!}{(M_A - N_A)!N_A!}}_{q_A} \cdot \underbrace{\frac{M_G!}{(M_G - N_G)!N_G!}}_{q_G} = \underbrace{\binom{M_A}{N_A}}_{q_A} \cdot \underbrace{\binom{M_G}{N_G}}_{q_G} \tag{1.7}$$

The first term, q_A, is the number of ways in which N_A adsorbed molecules can be placed onto the M_A adsorption sites. The second term, q_G, is the number of ways in which $N_G = N - N_A$ molecules in the gas can be placed onto the M_G sites in the gas. N is the total number of molecules.

Let us first work out the expression for q_A:

1. We start with the first of the N_A molecules on the surface and place it onto one of the M_A sites. Obviously this can be done in M_A different ways.
2. No we place the second molecule on one of the $M_A - 1$ remaining adsorption sites. Thus we have $M_A(M_A - 1)$ different ways to place two molecules onto the surface.

   If we assume that the molecules are indistinguishable, however, we have double-counted each configuration. Let us assume that two sites, A and B, are occupied. We obtained this configuration once by placing the first molecule on A and the second one on B, or vice versa. Thus we have to divide the number of possibilities by two.

   $$q_A(\theta = \frac{2}{M_A}) = \frac{M_A(M_A - 1)}{2}$$

3. If we distribute N_A molecules onto the M_A sites, we proceed analogusly.
   (a) We distribute N_A in all possible ways onto the M_A adsorption sites. For the first molecule there are M_A sites. For the second molecule there are $M_A - 1$ adsorption sites left. For the third one there are $M_A - 2$ sites left, and so on. Thus there are $M_A(M_A - 1) \cdots (M_A - N_A + 1) = \frac{M_A!}{(M_A - N_A)!}$ arrangements of the molecules.

---

[5] The factorial $N!$ of the non-negative integer N is defined as $N! \overset{\text{def}}{=} 1 \cdot 2 \cdots N$. The factorial of zero is defined as $0! = 1$.

[6] The binomial coefficient $\binom{a}{b}$ is defined as $\binom{a}{b} = \frac{a!}{b!(a-b)!}$.

(b) With this construction, however, we have created all possible permutations of the order of the molecules for each choice of N_A adsorption sites. There are $N_A!$ permutations of the N_A molecules, which is seen as follows: For N_A adsorbed molecules, let us distribute the numbers from 1 to N_A onto the N_A molecules. There are N_A possibilities to place the "1" onto the N_A molecules. For each choice for the position of the "1", there are $N_A - 1$ remaining molecules to place the number "2". For each choice of the first two numbers, there are $N_A - 3$ choices for the number "3". In total we obtain $N_A(N_A-1)\cdots 2\cdot 1 = N_A!$ ways to distribute the numbers onto the molecules. This is exactly the number of permutations that we have been looking for.

(c) By dividing the number $M_A!/(M_A - N_A)!$ of arrangements by the number $N_A!$ of permutations we obtain the number $M_A!/[(M_A - N_A)!N_A!]$ of choices irrespective of their order.

Thus, we obtain

$$q_A(\theta = \frac{N_A}{M_A}) = \frac{M_A(M_A-1)\cdots(M_A-N_A+1)}{N_A(N_A-1)\cdots 1} = \frac{M_A!}{(M_A-N_A)!N_A!} = \binom{M_A}{N_A} \tag{1.8}$$

Now we evaluate, analogously to Eq. 1.8, the number q_G of ways in which N_G molecules can be arranged in the M_G states of the gas.

$$q_G = \frac{M_G!}{(M_G-N_G)!N_G!} = \binom{M_G}{N_G} \tag{1.9}$$

By multiplying the number of ways the N_A adsorbed molecules from Eq. 1.8 can be arranged on the surface and with the number of ways the N_G gas molecules can be placed in the gas from Eq. 1.9, we obtain the above equation Eq. 1.7.

Next we express our result Eq. 1.7 for q by the fundamental variables, the coverage θ defined in Eq. 1.5, the number N of molecules, the number M_A of adsorption sites and the number N_G of states in the gas. This implies $N_A = M_A\theta$, $N_G = N - N_A = N - M_A\theta$.

We obtain

$$q(\theta) \overset{\text{Eq. 1.7}}{=} \underbrace{\frac{M_A!}{(M_A(1-\theta))!(M_A\theta)!}}_{q_A} \cdot \underbrace{\frac{M_G!}{(M_G - N + \theta M_A)!(N - M_A\theta)!}}_{q_G} \tag{1.10}$$

This is a nasty expression, but it can be evaluated. For our example with $N = 3$, $M_A = 7$ and $M_G = 21$, we obtain

$$q(\theta = 0) = 1330 \qquad q(\theta = \frac{1}{7}) = 1470 \qquad q(\theta = \frac{2}{7}) = 441 \qquad q(\theta = \frac{3}{7}) = 35$$

With this result we obtain the probability for each coverage using Eq. 1.6, namely

$$P(\theta_i) = \frac{q(\theta_i)}{\sum_{i=0}^{3} q(\theta_i)}$$

and hence

$$\begin{aligned}
P(\theta = 0) &= \frac{1330}{3276} \approx 0.406 \\
P(\theta = \frac{1}{7}) &= \frac{1470}{3276} \approx 0.4487 \\
P(\theta = \frac{2}{7}) &= \frac{441}{3276} \approx 0.1346 \\
P(\theta = \frac{3}{7}) &= \frac{35}{3276} \approx 0.0107
\end{aligned}$$

Finally we can determine the average coverage as

$$\langle\theta\rangle = \sum_{i=0}^{3} P(\theta_i)\theta_i$$

which in our case yields $\langle\theta\rangle = 0.1071$. This was the result we were looking for.

While this example is highly simplified, it demonstrates how we can estimate the coverage of a surface with molecules from a simple counting argument. Problems like these are typical for statistical physics and the approach towards a solution is rooted in the same basic principles.

1.3 Information and Entropy

Imagine the weather forecast[7] of Los Angeles. People joke about it because it nearly always has about 20°C so that a weather forecast is nearly useless. One can safely leave the umbrella at home, and get wet only rarely. In other words, our expectations about the weather expressed by probabilities for all types of weather conditions, are changed very little by the weather forecast.

In contrast, the weather forecast in northern Europe is quite useful, if correct, unless we want to carry around an umbrella all the time. The weather forecast of Los Angeles adds little information compared to the one in northern Europe. Thus if we had to pay for the information, we would be willing to pay more for it in northern Europe than in Los Angeles.

However, note the difference between the amount of information gained by the answer of a question and the value a message. Even though weather forecast in Los Angeles adds little information, the message that it snows in Los Angeles would be tremendously valuable, because the consequences of this unlikely event are important. It is important to distinguish between the amount of information and its value.

The concept of **information** developed in this chapter will allow us to determine the information content of a message.

Why do we need a measure for information? If we are able to quantify information, we can choose the probabilities in a more intelligent way. With a better choice of probabilities, we can make better predictions. The concept of information helps us to make good guesses.

At the basis of the concept of information is that the amount of information is additive. What does it mean that information is additive? Let us consider a safe-cracker, who typically opens a safe in one hour. It will take him about three hours to open three safes. We conclude that it takes about three times as much information to open all three safes than only one. How is the information related to the number of possibilities? To open a safe he has to turn a knob to the right setting. Let us say that he has to choose from one-hundred possibilities. To determine the setting of the three safes, he needs to identify one out of $100 \times 100 \times 100$, that is one-million, possibilities. If the amount of missing information $S(N)$ is related to the number N of possibilities, we would conclude that $S(10^6) = 3S(100)$, or more generally $S(N^M) = MS(N)$.

Hence the functional dependence of $S(N)$ on the number of possibilities N is $S(N) = k\ln[N]$, an expression written on the grave of Ludwig Boltzmann, who found it[8]. k_B is called **Boltzmann constant**. For the time being we consider it as an arbitrary proportionality constant.

We can use the argument to estimate the information content of the page you are reading: There are 26 characters and 10 numbers. A line contains about 80 characters. A page contains 40 lines.

---

[7] The German word for "forecast" is "Vorhersage".

[8] Ludwig Boltzmann. Austrian Physicist, Feb. 20, 1844 (Vienna, Austria) -Oct. 5, 1906 (Trieste, that time Austria). Invented statistical physics by explaining thermodynamic equations from a statistical treatment of the microscopic equations of motion for the atoms. He explained how irreversible macroscopic equations can result from reversible microscopic equations. He obtained the Maxwell-Boltzmann equation in 1871. He moved around a lot and held chairs for theoretical physics(Graz), mathematics (Vienna), experimental physics (Graz), theoretical physics (Vienna). Suffering from mood swings and possibly out of disappointment about lack of acceptance of his theories, he committed suicide during a vacation in Trieste, shortly before his theories were confirmed experimentally.

Thus a particular text is one out of $(26+10)^{(80*40)} = 10^{4980}$ possibilities, which corresponds to an information of 11467 k_B. Not bad!

ENTROPY

The **entropy**[a] S is a **measure of the uncertainty** over the outcome. The entropy is proportional to the average number of yes-no questions needed to specify a specific outcome. Entropy and (undiscovered) information are synonyms.

---

[a] The term entropy has been introduced by Clausius in 1865. It is derived form the Greek word $\tau\rho o\pi\eta$ for transformation.

We define the entropy via the following axioms:

PROPERTIES OF THE ENTROPY

1. The entropy $\mathcal{S}$ is a function of the probability distribution $\{P_i\}$, as the latter describes our expectation or our state of knowledge. If an experiment has N possible outcomes, the entropy is a function $\mathcal{S}(\{P_i\}) = \mathcal{S}(P_1, \dots, P_N)$ of the N probabilities. If there is a continuous range of possible outcomes x the entropy is a functional $\mathcal{S}[p(x)]$ of the probability density $p(x)$.
2. The entropy is scalar. It must be scalar so that we can compare the amount of information.
3. For equal probabilities $P_i = \frac{1}{N}$, the entropy is a monotonically increasing function of N.
4. The entropy for a system of **statistically uncorrelated subsystems**[a] is the sum of the entropies for the subsystems.

$$\mathcal{S}(\{P^{XY}_{i,j}\}) = \mathcal{S}(\{P^X_i P^Y_j\}) = \mathcal{S}(\{P^X_i\}) + \mathcal{S}(\{P^Y_j\}) \tag{1.11}$$

In order to acquire complete knowledge about a combined system, that is to determine the state (i,j) completely, one requires as much information as is needed to determine the states of the two subsystems individually.

---

[a] Two systems are statistically uncorrelated if the outcome of one system has no influence on the outcome of the other. For example, if we throw two dice, the outcome of one die is independent of the other. An example of a correlated system are the questions: (1) “Does it rain?” and (2) “Is it cloudy?”: if it rains it is almost certainly cloudy. If it is cloudy the chances of rainfall are higher than if it is not.

1.4 Functional form of the Entropy: Shannon’s derivation

ENTROPY

The **entropy** has the functional form

$$\mathcal{S}(\{P_i\}) \stackrel{\text{def}}{=} -k \sum_i P_i \ln[P_i] \tag{1.12}$$

where k is an arbitrary constant. This expression is one of the most fundamental expressions in physics. In physics the constant k is chosen to be the Boltzmann constant.

We can easily verify that the expression Eq. 1.12 fulfills the requirements for the entropy functional

- it is a function(al) of the probabilities.
- it is a scalar

- For equal probabilities $\frac{1}{N}$ it is monotonically increasing function of N.

$$\mathcal{S}(\frac{1}{N},\frac{1}{N},\dots)=-k\sum_{i=1}^{N}\frac{1}{N}\ln[\frac{1}{N}]=+k\ln[N]$$

  Because the logarithm is monotonically increasing, this prooves the imposed requirement.

- For uncorrelated systems the entropy is additive.

$$\begin{aligned}
\mathcal{S}(\{P_i^X P_j^Y\}) &= -k\sum_{i,j} P_i^X P_j^Y \ln(P_i^X P_j^Y)\\
&= -k\sum_{i,j} P_i^X P_j^Y \left(\ln(P_i^X)+\ln(P_j^Y)\right)\\
&= \underbrace{\left(-k\sum_{i} P_i^X \ln(P_i^X)\right)}_{\mathcal{S}(\{P_i^X\})}\underbrace{\left(\sum_j P_j^Y\right)}_{=1}+\underbrace{\left(\sum_i P_i^X\right)}_{=1}\underbrace{\left(-k\sum_{j} P_j^Y \ln(P_j^Y)\right)}_{\mathcal{S}(\{P_j^Y\})}\\
&= \mathcal{S}(\{P_i^X\})+\mathcal{S}(\{P_j^Y\})
\end{aligned}$$

1.4.1 Shannon's proof

In the following we will derive the functional form for the entropy Eq. 1.12 following the work of Shannon[2].

Step 1

The requirement Eq. 1.11, saying that the entropy is additive, leads to another important statement: Let us consider a system with N possible events. We group the events into m subsets A_j, where the j-th subset holds n_j events. The events shall have probabilities of $P_{j,i}$, where j identifies to the subset A_j and the index i identifies the i-th event in the subset j. Ultimately we will quantify the information with the number of yes-no questions that specify a certain event. From this principle we derive the information for more complex questions. Therefore we divide the identification process into two questions: The first question is that to which of the subsets the correct event belongs. The probability that the event is in subset A_i is $W_j=\sum_{i=1}^{n_j}P_{j,i}$. Thus the information gained in the first question is $S(W_1,\dots,W_m)$. Once we know this answer, we pose the second question to specify the correct event in this subset, which requires the information $S(\frac{P_{j,1}}{W_j},\dots,\frac{P_{j,n_j}}{W_j})$. The normalization is required to ensure that the probabilities in that subgroup add to one. The two questions are statistically uncorrelated. Because of the additivity of the entropy for statistically uncorrelated subsystems, Eq. 1.11, we add the two entropies

On average, we obtain the information

$$\begin{aligned}
S(P_{1,1},\dots,P_{1,n_1},P_{2,1},\dots,P_{m,n_m}) &= \sum_{j=1}^{m}\sum_{i=1}^{n_j}P_{j,i}\left[S(W_1,\dots,W_m)+S(\frac{P_{j,1}}{\sum_{k=1}^{n_j}P_{j,k}},\dots,\frac{P_{j,n_j}}{\sum_{k=1}^{n_j}P_{j,k}})\right]\\
&= \sum_{j=1}^{m}W_j\left[S(W_1,\dots,W_m)+S(\frac{P_{j,1}}{W_j},\dots,\frac{P_{j,n_j}}{W_j})\right]\\
&= S(W_1,\dots,W_m)+\sum_{j=1}^{m}W_jS(\frac{P_{j,1}}{W_j},\dots,\frac{P_{j,n_j}}{W_j}) \qquad (1.13)
\end{aligned}$$

Example: Consider a die as sketched in Fig. 1.4. It has six faces with values $1-6$. Each face occurs with equal probability $P_i=1/6$. If we want to obtain the correct face of the die in one question we ask "Which is the correct face of the die?". The information of the answer is the entropy $S(P_1, P_2, P_3, P_4, P_5, P_6)$. Now we ask in a different manner: First we divide the faces into three subsets: G_1, G_2 and G_3. G_1 contains the faces $\{1, 3, 4\}$, G_2 contains $\{2, 6\}$ and the last set, G_3, has only one element, namely $\{5\}$. Now we obtain the correct face from two questions: The first question is "Which set contains the correct face?". The second question is "Which of the faces in this particular set is the correct one?". The first question yields the information $S(W_1, W_2, W_3)$, where $W_1 = P_1 + P_3 + P_4 = 1/2$, $W_2 = P_2 + P_6 = 1/3$ and $W_3 = P_5 = 1/6$ are the probabilities that the correct face is in one of the three subsets. If the answer has been G_1, the remaining uncertainty is $S(P_1/W_1, P_3/W_1, P_4/W_1) = S(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. If it is in set G_2 the remaining uncertainty is $S(\frac{1}{2}, \frac{1}{2})$, and if the answer is G_3 then the remaining uncertainty is $S(1)$. Depending on the answer of the first question the information we receive from the answer to the second question differs. Thus, we can only average the entropy over all possible events

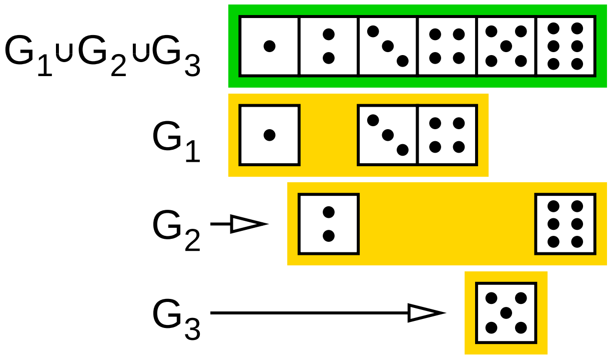


Fig. 1.4: Grouping of the six events of rolling a die into sets of 3,2 and 1 events. The probability that an event in G_1 takes place is $3 \times \frac{1}{6} = \frac{1}{2}$. For G_2 the probability is $2 \times \frac{1}{6} = \frac{1}{3}$ and that for G_3 is $\frac{1}{6}$.

face	probability	Information
1	$\frac{1}{6}$	$S(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) + S(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$
2	$\frac{1}{6}$	$S(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) + S(\frac{1}{2}, \frac{1}{2})$
3	$\frac{1}{6}$	$S(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) + S(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$
4	$\frac{1}{6}$	$S(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) + S(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$
5	$\frac{1}{6}$	$S(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) + S(1)$
6	$\frac{1}{6}$	$S(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) + S(\frac{1}{2}, \frac{1}{2})$

We obtain

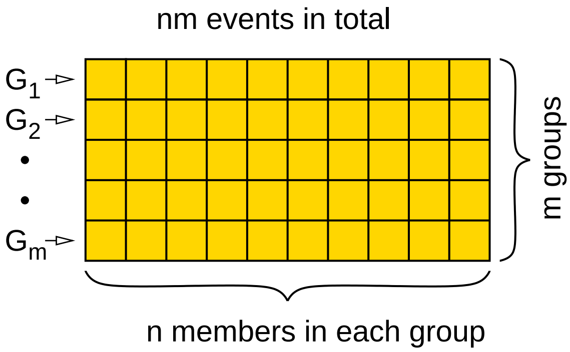
$$
\begin{aligned}
S\Big(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\Big) &= S\Big(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\Big) + \underbrace{\frac{3}{6}}_{W_1} S\Big(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\Big) + \underbrace{\frac{2}{6}}_{W_2} S\Big(\frac{1}{2}, \frac{1}{2}\Big) + \underbrace{\frac{1}{6}}_{W_3} S(1) \\
&= S(W_1, W_2, W_3) + W_1 S\Big(\frac{P_1}{W_1}, \frac{P_3}{W_1}, \frac{P_4}{W_1}\Big) + W_2 S\Big(\frac{P_2}{W_2}, \frac{P_6}{W_2}\Big) + W_3 S\Big(\frac{P_5}{W_3}\Big)
\end{aligned}
$$

which is identical to Eq. 1.13

Step 2

First we consider systems with equal probabilities. In order to simplify the discussion we introduce a new symbol f for the entropy with equal probabilities

$$f(n) \overset{\text{def}}{=} S\underbrace{\left(\frac{1}{n}, \dots, \frac{1}{n}\right)}_{n \text{ times}} \tag{1.14}$$



Now we specialize Eq. 1.13 to equal probabilities. Consider the case $P_{j,i} = \frac{1}{nm}$. That is, there are m groups with n events in each group. The first index j of $P_{j,i}$ is the group index and the second index i labels the individual event in the specified group. We obtain

$$\begin{aligned} \underbrace{S\left(\frac{1}{nm}, \dots, \frac{1}{nm}\right)}_{f(nm)} &\overset{\text{Eq. 1.13}}{=} \underbrace{S\left(\frac{1}{m}, \dots, \frac{1}{m}\right)}_{f(m)} + \sum_{j=1}^{m} \frac{1}{m} \underbrace{S\left(\frac{1}{n}, \dots, \frac{1}{n}\right)}_{f(n)} \\ \overset{\text{Eq. 1.14}}{\Rightarrow} \qquad f(nm) &= f(m) + f(n) \end{aligned} \tag{1.15}$$

This statement simply reproduces the additivity of the entropy for the special case of equal probabilities.

Step 3

From Eq. 1.15 we immediately obtain by successive application

$$\begin{aligned} & f(n^p) = f(n^{p-1}n) \overset{\text{Eq. 1.15}}{=} f(n) + f(n^{p-1}) \overset{\text{Eq. 1.15}}{=} 2f(n) + f(n^{p-2}) = \dots = pf(n) \\ \Rightarrow \qquad & f(n^p) = pf(n) \end{aligned} \tag{1.16}$$

This already shows that $f = k \ln[n]$ satisfies this requirement. In the following, we will show that this is the only possible solution.

Step 4

We start with $n = 2$ and obtain [9] for $p \in \mathbb{N}_0$

$$f(2^p) \overset{\text{Eq. 1.16}}{=} pf(2) = \underbrace{\frac{\ln(2^p)}{\ln(2)}}_{=p} f(2) = \underbrace{\frac{f(2)}{\ln(2)}}_{=:k} \ln(2^p)$$

---

[9] $\mathbb{N}_0$ denotes the set of natural numbers including zero.

Thus, with the definition

$$k \overset{\text{def}}{=} \frac{f(2)}{\ln(2)}$$

we obtain the desired relation for powers of 2

$$\Rightarrow \qquad f(n) = k\ln(n) \qquad \text{for} \qquad n = 2^p \tag{1.17}$$

Step 5

Now we need to show that we obtain the same result, Eq. 1.17, also, if n cannot be represented by a power of 2:

Let us pick an arbitrary $n \in \mathbb{N}$. For each integer m, we determine the next power $q_m \in \mathbb{N}$ of 2 which lies just below n^m, so that

$$2^{q_m} \leq \quad n^m \quad \leq 2^{q_m+1} \tag{1.18}$$

$$\overset{\ln}{\Rightarrow} \qquad q_m\ln(2) \leq m\ln(n) \leq (q_m+1)\ln(2)$$

$$\Rightarrow \qquad q_m \leq \; m\tfrac{\ln(n)}{\ln(2)} \; \leq (q_m+1)$$

$$\Rightarrow \qquad m\frac{\ln(n)}{\ln(2)} - 1 \leq \quad q_m \quad \leq m\frac{\ln(n)}{\ln(2)} \tag{1.19}$$

Now we use the requirement that the entropy for equal probabilities is monotonically increasing

$$\text{Eq. 1.18} \Rightarrow \qquad f(2^{q_m}) \leq f(n^m) \leq f(2^{q_m+1})$$

$$\overset{\text{Eqs. 1.16,1.17}}{\Rightarrow} \qquad kq_m\ln(2) \leq mf(n) \leq k(q_m+1)\ln(2)$$

$$\Rightarrow \qquad q_m\frac{k\ln(2)}{m} \leq \; f(n) \; \leq (q_m+1)\frac{k\ln(2)}{m} \tag{1.20}$$

Now we expand the limits by using Eq. 1.19, which can be converted as follows:

- the left inequality in Eq. 1.20 can be expanded to

$$f(n) \overset{\text{Eq. 1.20}}{\geq} q_m\frac{k\ln(2)}{m} \overset{\text{Eq. 1.19}}{\geq} \left(m\frac{\ln(n)}{\ln(2)} - 1\right)\frac{k\ln(2)}{m} = k\ln(n) - \frac{k\ln(2)}{m} \tag{1.21}$$

- the right inequality in Eq. 1.20 can be expanded to

$$f_n \overset{\text{Eq. 1.20}}{\leq} (q_m+1)\frac{k\ln(2)}{m} \overset{\text{Eq. 1.19}}{\leq} \left(m\frac{\ln(n)}{\ln(2)} + 1\right)\frac{k\ln(2)}{m} \leq k\ln(n) + \frac{k\ln(2)}{m} \tag{1.22}$$

Thus, we obtain

$$k\ln(n) - \frac{1}{m}k\ln(2) \overset{\text{Eq. 1.21}}{\leq} f(n) \overset{\text{Eq. 1.22}}{\leq} k\ln(n) + \frac{1}{m}k\ln(2)$$

For $m \to \infty$ the distance of the brackets vanishes, so that we can deduce uniquely

$$f(n) = k\ln(n) \qquad \text{for} \qquad n \in \mathbb{N} \tag{1.23}$$

Step 6

We now use the above result to derive the entropy for unequal probabilities.

We start with a system with N possible events, which we divide into subsets with n_j events each. We divide the task of identifying the correct event into (1) identifying the correct subset and (2) the effort to identify the correct event in that subset. We define

$$P_j \overset{\text{def}}{=} \frac{n_j}{N} \tag{1.24}$$

We use condition Eq. 1.13 for equal probabilities.

$$\begin{aligned}
\underbrace{S(\frac{1}{N},\ldots,\frac{1}{N})}_{k\ln[N]} &\overset{\text{Eq. 1.13}}{=} S(\frac{n_1}{N},\ldots,\frac{n_m}{N}) + \sum_{i=1}^{m}\frac{n_j}{N}\underbrace{S(\frac{1}{n_j},\ldots,\frac{1}{n_j})}_{k\ln[n_j]} \\
\overset{\text{Eq. 1.23}}{\Rightarrow} \quad S(\frac{n_1}{N},\ldots,\frac{n_m}{N}) &= k\ln[N] - k\sum_{i=1}^{m}\frac{n_j}{N}\ln[n_j] \\
\overset{\text{Eq. 1.24}}{\Rightarrow} \quad S(P_1,\ldots,P_m) &= k\ln[N] - k\sum_{j=1}^{m}P_j\underbrace{\ln[P_jN]}_{=\ln[P_j]+\ln[N]} \\
&= -k\sum_{j=1}^{m}P_j\ln[P_j]
\end{aligned}$$

This is Eq. 1.12 on p. 18, which we wanted to derive. *q.e.d.*

1.4.2 Discussion of the functional form of the entropy

In statistical mechanics, the constant k is chosen to be the Boltzmann constant $k_B = 1.380658 \times 10^{-23}$ J/K. Thus, the product of entropy and temperature has the unit energy. The choice for k, adopted in information theory, is $1/\ln(2)$. These choices are in principle arbitrary and have historical reasons. We will return to this point later.

The function $-P\ln[P]$ shown in Fig. 1.5 is asymmetric, because a probability close to one forces all other probabilities to remain close to zero, whereas a probability close to zero has no implications on the remaining probabilities. Note that the slope of $-P\ln[P]$ is infinite for $P = 0$. Thus, trying a Taylor expansion about the origin is a bad idea.

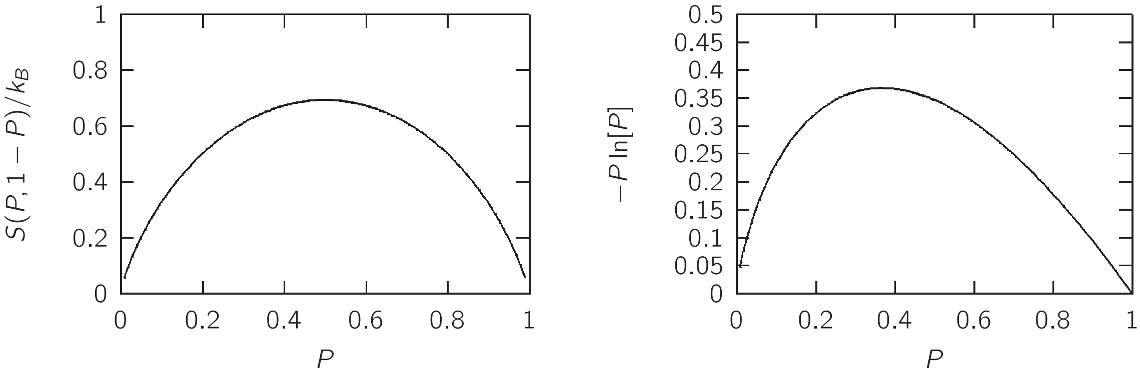


Fig. 1.5: Left: Entropy of a yes-no question $S(P, 1-P)$ as function of P. The maximum is located at $P = \frac{1}{2}$ and has the value $k\ln(2)$. Right: $-P\ln(P)$ as function of P. The entropy is always positive (if k is positive), because $-kP\ln(P) \geq 0$ if $0 < P < 1$. $-kP\ln(P) = 0$ for $P = 0$ and $P = 1$. The maximum of $-kP\ln(P)$ is k/e and is located at $P = 1/e$. The slope is infinite at $P = 0$ and and $-k$ at P=1.

To become familiar with the entropy, let us consider the entropy related to a yes-no question, a question with only two possible answers. Each answer, will be given with a certain probability. Owing to the normalization constraint, this entropy can be expressed as a function of only one variable. The result $S(P_1, P_2) = -k_B P_1 \ln(P_1) - k_B P_2 \ln(P_2)$ is shown in Fig. 1.5.

The entropy with $k\ln[2]$ is largest if both possible answers have the same probability. This shows for example that the most efficient strategy to reach certainty with yes-no questions is to divide the possible events into subsets of approximately equal initial probabilities. Thus, we reduce the entropy of the problem with each question by the maximum amount. Note, however, that the entropy is a property of the question, not the answer. By directly asking if a particular event is the correct one, we can, by chance, reach certainty within one step. On average, however, we will require more questions so that this is a bad strategy.

Fig. 1.5 also shows that the answer to a question, for which we already expect a certain answer with large probability, has less of a value than that to a question for which we have no expectations.

Let us consider two special cases:

- Certainty: The information is complete, if the state in which the system is realized, namely state i_0, is known. The probability distribution is then $P_i = \delta_{i,i_0}$ and the entropy vanishes according to Eq. 1.12.[10]

  $$S = \mathcal{S}(\{P_i = \delta_{i,i_0}\}) = -k \sum_i \delta_{i,i_0} \ln[\delta_{i,i_0}] = 0$$

  This is obviously the smallest entropy a system can possess. Since the system is completely known there is no further information that can be gained about it. Note that we have introduced a new symbol S as opposed to $\mathcal{S}$. $\mathcal{S}$ is used for the functional depending on the probabilities, whereas S is used for a particular value of the entropy.

- Uncertainty: Nothing is known about a finite system with N states. The probabilities are chosen[11] as

  $$P_i = \frac{1}{N}$$

  The entropy is

  $$S = \mathcal{S}(\{P_i = \frac{1}{N}\}) = -k \sum_{i=1}^{N} \frac{1}{N} \ln[\frac{1}{N}] = k\ln[N]$$

  This is the maximum entropy a system with N states can have.

1.4.3 Physical motivation for the entropy:

Here we show that the entropy Eq. 1.12 is directly related to the number of configurations. To be more specific we will demonstrate this relation for a simple example, namely a large set of N identical harmonic oscillators as sketched in fig. 1.6.

Given a vector $(N_0, N_1, N_2, \ldots)$ of occupation numbers N_j, where N_j is the number of harmonic oscillators in their j-th excited states, we will work out the number $Q(N_0, N_1, \ldots)$ of configurations

---

[10] Note that $x\ln[x] = 0$ for $x = 0$

[11] The probability can be obtained more rigorously by maximizing the entropy functional. The condition that the probabilities add up to one is enforced using the *method of Lagrange multipliers.*

$$\begin{aligned}
\frac{d\mathcal{S}(\{P_j\}) - \lambda \sum_j P_j}{dP_i} &= -k\ln[P_i] - k - \lambda = 0 \\
\ln[P_i] &= -1 - \frac{\lambda}{k} \\
P_i &= \exp(-1 - \frac{\lambda}{k}) = \frac{1}{N}
\end{aligned} \tag{1.25}$$

In the last step we used that value of λ that fulfills the norm.

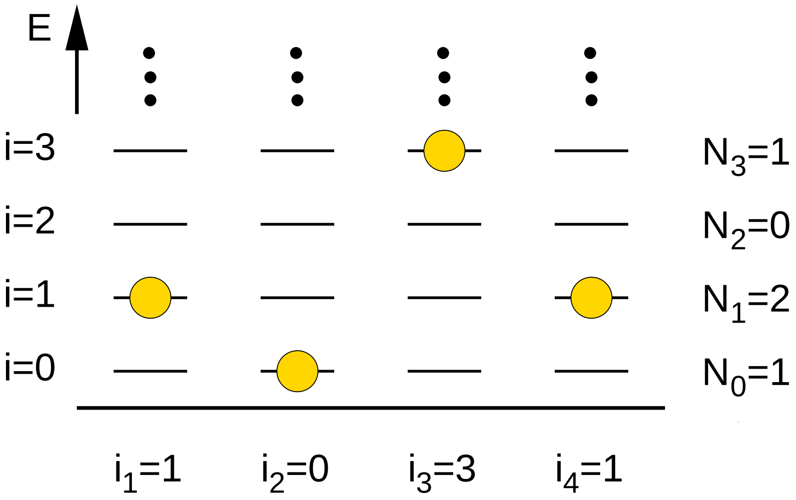


Fig. 1.6: Definition of the variables used in the model for oscillators for $N = 4$ oscillators. Golden balls indicate the state of the oscillator.

with this set of occupation numbers.[12] Each individual configurations of the $N_{tot} = \sum_j N_j$ harmonic oscillators is characterized by a N-dimensional vector of excitations $(i_1, i_2, \ldots, i_N)$.

We will find, that the number of configurations can be expressed as

$$Q(N_1, N_2 \ldots) \overset{\text{Eq. 1.30}}{=} \exp\left[\frac{N}{k}\mathcal{S}\left(\frac{N_0}{N}, \frac{N_1}{N}, \ldots\right)\right]$$

where $\mathcal{S}$ is the Shannon entropy from Eq. 1.12. This gives the entropy another meaning besides the missing information.

Let us begin: We count the number $Q(N_0, N_1, N_2, \ldots)$ of ways one can prepare, out of $N = \sum_i N_i$ oscillators, N_0 oscillators in the ground state, N_1 oscillators in the first excited state, N_2 oscillators in the second excited state and so on.

1. Given that N_0 out of N oscillators are in the ground state we need to find out the number of configurations. We begin by choosing one out of N oscillators which we determine to be in the ground state. Clearly there are N possibilities. Then we select the second oscillator, which shall be in the ground state. We have $N - 1$ possibilities, since one is already selected. Thus we have $N(N - 1)$ possibilities to fix the first two states. However, we have counted each configuration twice, because one of the two selected oscillators could have been the first or the second to be selected. Both choices correspond to the same configuration. Thus we have to divide the result by two.

   Let us demonstrate the same idea on an example of three oscillators with two oscillators in the ground state. Fig. 1.7 illustrates the argument.

   Avoiding double counting we obtain $N\frac{N-1}{2}$ possibilities to select the first two oscillators as being in the ground state. For the third one we have $N - 2$ possibilities. Now we need to divide by three to avoid double counting.

   Once we selected all oscillators which are in the ground state we count

$$\frac{N(N-1)(N-2)\cdots(N-N_0+1)}{1\cdot 2\cdot 3\cdots N_0} = \frac{N(N-1)\cdots(N-N_0+1)}{N_0!} \qquad (1.26)$$

---

[12] The occupation numbers imply that N_0 of these oscillators are in the ground state, N_1 harmonic oscillators are in the first excited state and so on. There are infinitely many excitation levels and the occupation numbers add up to the total number N of oscillators.

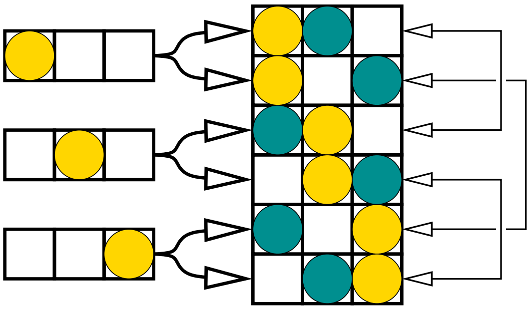


Fig. 1.7: Explanation of the double counting correction. Let us consider that there are three oscillators, of which two are in the ground state. Each oscillator is represented by a square and the three squares side-by-side represent all three oscillators. By placing a sphere in a box, we specify this harmonic oscillator to be in the ground state. First we select the first oscillator to be in the ground state, which we indicate by placing a yellow sphere in one of the boxes. There are three possibilities, shown on the left. On the right we select the second harmonic oscillator to be in the ground state. For each of the three possibilities there are two possibilities left, resulting in six possibilities over all. However, we have double counted each configuration, because interchanging the two colors does not correspond to a different choice: It is still the same two harmonic oscillators that are in the ground state. Thus we need to divide our result for the number of configurations by two to avoid double counting.

2. Now we select all oscillators which are in the first excited state. For each selection of oscillators in the ground state, we have now

$$\frac{(N-N_0)(N-N_0-1)\cdots(N-N_0-N_1+1)}{N_1!} \tag{1.27}$$

arrangements. If we multiply the result Eq. 1.27 with the number of choices for oscillators ion the ground state from Eq. 1.26, we obtain

$$\frac{N(N-1)\cdots(N-N_0-N_1+1)}{N_0!N_1!}$$

3. Now we proceed analogously for the second, third, etc excited state and obtain

$$Q(N_0,N_1,N_2,\ldots)=\frac{N(N-1)\cdots(N-\overbrace{\sum_{i=0}^{\infty}N_i}^{=N}+1)}{N_0!N_1!N_2!\ldots}=\frac{N!}{\prod_{i=0}^{\infty}N_i!}$$

Since it is difficult to work with the factorials, let us simplify the expression using Stirling's formula

STIRLING'S FORMULA

$$\ln(n!)\approx\left(n+\frac{1}{2}\right)\ln(n)-n+\frac{1}{2}\ln(2\pi)\approx n\ln(n)-n \tag{1.28}$$

The functional form of $x\ln(x)-x$ is shown in Fig. 1.8. A derivation of Stirling's formula can be found in appendix D.1

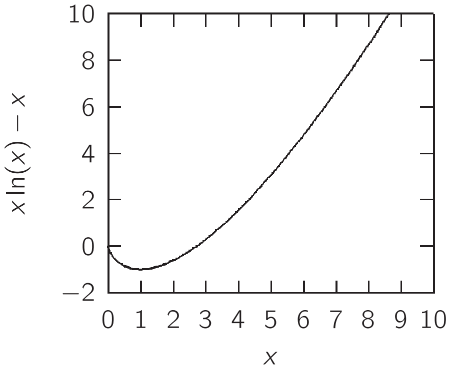


Fig. 1.8: Functional form of $x\ln(x) - x$ which appears in Stirling's formula Eq. 1.28.

We obtain $\ln(Q)$ as

$$
\begin{aligned}
\ln[Q(N_0, N_1, \ldots)] &= \ln\left[\frac{N!}{\prod_{i=0}^{\infty} N_i!}\right] \overset{\ln(ab)=\ln(a)+\ln(b)}{=} \ln[N!] - \sum_{i=0}^{\infty} \ln[N_i!] \\
&\overset{Stirling}{\approx} N\ln[N] - N - \sum_{i=0}^{\infty} (N_i \ln[N_i] - N_i) \\
&= \underbrace{N}_{\sum_{i=0}^{\infty} N_i} \ln[N] - N - \sum_{i=0}^{\infty} N_i \ln[N_i] + \underbrace{\sum_{i=0}^{\infty} N_i}_{N} \\
&= -\sum_{i=0}^{\infty} N_i \left(\ln[N_i] - \ln[N]\right) \\
&= N \sum_{i=0}^{\infty} \left(-\frac{N_i}{N} \ln[\frac{N_i}{N}]\right)
\end{aligned}
$$

If we identify the relative occupation by the probability of finding an oscillator in a particular state, that is

$$
P_i = \frac{N_i}{N}
$$

we obtain

$$
\frac{1}{N} \ln[Q(NP_0, NP_1, \ldots)] = -\sum_{i=0}^{\infty} P_i \ln[P_i] \overset{Eq.\ 1.12}{=} \frac{1}{k} S(P_0, P_1, \ldots) \tag{1.29}
$$

which leads us to one definition of the entropy $\mathcal{S}(P_0, P_1, \ldots)$. Conversely we can express the number of arrangements by the entropy as

$$Q(NP_0, NP_1, \ldots) \overset{\text{Eq. 1.29}}{=} \mathrm{e}^{\frac{N}{k}\mathcal{S}(P_0,P_1,\ldots)} \tag{1.30}$$

Thus we obtain a second interpretation of the entropy. Given N identical systems[a], the number Q of possible arrangements with a given set of relative frequencies P_i is given by the entropy as e^{NS}. The entropy is given exactly by Shannon's formula, Eq. 1.12.

---

[a] While we used the model of harmonic oscillators, we never used any specific property of harmonic oscillators. The result is therefore more general.

An important result is that the relative number of two different sets $(N_0, N_1, \ldots,)$ of occupations grows very rapidly with increasing N.

$$\frac{Q(NP'_0, NP'_1, \ldots)}{Q(NP_0, NP_1, \ldots)} \overset{\text{Eq. 1.29}}{=} \exp\left[\frac{N}{k}\Big(\mathcal{S}(P'_0, P'_1, \ldots) - \mathcal{S}(P_0, P_1, \ldots)\Big)\right] \tag{1.31}$$

If the ratio is of order one for a microscopic system with about 10 oscillators, the configuration with smaller entropy becomes entirely neglegible for a macroscopic system with N in the range of 10^{23}. This is the origin for the power of thermodynamics.

As N becomes very large, there is a single distribution $\{P_i\}$, the one that maximizes the entropy S, that describes so many more different configurations than any other, that only this probability distribution is of any relevance. This allows us to select one probability distribution $\{P_0, P_1, \ldots\}$ from all the others. This conclusion is at the heart of the maximum-entropy principle, that we will introduce in the following.

1.5 Maximum-Entropy Principle

The maximum-entropy principle[3, 4] is at the heart of statistical mechanics and thermodynamics. It allows one to select one probability distribution that is optimally unbiased, while considering the knowledge that is available for the system.

In order to make a prediction, we need to estimate the probability that a physical system is in a given state. This is done using the maximum-entropy principle. The probabilities allow us to make predictions about the outcome of an experiment. (The predictions will not be very good if the probabilities are chosen poorly.)

Fig. 1.9: Edwin Thompson Jaynes, 1922-1989. American Physicist. Invented the maximum-entropy principle and showed its relation to thermodynamics.

For example we can estimate the **mean value** $\langle A\rangle$ of a quantity A measured repeatedly by an experiment as

$$\langle A\rangle = \sum_i A_i P_i \tag{1.32}$$

where A_i is the value of A when the system is in its i-th state. We can not only predict the mean value of a property, but also estimate the certainty of the prediction, by estimating the **standard deviation**

$$\sigma = \sqrt{\langle (A - \langle A\rangle)^2\rangle} = \sqrt{\langle A^2\rangle - \langle A\rangle^2} \tag{1.33}$$

The **maximum-entropy principle**[3, 4] says that the best probabilities are those, which (1) are compatible with the state of knowledge and (2) maximize the entropy $\mathcal{S} = -k\sum_i P_i \ln[P_i]$ under the restrictions of (1). Since the entropy is a measure of the lacking information about a system, any set of probabilities that does not maximize the entropy contains additional unknown assumptions. Thus, it does not correspond to an unbiased choice.

From the maximum-entropy principle follows immediately that the probabilities are equal for a system with N possible events and no additional knowledge.

The maximum-entropy principle can, however, be applied to much more complex problems. For example, the state of knowledge about an experiment may be determined by the mean values of some quantities $A, B, \ldots$. Then we can estimate the probabilities for the outcome of the experiment for variables that have not been explored before. This is the typical type of problems we encounter in statistical physics, which we will discuss in this section.

1.5.1 Derivation of the maximum-entropy principle

Let us imagine a situation where nothing but an average property $\bar{A} = \sum_i P_i A_i$ is known about the state of a system. This case is most important for large systems, where only average properties are measurable. It is also relevant for dynamical systems with transitions between states that are too rapid to be observed, while an average property can still be measured. (Of course, the values A_i for each state must be known as well.)

We will now determine the probability and the entropy of the system by maximizing the entropy under the constraints of the known fact, namely

$$\bar{A} = \sum_i P_i A_i . \tag{1.34}$$

Following the method of Lagrange multipliers[13], we extend the entropy functional by the constraints

$$\begin{aligned}\mathcal{Y}(\bar{A}, \{P_i\}, \alpha, \lambda) &\stackrel{\text{def}}{=} \mathcal{S}(\{P_i\}) - \lambda(\sum_i P_i - 1) - \alpha(\sum_i P_i A_i - \bar{A}) \\ &= -k \sum_i P_i \ln[P_i] - \lambda(\sum_i P_i - 1) \\ &\quad -\alpha(\sum_i P_i A_i - \bar{A})\end{aligned} \tag{1.35}$$

where λ is the Lagrange multiplier for the condition that the probabilities must sum to 1, and α is the Lagrange multiplier for the constraint Eq. 1.34 saying that the probabilities must produce the selected mean value $\bar{A}$. If the constraints are fulfilled, the **extended entropy** $\mathcal{Y}$ is identical to the entropy $\mathcal{S} = -k \sum_i P_i \ln[P_i]$, because, in that case, the additional terms, describing the constraints, vanish.

According to the method of Lagrange multipliers, the maximum of the entropy under the constraints is obtained by solving the **equilibrium condition**[14]

$$0 = \frac{\partial \mathcal{Y}}{\partial P_i} = -k_B \ln(P_i) - k_B - \lambda - \alpha A_i \tag{1.36}$$

and the **constraint conditions**

$$0 = \frac{\partial \mathcal{Y}}{\partial \lambda} = \left(\sum_i P_i \right) - 1 \tag{1.37}$$

$$0 = \frac{\partial \mathcal{Y}}{\partial \alpha} = \left(\sum_i P_i A_i \right) - \bar{A} \tag{1.38}$$

---

[13] For a description of the method of Lagrange multipliers see ΦSX: Klassische Mechanik

[14] The term equilibrium condition is not standard terminology.

In the following we will determine probabilities that fulfill these equations.

Let us start with the equilibrium condition Eq. 1.36, from which we obtain a parameterized form for the probabilities

$$P_i^{eq}(\alpha,\lambda) = \mathrm{e}^{-1-\frac{\lambda}{k}} \cdot \mathrm{e}^{-\frac{\alpha}{k}A_i} \tag{1.39}$$

The Lagrange multiplier λ for the normalization can be determined from the normalization constraint Eq. 1.37, that the probabilities add up to one. [15] Thus, we obtain the probabilities that only depend on α.

$$P_i^{eq}(\alpha) = \frac{1}{Z(\alpha)}\mathrm{e}^{-\frac{\alpha}{k}A_i} \;, \tag{1.40}$$

where

$$Z(\alpha) \stackrel{\mathrm{def}}{=} \sum_i \mathrm{e}^{-\frac{\alpha}{k}A_i} \tag{1.41}$$

is called the **partition function**. The partition function plays a central role in statistical mechanics far beyond that of a simple normalization factor, but we will come to that later. We call probability distributions of the form Eq. 1.40 **equilibrium distributions**.

The exponential factor $\mathrm{e}^{-\frac{\alpha}{k}A_i}$ in Eq. 1.40 is related to the **Boltzmann factor** $\mathrm{e}^{-\frac{1}{k_BT}E_i}$, that is ubiquitous[16] in statistical processes. The expression given above is actually the conceptual origin for the Boltzmann factor.

Since the correct probability distribution is of the form Eq. 1.40, we restrict our search from here on to these distributions, which we call **equilibrium distributions**.[17]

Let us evaluate the function Y which is obtained from $\mathcal{Y}$ by inserting the equilibrium probabilities:

$$\begin{aligned}
Y(\alpha,\bar{A}) &\stackrel{\mathrm{def}}{=} \mathcal{Y}(\bar{A},\{P_i^{eq}(\alpha)\},\alpha,\lambda(\alpha)) \\
&\stackrel{\text{Eqs. 1.35,1.40}}{=} \underbrace{-k\sum_i P_i^{eq}(\alpha)\overbrace{\left(-\ln[Z(\alpha)] - \frac{\alpha}{k}A_i\right)}^{\ln[P_i^{eq}(\alpha)]}}_{\mathcal{S}(\{P_i(\alpha)\})} - \alpha\left(\sum_i P_i^{eq}(\alpha)A_i - \bar{A}\right) \\
&= k\ln[Z(\alpha)] + \alpha\bar{A}
\end{aligned}$$

The dependence on λ drops out, because the corresponding constraint is already fulfilled. Finally, we have to fulfill the second constraint equation, which leads us to[18]

$$0 = \frac{\partial Y}{\partial\alpha} \qquad \Rightarrow \qquad \bar{A} = -k\frac{\partial\ln[Z(\alpha)]}{\partial\alpha} \tag{1.42}$$

---

[15] We start from the constraint condition for λ to evaluate the prefactor.

$$\frac{\partial\mathcal{Y}}{\partial\lambda} \stackrel{Eq.\ 1.37}{=} 0 \Rightarrow 1 = \sum_i P_i \stackrel{Eq.\ 1.39}{=} \sum_i \mathrm{e}^{-1-\frac{\lambda}{k}} \cdot \mathrm{e}^{-\frac{\alpha}{k}A_i} = \mathrm{e}^{-1-\frac{\lambda}{k}}\sum_i \mathrm{e}^{-\frac{\alpha}{k}A_i}$$

$$\Rightarrow \mathrm{e}^{-1-\frac{\lambda}{k}} \quad = \quad \frac{1}{\sum_i \mathrm{e}^{-\frac{\alpha}{k}A_i}} = \frac{1}{Z(\alpha)} \qquad \Rightarrow \qquad \lambda(\alpha) = k\ln[Z(\alpha)] - k$$

The prefactor can be resolved for λ. Since we need λ in the exponential we do not need to resolve for it.

[16] german(“ubiquitous”)=”allgegenwärtig”

[17] The equilibrium distributions satisfy the equilibrium conditions, but not yet the constraint conditions.

[18]

$$\frac{\partial Y}{\partial\alpha} = \sum_i \underbrace{\frac{\partial\mathcal{Y}}{\partial P_i^{eq}}}_{=0 \quad \text{Eq. 1.36}} \frac{\partial P_i^{eq}}{\partial\alpha} + \frac{\partial\mathcal{Y}}{\partial\alpha} + \underbrace{\frac{\partial\mathcal{Y}}{\partial\lambda}}_{=0 \quad \text{Eq. 1.37}} \frac{\partial\lambda}{\alpha} = \frac{\partial\mathcal{Y}}{\partial\alpha} \stackrel{\text{Eq. 1.38}}{=} 0$$

This equation can be resolved for $\alpha(\bar{A})$, which provides us with the desired probability distribution

$$P_i(\bar{A}) = \frac{1}{Z(\alpha(\bar{A}))} \mathrm{e}^{-\frac{\alpha(\bar{A})}{k} A_i} \tag{1.43}$$

Thus we have obtained our final result, namely the most unbiased probabilities consistent with our current state of knowledge. The latter has been incorporated via constraints, that fix certain mean values.

Given the probabilities from Eq. 1.43 we can make predictions about other physical quantities B, via

$$\bar{B} \stackrel{\text{def}}{=} \langle B \rangle = \sum_i P_i(\bar{A}) B_i$$

Such a prediction is not a law of nature, but it rather plays the role of an "educated guess". Later, we will see that the concept just demonstrated allows to make surprisingly accurate predictions for many physical systems.

Note, that once all constraint equations are fulfilled, the entropy is equal to the function $\mathcal{Y} = Y$. Thus, the entropy has the form

$$S(\bar{A}) = k \ln[Z(\alpha(\bar{A}))] + \alpha(\bar{A})\bar{A} \tag{1.44}$$

The entropy will be important to extend the principle of maximum entropy to composite[19] systems. In that case we do not need to refer to the individual probabilities in each subsystem, but one can work directly with mean values. This abstraction from the individual states of each subsystem is what represents later the step from statistical physics to thermodynamics.

COOKING RECIPE

Let us summarize the main steps that actually need to be done:

1. evaluate the partition function

$$Z(\alpha) = \sum_i \mathrm{e}^{-\frac{\alpha}{k} A_i} .$$

2. determine the relation between the expectation value and the Lagrange multiplier

$$\bar{A} = -k \frac{\partial \ln[Z(\alpha)]}{\partial \alpha} \rightarrow \alpha(\bar{A}) .$$

3. determine the probability distribution using

$$P_i(\bar{A}) = \frac{1}{Z(\alpha(\bar{A}))} \mathrm{e}^{-\frac{\alpha(\bar{A})}{k} A_i} ,$$

   which is the desired result

4. determine the thermodynamic expectation values of any observable B as

$$\langle B \rangle = \sum_i P_i^{eq}(\bar{A}) B_i$$

   for comparison with experiment or the entropy

$$S(\bar{A}) = -k \sum_i P_i^{eq}(\bar{A}) \ln\left(P_i^{eq}(\bar{A})\right) = k \ln\left(Z(\alpha(\bar{A}))\right) + \alpha(\bar{A})\bar{A} .$$

---

[19] german(composite)=zusammengesetzt

1.5.2 Physical meaning of Lagrange multipliers

Let us consider two systems X and Y. We do know the initial values $\bar{A}_X$ and $\bar{A}_Y$ of the quantity A in the two systems. When the two systems are brought into contact, the values of A in the two subsystems may change individually. However, we require that the total amount $A_{tot} \overset{\text{def}}{=} \bar{A}_X + \bar{A}_Y$ remains constant. A_{tot} is a conserved quantity. Hence, the quantity A can "flow" from one system to the other, but, overall, it is conserved.

Many quantities, we encounter in physics, are conserved quantities such as energy, momentum, charge, particle number, volume, etc.[20] For example, energy can flow from one subsystem into the other, but the total energy of both systems is conserved.

We make use of the fact that the entropy of the two systems is additive, and that the total entropy is maximized in thermal equilibrium.

$$S_{XY}(A_{tot},\Delta) = S_X(\overbrace{\frac{1}{2}A_{tot} + \frac{1}{2}\Delta}^{\bar{A}_X}) + S_Y(\overbrace{\frac{1}{2}A_{tot} - \frac{1}{2}\Delta}^{\bar{A}_Y})$$

where $\Delta \overset{\text{def}}{=} \bar{A}_X - \bar{A}_Y$. Δ is a free parameter, which will adjust via transitions between states in order to maximize the entropy.

$$\begin{aligned} \frac{d}{d\Delta}S_{XY}(A_{tot},\Delta) &= \frac{1}{2}\left.\frac{dS_X}{d\bar{A}}\right|_{\bar{A}_x} - \frac{1}{2}\left.\frac{dS_Y}{d\bar{A}}\right|_{\bar{A}_y} = 0 \\ \Rightarrow \qquad \frac{dS_X}{d\bar{A}} &= \frac{dS_Y}{d\bar{A}} \end{aligned}$$

Hence the conserved quantity A will move back and forth between the two subsystems until the system reaches a point, where the increase of the entropy of one system is identical to the loss of entropy of the other.

We will show now that the derivative of the entropy can be identified directly with the Lagrange multiplier. We use Eq. 1.44 and Eq. 1.42

$$\frac{dS}{d\bar{A}} \overset{\text{Eq. 1.44}}{=} k\frac{d\ln[Z(\alpha)]}{d\alpha}\frac{d\alpha}{d\bar{A}} + \frac{d\alpha}{d\bar{A}}\bar{A} + \alpha(\bar{A}) = \underbrace{\left(k\frac{d\ln[Z(\alpha)]}{d\alpha} + \bar{A}\right)}_{=0 \quad \text{due to Eq. 1.42}}\frac{d\alpha(\bar{A})}{d\bar{A}} + \alpha$$

$$\Rightarrow \frac{dS}{d\bar{A}} = \alpha \tag{1.45}$$

EQUILIBRIUM AND LAGRANGE MULTIPLIERS

Thus, if two systems X and Y are in contact, so that they can exchange one globally conserved quantity A, the Lagrange multipliers α_X and α_Y related to the quantity A in the two systems are identical, if the two systems are in equilibrium.

$$\alpha_X = \alpha_Y$$

Hence, two systems that are in contact, such that a globally conserved quantity A can be exchanged,

---

[20] A philosophical note: We encounter so many conserved quantities, more because physicists like them, rather than because the real world contains so many conserved quantities. Most theoretical studies become feasible only when symmetries are imposed as model assumptions, and symmetries result in conserved quantities. Most experiments become understandable only under simplified conditions. This, however, is not a limitation of the physical theory, because violations of conservation laws are described as open systems, where explicit drains and sources are added to describe these deviations.

will adjust the amount of the quantity A in the two subsystems so that their Lagrange multipliers are identical. This implies for example, that two systems in thermal contact exchange heat (energy) until their temperatures are identical.

While the Lagrange multipliers appear at first as a purely mathematical byproduct, Lagrange multipliers are actually much more accessible to experiment than the original variable A:

The Lagrange multiplier can be measured by bringing a system into contact with a – very small – **test probe**, for which the Lagrange multipliers can be determined, e.g. a thermometer. Bringing the test probe into contact with the physical system, the Lagrange parameters will equilibrate and one can read the Lagrange multiplier from the test probe. A measurement of A, on the other hand, requires transferring the quantity A completely out of the system, and measuring the total amount of A removed.

Imagine a system coupled to a another – very large – system such that the conserved quantity A can be exchanged. Such a system is large **reservoir**. The reservoir is sufficiently large so that an exchange of A with the small system does not change the properties, namely α of the large system in any appreciable way. The reservoir is uniquely characterized by the Lagrange multiplier α.[21]

The typical example for a reservoir is a **heat bath**, which exchanges energy with a system such as to maintain a constant temperature. A whine cellar implements a heat bath, by maintaining a constant temperature of the whine. The heat bath is the surrounding soil, which maintains an approximately constant temperature. The temperature of the cellar does not rise significantly, if a warmer bottle is introduced into the cellar. Rather the bottle will adjust to the temperature of the cellar, as it is in thermal contact with the cellar.

1.6 From games to physical systems

So far we have considered games. The fundamental problem was that of determining the probabilities in a random experiment, and to include prior knowledge about the experiment into these predictions.

Now, we are changing our focus towards another "game", namely the one to predict properties of a physical system. The states of the physical system are the quantum mechanical states. We assume that we know the energy eigenstates $|\vec{n}\rangle$ of the system. The states are labeled by a set of quantum numbers $\vec{n}$, which we combined into a vector. Furthermore we need to know the expectation values of the energy, namely $E_{\vec{n}} = \langle\vec{n}|\hat{H}|\vec{n}\rangle$, and the expectation values[22] $X_{i,\vec{n}} = \langle\vec{n}|\hat{X}_i|\vec{n}\rangle$ of some other observables X_i.

With this knowledge, the goal of statistical physics is to predict the average values of the observables, or the change of the mean-values under certain physical conditions.

One may ask, how a purely cognitive concept, the entropy, can lead to reliable and almost certain predictions about a physical system. Why should a physical system care about our mind set? This question is related to an interesting discussion that still goes on. One direction of the argument leads to the H-theorem of Boltzmann described later.

When a system is coupled to an environment, transitions between the quantum mechanical states of the system lead to an effective randomization. This is an effect similar to the decay of a quantum mechanical wave packet. While an element of randomization is introduced, the transitions obey certain conservation laws, which avoid complete disorder. This introduces a new, no more subjective but objective, meaning of the probabilities. The probability describes the quantum mechanical probability averaged over time. In this objective interpretation, the probabilities are accessible to experiments.

We assume that the a-priori probabilities for the quantum-mechanical many-particle states are identical. Note that this assumption is one of the few basic assumptions in physics, which must be

---

[21] This is of course an oversimplification. In practice, a reservoir is characterized by its energy spectrum and coupling matrix elements. Those influence the time scales of the processes, but not the final thermodynamic states.

[22] We assume that the states are orthonormal and that they are eigenstates of the corresponding operators, that is: $\hat{H}|\vec{n}\rangle = |\vec{n}\rangle E_{\vec{n}}$ and $\hat{X}_i|\vec{n}\rangle = |\vec{n}\rangle X_{i,\vec{n}}$. We will see later that this requirement is not necessary. It is used here to simplify the discussion in the beginning.

confirmed by experiment. (It is.[23])

BASIC ASSUMPTION OF EQUILIBRIUM STATISTICAL MECHANICS

The *a-priori* probabilities of quantum mechanical many-particle states are identical.

That is, if we do not constrain the system every quantum mechanical state occurs with equal probability. We are free to make this choice in the subjective interpretation of the probabilities. However, the assumption can, and must be, tested for the objective interpretation, because, in principle, it can be falsified by comparison with frequencies. Surprisingly, the assumption of equal probabilities turns out to be correct.

1.6.1 Energy representation

Up to now, the entropy has been the fundamental variable. While the entropy is the fundamental variable of information theory, it is not a measurable quantity. For a physical description, the fundamental quantity is usually the energy, because we can make use of energy conservation. Energy conservation is one of the most fundamental elements of physical theory, as its roots lie in the fact that laws of nature do not change with time, a requirement for any predictive theory.

Therefore, we introduce here the energy representation, where energy-related variables play the dominant role. The entropy, on the other hand, is treated as a variable. In addition, the Lagrange multipliers are described by a combination of other variables, so that the variables of the theory match those, which developed historically. Note that the concepts of temperature and pressure were in use long before the formal theory of thermodynamics presented here was in place. Historically the formulation based on entropy, which was invented by Massieu in 1868 actually predated that of the energy representation introduced by Gibbs in 1875.

The change to the energy representation requires a redefinition of variables, while the information content of the theory is identical to what we learned so far. Let us assume a system, for which we know the mean value of the energy $E = A_0$ and that of other observables $X_j = A_j$. The Lagrange multiplier for the energy is related to the temperature T by $\alpha_0 = \frac{1}{T}$. The remaining Lagrange multipliers are written as $\alpha_i = \frac{f_i}{T}$.

$$\Big\{A_0, A_1, A_2, \ldots\Big\} \rightarrow \Big\{E, X_1, X_2, \ldots\Big\}$$
$$\Big\{\alpha_0, \alpha_1, \alpha_2, \ldots\Big\} \rightarrow \Big\{\frac{1}{T}, \frac{f_1}{T}, \frac{f_2}{T}, \ldots\Big\}$$

Note that the index in this expression does not refer to individual states of the system, but to different observables, such as energy, momentum etc.

The sign convention for the f_i is chosen such that $\bar{X}_i$ behaves like a coordinate and f_i behaves like a corresponding force[24]. We will later see that an unconstrained system will spontaneously adjust its parameters $\bar{X}_i$ in the direction of the corresponding force f_i. Because the spontaneous change is caused by the "attempt" of the system to maximize its entropy, the force points in the direction of increasing entropy.

In the following, we will need the definition of intensive and extensive variables

- **Extensive variable:** For a system combined out of N identical subsystems the value of an extensive quantity is N times the value of one of the subsystems. Extensive variables are the observables such as E or the X_i, respectively their expectation values $\bar{E}$ and $\bar{X}_i$.

---

[23] We mean that we need input from outside the present theory of probabilities. In this case it is one derived from the dynamics of quantum mechanical systems. Thus, it actually is again a theoretical input.

[24] This definition of the force may differ from what is called the thermodynamic force. As we will see later, the force used here is a direct generalization of that of a mechanical force, so that $f_i = -\frac{\partial \bar{E}}{\partial X_i}$.(Compare Eq. 1.73 with $U = \bar{E}$)

- **Intensive variable:** For a system combined out of N identical subsystems the value of an intensive quantity is identical to the value of one of the subsystems. Intensive variables are the Lagrange multipliers such as the temperatur or pressure.

The X_i can be any observable quantity. In principle we can use any quantum mechanical observable for X_i. In thermodynamics, only those observables are relevant that can be linked to a conservation law. [25] The most common pairs of extensive and intensive variables are summarized in table 1.1.

extensive	X	f	intensive
volume	V	p	pressure
strain	$\epsilon_{i,j}$	$-\sigma_{j,i}$	stress
particle number	N	$-\mu$	chemical potential
magnetization	$\vec{M}$	$-\vec{H}$	magnetic field
charge density	$\rho(\vec{r})$	$\Phi(\vec{r})$	electric potential
polarization	$\vec{P}$	$-\vec{E}$	electric field

Table 1.1: Most common pairs of extensive and intensive variables.

If we translate Eq. 1.34 we obtain the expectation values: The **internal energy** U is the mean value of the energy

INTERNAL ENERGY

$$U = \bar{E} = \sum_{\vec{n}} P_{\vec{n}} E_{\vec{n}} \tag{1.46}$$

and the mean values[26] of the other variables are

$$\bar{X}_i = \sum_{\vec{n}} P_{\vec{n}} X_{i,\vec{n}} \tag{1.47}$$

Because the temperature often occurs as $1/(k_B T)$ one defines a special symbol for it, namely

$$\beta \stackrel{\text{def}}{=} \frac{1}{k_B T} \tag{1.48}$$

We will try to avoid this notation, but we will come back to it, when it is necessary to simplify the notation. The Boltzmann constant k_B converts the temperature into an energy, so that $\alpha_0 A_0 = \frac{E}{k_B T}$ appearing in the exponent of the probability distribution is a dimension-less quantity. Its numerical value actually defines the temperature unit, the Kelvin, by relating it to an energy unit.

In order to get used to the above notation, let us again follow through the maximum-entropy principle.

---

[25] Observables that cannot be traced to a conservation law, are of less use, because the corresponding information would get lost over time due to transitions between the quantum mechanical states.

[26] We use the same symbol X_j for the quantity and its mean value. The mean value could be distinguished by writing $\langle X_j \rangle$, but we find that this would clutter the notation. The difference should become clear from the context. It is important to pay some attention to this point to keep track of the difference.

Maximum-Entropy principle

Let us go through the maximum-entropy principle again with the new notation. This will also give us an opportunity to generalize the results to several constraints instead of only one.

$$\begin{aligned}\mathcal{Y}(\{P_{\vec{n}}\}, U, T, \{\bar{X}_k\}, \{f_k\}) = -k_B \sum_{\vec{n}} P_{\vec{n}} \ln(P_{\vec{n}}) - \lambda \left(\sum_{\vec{n}} P_{\vec{n}} - 1 \right) \\ -\frac{1}{T} \left(\sum_{\vec{n}} P_{\vec{n}} E_{\vec{n}} - U \right) - \sum_k \frac{f_k}{T} \left(\sum_{\vec{n}} P_{\vec{n}} X_{k,\vec{n}} - \bar{X}_k \right) \qquad (1.49)\end{aligned}$$

The equilibrium condition is

$$0 = \frac{\partial \mathcal{Y}}{\partial P_{\vec{n}}} = -k_B \ln(P_{\vec{n}}) - k_B - \lambda - \frac{1}{T} E_{\vec{n}} - \sum_k \frac{f_k}{T} X_{k,\vec{n}} \qquad (1.50)$$

$$\Rightarrow P_{\vec{n}} = \mathrm{e}^{-(1+\frac{\lambda}{k})} \cdot \mathrm{e}^{-\frac{1}{k_B T}\left(E_{\vec{n}} + \sum_k f_k X_{k,\vec{n}}\right)} \qquad (1.51)$$

and the constraint conditions are

$$0 = \frac{\partial \mathcal{Y}}{\partial \lambda} = \sum_{\vec{n}} P_{\vec{n}} - 1 \qquad (1.52)$$

$$0 = \frac{\partial \mathcal{Y}}{\partial T} = \left[\left(\sum_{\vec{n}} P_{\vec{n}} E_{\vec{n}} - U \right) + \sum_k f_k \left(\sum_{\vec{n}} P_{\vec{n}} X_{k,\vec{n}} - \bar{X}_k \right) \right] \left(+\frac{1}{T^2} \right) \qquad (1.53)$$

$$0 = \frac{\partial \mathcal{Y}}{\partial f_k} = -\frac{1}{T} \left[\sum_{\vec{n}} P_{\vec{n}} X_{k,\vec{n}} - \bar{X}_k \right] \qquad (1.54)$$

Note that we obtain the same set of equations, whether we set the derivatives with respect to the original Lagrange multipliers $\alpha_0 = \frac{1}{T}$ and $\alpha_k = -\frac{f_k}{T}$ to zero or the derivatives with respect to T and f_k. While the equations, one-by-one, differ, the entire set contains the same information.[27]

Firstly, we enforce only the normalization condition Eq. 1.52 in addition to the equilibrium condition Eq. 1.50, which yields, analogously to Eq. 1.40, the equilibrium probabilities

EQUILIBRIUM PROBABILTIES

$$P^{eq}_{\vec{n}}(T, \{f_k\}) = \frac{1}{Z(T, \{f_k\})} \mathrm{e}^{-\frac{1}{k_B T}\left(E_{\vec{n}} + \sum_k f_k X_{k,\vec{n}}\right)} \qquad (1.55)$$

where the normalization constant, the so-called **partition function**, is defined as

PARTITION FUNCTION

$$Z(T, \{f_k\}) = \sum_{\vec{n}} \mathrm{e}^{-\frac{1}{k_B T}\left(E_{\vec{n}} + \sum_k f_k X_{k,\vec{n}}\right)} \qquad (1.56)$$

Note that the probabilities are now uniquely determined by the intensive variables T, f_k. We call probability distributions of Eq. 1.55 **equilibrium distributions**. They are entirely characterized by the intensive variables.

---

[27] A simple explanation is that if the function has a zero gradient, the variation of the function by small changes of the variables vanishes to first order irrespective of the direction of the change in the variable space. In a more formal way we make the same argument with the chain rule.

At this point it may also become clear, why the Lagrange multiplier for the energy, namely $\frac{1}{T}$ occurs in all other Lagrange multipliers $-f_k/T$ as well. Instead of treating energy and other variables separately, we could similarly define a new energy $E'_{\vec{n}}(\{f_k\}) = E_{\vec{n}} + \sum_k f_k X_{k,\vec{n}}$, which simply depends on the intensive variables. Thus, $X_{k,\vec{n}} = \partial E'_{\vec{n}}/\partial f_k$. In this case the partition sum has the simple form $Z = \sum_{\vec{n}} \mathrm{e}^{-\frac{1}{k_BT}E'_{\vec{n}}}$. Carrying over the temperature into the Lagrange multipliers helps to treat both choices consistently.

The remaining constraint conditions Eqs. 1.53,1.54 relate the intensive variables T, f_k to the extensive variables U, X_k.

$$\begin{aligned}
\bar{X}_k(T,\{f_k\}) &\overset{\text{Eq. 1.54}}{=} \sum_{\vec{n}} X_{k,\vec{n}} P^{eq}_{\vec{n}}(T,\{f_k\}) \overset{\text{Eq. 1.55}}{=} \frac{1}{Z(T,\{f_k\})} \sum_{\vec{n}} X_{k,\vec{n}}\, \mathrm{e}^{-\frac{1}{k_BT}\left(E_{\vec{n}}+\sum_k f_k X_{k,\vec{n}}\right)} \\
U(T,\{f_k\}) &\overset{\text{Eqs. 1.53,1.54}}{=} \sum_{\vec{n}} E_{k,\vec{n}} P^{eq}_{\vec{n}}(T,\{f_k\}) \overset{\text{Eq. 1.55}}{=} \frac{1}{Z(T,\{f_k\})} \sum_{\vec{n}} E_{\vec{n}}\, \mathrm{e}^{-\frac{1}{k_BT}\left(E_{\vec{n}}+\sum_k f_k X_{k,\vec{n}}\right)}
\end{aligned}$$

These are the microscopic versions of the equations of state.

MICROSCOPIC EQUATIONS OF STATE

$$U(T,\{f_k\}) = \frac{1}{Z(T,\{f_k\})} \sum_{\vec{n}} E_{\vec{n}}\, \mathrm{e}^{-\frac{1}{k_BT}\left(E_{\vec{n}}+\sum_k f_k X_{k,\vec{n}}\right)} \tag{1.57}$$

$$\bar{X}_k(T,\{f_k\}) = \frac{1}{Z(T,\{f_k\})} \sum_{\vec{n}} X_{k,\vec{n}}\, \mathrm{e}^{-\frac{1}{k_BT}\left(E_{\vec{n}}+\sum_k f_k X_{k,\vec{n}}\right)} \tag{1.58}$$

Maximum-Entropy principle for equilibrium states

Eq. 1.55 and is the link between statistical mechanics and thermodynamics. Whereas statistical physics is concerned with the individual states of the quantum mechanical system, thermodynamics is only concerned with the macroscopic properties of maximum-entropy states, which are embodied in the equations of state, that is the relation between intensive and extensive variables.

There are also **non-equilibrium distributions**, where the maximum entropy principle will not be sufficient: There are probability distributions that cannot be mapped into the form of Eq. 1.55. We cannot attribute intensive variables such as T and f_k to those. The intensive variables of non-equilibrium distributions are, strictly speaking, undefined. Nevertheless, we will learn later how these concepts can be carried over to non-equilibrium distributions, if they are only sufficiently close to equilibrium.

Since we know that probability distributions that maximize the entropy will be of the form Eq. 1.55, we can restrict our search for the maximum to such equilibrium distributions. If we restrict ourselves to equilibrium distributions, we can rewrite the entropy as a function of the intensive variables, which leads us to the **equilibrium entropy** defined as

$$Y(T,\{f_k\},U,\{\bar{X}_k\}) \overset{\text{def}}{=} \mathcal{Y}(P^{eq}_{\vec{n}}(T,\{f_j\}),T,\{f_k\},U,\{\bar{X}_k\}) \tag{1.59}$$

We can work out the expression for the equilibrium entropy as follows:

$$
\begin{aligned}
Y(T,\{f_k\},U,\{\bar{X}_k\}) &= -k_B\sum_{\vec{n}} P^{eq}_{\vec{n}}(T,\{f_k\})\ln(P^{eq}_{\vec{n}}(T,\{f_k\})) \\
&\quad -\frac{1}{T}\left(\sum_{\vec{n}} P^{eq}_{\vec{n}}(T,\{f_k\})E_{\vec{n}}-U\right)-\sum_k\frac{f_k}{T}\left(\sum_{\vec{n}} P^{eq}_{\vec{n}}(T,\{f_k\})X_{k,\vec{n}}-\bar{X}_k\right) \\
&\overset{Eq.\ 1.55}{=} -k_B\sum_{\vec{n}} P^{eq}_{\vec{n}}(T,\{f_k\})\left[-\ln[Z(T,\{f_k\})]-\frac{1}{k_BT}\left(E_{\vec{n}}-\sum_k f_kX_{k,\vec{n}}\right)\right] \\
&\quad -\frac{1}{T}\left(\sum_{\vec{n}} P^{eq}_{\vec{n}}(T,\{f_k\})E_{\vec{n}}-U\right)-\sum_k\frac{f_k}{T}\left(\sum_{\vec{n}} P^{eq}_{\vec{n}}(T,\{f_k\})X_{k,\vec{n}}-\bar{X}_k\right) \\
&= -\frac{1}{T}\left[-k_BT\ln[Z(T,\{f_k\})]-U-\sum_k f_k\bar{X}_k\right]
\end{aligned}
\tag{1.60}
$$

Note that the probabilities drop out of the equation, if we only resolve $\ln(P_n(T,f_k))$. Exploiting this fact is a mathematical trick that greatly simplifies the derivation.

Free energy

Note that all the information specific to the system is contained solely in the partition function. Therefore we define a new quantity, namely the free energy

FREE ENERGY

$$
F(T,\{f_k\})\overset{\text{def}}{=}-k_BT\ln[Z(T,\{f_k\})]
\tag{1.61}
$$

so that the partition function can be written as

$$
Z(T,\{f_i\})\overset{Eq.\ 1.61}{=}\mathrm{e}^{-\frac{1}{k_BT}F(T,\{f_i\})}
$$

and the probabilities can directly be written as

$$
P_{\vec{n}}(T,\{f_i\})=\mathrm{e}^{-\frac{1}{k_BT}\left(E_{\vec{n}}+\sum_{i=1}^n f_iX_{i,\vec{n}}-F(T,\{f_i\})\right)}
$$

We express the entropy by the free energy

$$
Y(T,\{f_k\},U,\{\bar{X}_k\})\overset{Eq.\ 1.60,1.61}{=}-\frac{1}{T}\left[F(T,\{f_k\})-U-\sum_k f_k\bar{X}_k\right]
\tag{1.62}
$$

Note that Y is equal to the entropy S, when the the constraints are fulfilled. This can be seen from the definition of $\mathcal{Y}$ in Eq. 1.49 and the definition of Y in Eq. 1.59.

Equations of state

Still we need to fulfill the constraint equations Eqs. 1.53,1.54 , which have lead us to the equations of state. Here we derive them without referring to the microscopic states, but only assume that the free energy is known.

Since the equilibrium probabilities have been obtained from the extremum condition $\frac{\partial\mathcal{Y}}{\partial P_{\vec{n}}}=0$, we can identify the constraint condition directly with the derivative of the equilibrium entropy with respect to the intensive variables.

$$
\frac{\partial Y}{\partial T}=\frac{\partial\mathcal{Y}}{\partial T}+\sum_{\vec{n}}\underbrace{\frac{\partial\mathcal{Y}}{\partial P_{\vec{n}}}}_{=0}\frac{dP_{\vec{n}}(T,\{f_k\})}{dT}=\frac{\partial\mathcal{Y}}{\partial T}
$$

Thus, the constraint conditions can be cast in the form

$$
\begin{aligned}
0 &= \frac{\partial Y(T,\{f_k\},U,\{\bar{X}_k\})}{\partial T} \\
0 &= \frac{\partial Y(T,\{f_k\},U,\{\bar{X}_k\})}{\partial f_i}
\end{aligned}
$$

Now we can evaluate the equations of state

$$
0 = \frac{\partial Y(T,\{f_k\},U,\{\bar{X}_k\})}{dT} \overset{\text{Eq. 1.62}}{=} \frac{1}{T^2}\underbrace{\left[F(T,\{f_k\})] - U - \sum_k f_k \bar{X}_k\right]}_{=-TY=-TS} - \frac{1}{T}\frac{\partial F}{\partial T} = -\frac{1}{T}\left[S + \frac{\partial F}{\partial T}\right]
$$

$$
\Rightarrow \qquad S(T,\{f_k\}) = -\frac{\partial F}{\partial T} \tag{1.63}
$$

$$
0 = \frac{\partial Y(T,\{f_k\},U,\{\bar{X}_k\})}{df_i} \overset{\text{Eq. 1.62}}{=} -\frac{1}{T}\left[\frac{\partial F(T,\{f_k\})}{\partial f_i} - \bar{X}_i\right]
$$

$$
\Rightarrow \qquad \bar{X}_i(T,\{f_k\}) = \frac{\partial F}{\partial f_i} \tag{1.64}
$$

Thus, we obtain three equations that together with the functional form of the free energy contain the complete thermodynamical information on a system.

EQUATIONS OF STATE

If the free energy is known as function of the intensive variables, we obtain entropy and other observables from the equations of state

$$
\begin{aligned}
S(T,\{f_k\}) &\overset{\text{Eq. 1.63}}{=} -\frac{\partial F(T,\{f_k\})}{\partial T} && (1.65) \\
\bar{X}_i(T,\{f_k\}) &\overset{\text{Eq. 1.64}}{=} \frac{\partial F(T,\{f_k\})}{\partial f_i} && (1.66) \\
U(T,\{f_k\}) &\overset{Eq.\ 1.62}{=} F(T,\{f_k\}) + TS(T,\{f_k\}) - \sum_i f_i \bar{X}_i(T,\{f_k\}) && (1.67)
\end{aligned}
$$

The last equation is equivalent with the maximum condition. The equations Eq. 1.65 and 1.66 are the **equations of state**. The equations of state impose a relation between the intensive and the extensive variables.

To be more specific we will use in the next chapter the volume V and the particle number N as extensive variable. The volume plays a special role as it cannot be attributed to a quantum mechanical observable. Nevertheless, we will find that it can be treated just like one of the X_k. The corresponding intensive variables are the negative pressure $f_k = p$ and the chemical potential $f_k = -\mu$.

Total differentials

Thermodynamic processes involve changes of the state variables, such as U, V, N. We can then define a thermodynamic process via the time dependent thermodynamic variables $(U(t), V(t), N(t))$. The corresponding intensive variables are then determined by the equations of state. We may be interested in how the entropy, which may be given as function of the extensive variables $S(U, V, N)$, changes during this process.

$$
\frac{dS}{dt} = \left.\frac{dS}{dU}\right|_{V,N}\frac{dU}{dt} + \left.\frac{dS}{dV}\right|_{U,N}\frac{dV}{dt} + \left.\frac{dS}{dN}\right|_{U,V}\frac{dN}{dt}
$$

The expression such as

$$\left.\frac{dS}{dN}\right|_{U,V} \stackrel{\text{def}}{=} \frac{\partial S(U,V,N)}{\partial N}$$

describes nothing but a partial derivative. It describes the infinitesimal change of S in a process where N is changed by a small amount while U and N are fixed. Therefore the mathematical symbols provide an immediate connection to some thermodynamic process.

What we did so-far is equilibrium thermodynamics, which should better be called "thermostatics". We can only make predictions about the system, once it reached thermal equilibrium. We cannot say anything about the dynamics of a system. Therefore it does not make sense to drag along the time coordinate, and we better write the above equation in the form

$$dS = \left.\frac{dS}{dU}\right|_{V,N} dU + \left.\frac{dS}{dV}\right|_{U,N} dV + \left.\frac{dS}{dN}\right|_{U,V} dN$$

We call this expression also the **total differential** of the entropy, implying that it describes all changes of the entropy during an infinitesimal process. Each of the three terms is a differential, but it is only a partial differential. This difference is analogous to the total and partial derivatives.

We might consider a differential, which has the form

$$df = \sum_i a_i(\vec{x})dx_i \tag{1.68}$$

describing the changes of a quantity f as the variables x_i are changed. We can convert this into a line integral $f_B = f_A + \int_{\vec{x}_A}^{\vec{x}_B} a_i dx_i$. If this integral is independent of the path from $\vec{x}_A$ to $\vec{x}_B$, we can define a function $f(\vec{x})$ such that $f_A = f(\vec{x}_A)$ and $f_B = f(\vec{x}_B)$. If this is the case, we call the differential Eq. 1.68 a **total differential**. The condition for a differential to be a total differential is

$$\frac{\partial a_i}{\partial x_j} = \frac{\partial a_j}{\partial x_i}$$

It can be shown via Stokes' theorem, that the path integral for any closed path vanishes, if the above condition is fulfilled. As a consequence, the integral is independent of the specific path taken between the end points. In three dimensions the condition requires the rotation to vanish.

For a total differential, the coefficients are then related to the gradient $a_i(\vec{x}) = \frac{\partial f}{\partial x_i}$ of the function f. Thus, the function $f(\vec{x})$ contains all necessary information to describe the respective processes. For thermodynamics these functions are the **thermodynamical potentials**, such as the entropy, the free energy or the internal energy.

As an example we will see that neither heat or work are total differentials.

1.6.2 Work and Heat

If we consider a process which changes T and f_i. We start out with the expression for $U(T, f_i)$ from Eq. 1.67 and determine the change of the internal energy

$$dU(T,\{f_k\}) = \frac{\partial U}{\partial T}dT + \sum_i \frac{\partial U}{\partial f_i}df_i \tag{1.69}$$

Let us work out the two derivatives using the equations of state Eqs. 1.65, 1.66, and Eq. 1.67.

$$\frac{\partial U}{\partial T} \stackrel{Eq.\ 1.67}{=} \underbrace{\frac{\partial F}{\partial T} + S}_{=0} + T\frac{\partial S}{\partial T} - \sum_k f_k \frac{\partial \bar{X}_k}{\partial T} \stackrel{Eq.\ 1.65}{=} T\frac{\partial S}{\partial T} - \sum_k f_k \frac{\partial \bar{X}_k}{\partial T} \tag{1.70}$$

$$\frac{\partial U}{\partial f_i} \stackrel{Eq.\ 1.67}{=} \frac{\partial F}{\partial f_i} + T\frac{\partial S}{\partial f_i} - \bar{X}_i - \sum_k f_k \frac{\partial \bar{X}_k}{\partial f_i} \stackrel{Eq.\ 1.66}{=} T\frac{\partial S}{\partial f_i} - \sum_k f_k \frac{\partial \bar{X}_k}{\partial f_i} \tag{1.71}$$

Insertion of the derivatives into Eq. 1.69 yields

$$\begin{aligned}
dU &= \frac{\partial U}{\partial T}dT+\sum_i \frac{\partial U}{\partial f_i}df_i \\
&\overset{\text{Eqs. 1.70,1.71}}{=} \left[T\frac{\partial S}{\partial T}-\sum_k f_k\frac{\partial \bar{X}_k}{\partial T}\right]dT+\sum_i\left[T\frac{\partial S}{\partial f_i}-\sum_k f_k\frac{\partial \bar{X}_k}{\partial f_i}\right]df_i \\
&= T\underbrace{\left[\frac{\partial S}{\partial T}dT+\sum_i\frac{\partial S}{\partial f_i}df_i\right]}_{dS}-\sum_k f_k\underbrace{\left[\frac{\partial \bar{X}_k}{\partial T}dT+\sum_i\frac{\partial \bar{X}_k}{\partial f_i}df_i\right]}_{d\bar{X}_k} \\
&= TdS-\sum_k f_k d\bar{X}_k
\end{aligned} \tag{1.72}$$

This is the so-called **fundamental relation of thermodynamics**

FUNDAMENTAL RELATION OF THERMODYNAMICS

$$dU \overset{\text{Eq. 1.72}}{=} \underbrace{TdS}_{\text{Heat}}\underbrace{-\sum_k f_k d\bar{X}_k}_{\text{Work}} = dQ+dW \tag{1.73}$$

HEAT AND WORK

The heat $\Delta Q_{A,B}$ transferred to a system as the system progresses from one state A to another one B is the increase of the internal energy caused by a change in entropy.

$$\Delta Q_{A,B}=\int_A^B TdS \tag{1.74}$$

The work $\Delta W_{A,B}$ transferred to a system as the system progresses from one state A to another one B is the increase of the internal energy due to a change of the extensive variables, but not caused by a change in entropy.

$$\Delta W_{A,B}=-\sum_k\int_A^B f_k d\bar{X}_k \tag{1.75}$$

The sign convention is as follows:

- $dQ > 0$ the system is heated up by the environment, i.e. energy flows from the environment into the system.
- $dQ < 0$ the system is releases heat into the environment, i.e. energy flows from the system into the environment.
- $dW > 0$ work is done on the system, i.e. energy flows from the system into the environment.
- $dW < 0$ the system is does work into the environment, i.e. energy flows from the system into the environment.

Note that both integrals depend on the path taken. Only the sum of heat and work is a total differential.
The term heat and work are related to energy transported from one system to another. Deviating from the common definitions I generalize these terms to allow for a conversion from work into heat. This is discussed below.

Here, I will discuss a matter that leads to a lot of confusion. It is based on an unfortunate definition of the term “heat”. Commonly heat is defined as energy transported during heat conduction. The energy stored in a body while heating it up to a higher temperature, is called **thermal energy**. One common mistake is to mix the terms “heat” and “thermal energy”.

The notions of heat and work developed historically to describe energy conservation, namely that difference of energy that flows in and out of the system is identical to the changes in the internal energy. Thus heat and work are commonly defined to describe forms of energy transfer.

In constrast to the common definition, we allow here also the notion of **heat generation**, which corresponds to a conversion of mechanical energy into thermal energy. Heat generation is always accompanied by **energy dissipation**. Heat generation and energy dissipation are two sides of the same medal: Because the energy is conserved, the amount of dissipated energy is equal to the heat generated. Energy is dissipated as a consequence of mechanical friction, electrical resistance, etc. During this process the energy is not lost, but it is transformed from a “useful” form into a “useless” form, namely thermal energy. The entropy of a body can change either due to heat transfer between the body and the environment or by heat generation, resulting from a spontaneous increase of entropy. Both terms are represented by TdS. In my notation the equation $dQ = TdS$ is used also for processes where heat is generated. If the term Q is reserved for heat exchange only, we must instead write $dQ \leq TdS$, which is often found in text books [28].

The fundamental relation also leads us to an alternative set of equations of state, namely

---

[28] See, for example, Eq.1.36 in the book of Chowdhury and Stauffer[5], Eq. 2.39 in the book by Schnakenberg[6].

EQUATIONS OF STATE:

If the internal energy is known as function of the extensive variables we obtain the corresponding intensive variables as functions of the extensive variables as:

$$f_j = -\frac{\partial U(S, \{\bar{X}_k\})}{\partial \bar{X}_j} \tag{1.76}$$

$$T = \frac{\partial U(S, \{\bar{X}_k\})}{\partial S} \tag{1.77}$$

Note that f_k plays the role of a force. For a classical system the force is given by $F_i = -\frac{dV}{dx_i}$, where $V(\vec{x})$ is the potential energy[29]. In analogy, we may call $\bar{X}_k$ a coordinate and $f_k = -\frac{\partial U}{\partial \bar{X}_k}$ the corresponding force.

1.6.3 Intensive and extensive variables

In thermodynamics we distinguish between intensive and extensive quantities.

An **extensive quantity** is one which scales with the size of the system. If we consider a combined system of subsystems, the value of an extensive quantity is the sum of the values of the subsystems. An extensive variable is directly related to an observable. $U, F, S, \bar{X}_i$ are extensive quantities,

The Lagrange multipliers used in the maximum-entropy principle are the **intensive quantities**. They do not scale with the size of the system system. If we combine two subsystems with different intensive variables, we can not assign an intensive variable to the global system. However, if we bring two subsystems into contact, the intensive parameters of the two subsystems become identical, so that we attribute this value to the combined system. Thus, T and $\{f_i\}$ are intensive quantities.

That there must be clear distinction between extensive and intensive variables can be seen from Eq. 1.67. If we consider two independent systems, the extensive variables scale up by a factor of two. Since Eq. 1.67 still holds for the combined system the temperature T and $\{f_i\}$ must remain constant.

1.7 Summary

In this section we have learned about the concepts of probability and entropy. Entropy corresponds to the information that can still be gained about a system. By maximizing the entropy, while incorporating all known information we can *obtain* a good choice for the probabilities. We have learned about Lagrange multipliers, thermodynamic potential, which depends only on the Lagrange multipliers, and, the partition function. Lagrange multipliers become equal, when systems are brought into contact so that they can exchange a globally conserved quantity.

This chapter has been formulated without specifying any system, in order to show the generality of the arguments.

A good reading are the lucid articles of Jaynes[**?**, **?**], which describe the connection between thermodynamics and information theory. The basis for information theory has been laid by Shannon[30][2].

---

[29] More generally we could have defined the force by the Hamilton function as $F_i = \dot{p}_i = -\frac{dH}{dx_i}$

[30] Claude Elwood Shannon. American Electrical engineer and Mathematician 1916-2001. Undergraduate studies at University of Michigan, Graduate studies and PhD at MIT. Worked at AT&T Bell Laboratories. Later professor at MIT. Founded the field of information theory. He was the first to use the word “bit”. Built the first chess game

1.7.1 Important formulas

$$
\begin{aligned}
\ln(N!) &\approx N\ln(N)-N\\
S[P_{\vec n}] &= -k_B\sum_{\vec n} P_{\vec n}\ln[P_{\vec n}]\\
U &= \sum_{\vec n} P_{\vec n}E_{\vec n}\\
\bar X &= \sum_{\vec n} P_{\vec n}\bar X_{\vec n}\\
Z(T,\{f_k\}) &= \sum_{\vec n} \mathrm{e}^{-\frac{1}{k_BT}\left(E_{\vec n}+\sum_k f_k\bar X_{k,\vec n}\right)} = \mathrm{e}^{-\frac{1}{k_BT}F(T,f_i)}\\
F(T,f_i) &= -k_BT\ln[Z(T,f_i)]\\
S &= -\frac{\partial F(T,f_i)}{\partial T}\\
\bar X_i &= \frac{\partial F(T,f_i)}{\partial f_i}\\
U &= F+TS-\sum_i f_i\bar X_i\\
dU &= TdS-\sum_i f_i d\bar X_i\\
T &= \frac{dU}{dS}\\
f_k &= -\frac{dU}{d\bar X_k}
\end{aligned}
$$

1.8 Exercises

Dice

Imagine a game of dice, where the person wins, that chooses the right value for the sum of figures shown on the dice. If nobody wins, the casino wins. Is there a safe winning strategy? For a set of two dice with states (i,j), work out the probabilities to obtain a given value for $(i+j)/2$. Repeat the same for three dice.

Solution: $P((i+j)/2) = \min[i+j-1; 13-i-j]/36$ There are $6\times 6 = 36$ states.

Lotto

Lotto is a game of luck, in which a participant selects six different numbers out of 48 numbers. Once all participants have made their selection, six numbers are randomly drawn. Depending on how many of the selected numbers coincide with the drawn numbers, an amount is paid back to the participant.

One selection (of a sextuple) costs 3.75 Euro. The following list determines the amount paid back as depending on the number of coinciding numbers

6	274,684.80 Euro	5	2,688.90 Euro	4	33.80 Euro	3	8.70 Euro

- Determine the real value of one ballot.
- Determine the probability that all six numbers coincide
- Assume that you invest 100 Euro per month for a period of 10 years to play lotto. What is the probability that you gain money

- Assume that you invest 100 Euro per month for a period of 10 years to play lotto. What is the probability that you draw once six correct numbers

Value of information

On average there are 60 rainy days in Los Angeles per year. The owner of a restaurant with an open garden may lower her losses by 1000 $, if she reduces her personell on a rainy day, where the garden cannot be used. Determine the value the information of the weather forecast for the owner of restaurant.

1. Determine the probability for rain in Los Angeles on the next day.
2. How much money can the restaurant owner make if he obtains the daily forecast as compared to reducing the personell at random.
3. Determine the possible states, and the probability distribution of the restaurant owner before and after the broadcast message.
4. Determine the entropy of the weather forecast.
5. Determine the value of the broadcast message for the restaurant.

Entropy of vacancies

Consider a crystal with M lattice sites on a periodic lattice. Not all of the lattice sites are occupied with atoms. Some atoms near the surface of the crystal may diffuse out of the interior of the crystal onto the surface. This creates a vacant lattice site, a so-called vacancy. An atom next to the vacancy can jump into the vacancy, leaving its former position vacant. Thus, the vacancy has moved. This is how vacancies migrate. Even though it costs energy for the atom to move to the surface, this is a possible process that does occur with a certain probability.

Vacancies often have dangling bonds, which appear in the forbidden band gap of an insulator or semiconductor and thus influence the optical, electrical and electrochemical properties. Since many dopants diffuse by a vacancy mechanism, vacancies determine the diffusion of dopant atoms in semiconductors. Vacancies can also hinder the migration of dislocations in metals by their strain field and thus affect the plastic behavior of materials. Thus, we want to know the concentration of vacancies at a given temperature.

In order to estimate the concentration we need to know the entropy of a state with a certain vacancy concentration. The goal is to estimate the entropy of a vacancy in the lattice. Assume that the crystal has M lattice sites, of which N sites are vacant.

1. determine the number of different arrangements of N vacancies in a crystal with M lattice sites. Tip: Start with an explicit calculation for $M = 1, 2, 3$ to become familiar with the problem, and then generalize the result to arbitrary M.
2. Determine the entropy $S(N, M)$ of this configuration.
3. Determine the entropy per vacancy in the limit $M \to \infty$ for a constant vacancy concentration $c = \frac{N}{M}$
4. Draw the functional dependence of of the entropy as function of the vacancy concentration and discuss the result.
5. Let the formation energy of a vacancy be E_F. The formation energy is the energy to move one atom to the surface of the crystal and thus to create a vacancy. Note that the surface only

shifts as atoms are deposited on it, so that more bulk atoms are created. The surface energy drops out of the equation. The probability for having a certain number of vacancies is given by

$$P(N)=\frac{\mathrm{e}^{-\frac{1}{k_BT}(E(N,M)-S(N,M))}}{\sum_N \mathrm{e}^{-\frac{1}{k_BT}(E(N,M)-S(N,M))}}$$

Determine the mean concentration of vacancies in the crystal.

Use $N!\stackrel{\text{def}}{=}1\cdot 2\cdots N$ and Stirling's formula $\ln[N!]\approx N\ln[N]-N$ for $N\rightarrow\infty$.
This exercise shall train the basic rules of combinatorics.

Chapter 2

The ideal Boltzmann gas

The ideal Boltzmann gas or simply the ideal gas is the prototypical system for thermodynamics. It describes a dilute system of non-interacting gas particles. This is the case for our atmosphere, and can also be used to describe the processes in heat engines, such as a Otto or Diesel motor, or refrigerators. We will see that the Boltzmann gas even gives a good description of the conduction electrons in a semiconductor, if the latter are sufficiently dilute.

In this section we will derive the thermodynamic functions such as free energy, internal energy and entropy, which we can use to illustrate the concepts of thermodynamic processes on a real system.

The first step in a thermodynamical treatment is an accurate description of the system and its possible states. Once we know the possible states of the system, we can consider the probabilities for the system being in a particular state.

We consider a rectangular container with side-lengths L_x, L_y, L_z, which is filled by N gas molecules. The results will not strongly depend on the shape of the box, which is why we have chosen a particularly simple form.

Fig. 2.1: Ludwig Boltzmann, 1844-1906

- The gas in the container is called **ideal**, if the particles do not interact with each other. In a real gas this is, of course, not true, because the gas molecules repel each other, when they collide, and they also attract each other, so that they become liquid or solid at low temperatures. However, as long as the system is sufficiently dilute, these interactions do not have a large influence on the energetics.

- A second approximation of the ideal Boltzmann gas is that we consider the gas molecules as point particles. Thus, we ignore the internal degrees of freedom such as rotations of the molecule, intramolecular vibrations or electronic excitations. We will see that these contributions can be treated separately.

It is at first surprising, how well real gases are approximated by the ideal Boltzmann gas.

We treat the gas here quantum mechanically. In the previous chapter, we introduced the principle of entropy based on discrete states of the system. In classical mechanics the states are continuous, so that it is not straight forward to assign a-priori probabilities in classical mechanics. In quantum mechanics the energy eigenstates are discrete, at least if the system is bound and thus confined to a finite volume. This makes the quantum mechanical description of statistical physics conceptually easier than the classical one. We will later see, how the rules can be transferred to classical systems by taking the classical limit $\hbar \to 0$.

2.1 States

2.1.1 Particle in a box

As the system consists of non-interacting particles, we first consider a single particle in the box. This will allow us later to work on the many-particle systems. For one particle we have to work on the "particle-in-a-box", a well known problem of quantum mechanics,

Each particle can be in one of the eigenstates of the Schrödinger equation

$$\left[-\frac{\hbar^2}{2m}\vec{\nabla}^2+V(\vec{r})\right]\phi(\vec{r})=\phi(\vec{r})\epsilon$$

where $V(\vec{r})=0$ inside the box and infinite elsewhere. The box is defined by $0<x<L_x$, $0<y<L_y$ and $0<z<L_z$.

The normalized wave functions are, inside the box,

$$\phi_{\vec{n}}(\vec{r})=\frac{2^{\frac{3}{2}}}{\sqrt{L_xL_yL_z}}\sin(k_xx)\sin(k_yy)\sin(k_zz) \tag{2.1}$$

where the components of the wave vector $\vec{k}_{i,j,k}=(\frac{\pi}{L_x}i,\frac{\pi}{L_y}j,\frac{\pi}{L_z}k)$ with $i,j,k=1,2,3,\dots$ are derived from the condition that the wave function vanishes at the boundaries of the box. The energy eigenstates of the one-particle Hamiltonian are

$$\epsilon=\int d^3r\,\phi^*(\vec{r})\left[-\frac{\hbar^2}{2m}\vec{\nabla}^2+V(\vec{r})\right]\phi(\vec{r})=\frac{(\hbar\vec{k})^2}{2m} \tag{2.2}$$

Thus,

$$\epsilon_{i,j,k}=\frac{(\hbar\vec{k}_{i,j,k})^2}{2m}=\frac{\hbar^2\pi^2}{2mL_x^2}i^2+\frac{\hbar^2\pi^2}{2mL_y^2}j^2+\frac{\hbar^2\pi^2}{2mL_z^2}k^2 \tag{2.3}$$

To simplify the discussion in the following, we label the eigenstates with a single index n. We could consider an ordering scheme of the following type

n	i	j	k
1	1	1	1
2	1	1	2
3	1	2	1
4	2	1	1
5	1	1	3
6	1	2	2
7	1	3	1
8	2	1	2
9	2	2	1
10	3	1	1

where we first consider all states with $i+j+k=3$ then the ones with $i+j+k=4$ and so forth.

2.1.2 From one-particle states to many-particle states

Product wave functions

Let us now consider a state of N independent particles in a box. Independent particles are those that do not interact with each other. The Hamiltonian for independent particles is the sum of the one-particle Hamiltonians. Thus, the N-particle Schrödinger equation has the form

$$\sum_{i=1}^{N}\left[\frac{-\hbar^2}{2m}\vec{\nabla}_i^2+V(\vec{r}_i)\right]\Psi(\vec{r}_1,\ldots,\vec{r}_N)=\Psi(\vec{r}_1,\ldots,\vec{r}_N)E$$

where $\vec{r}_i$ is the position of the i-th particle and $\vec{\nabla}_i$ acts on the coordinate of the i-t particle.

For independent particles, the eigenstates of the many-particle Hamilton operator can be represented as a product of one-particle states

$$\Psi_{n_1,\ldots,n_N}(\vec{r}_1,\ldots,\vec{r}_N)=\phi_{n_1}(\vec{r}_1)\phi_{n_2}(\vec{r}_2)\cdots\phi_{n_N}(\vec{r}_N)=\prod_{i=1}^{N}\phi_{n_i}(\vec{r}_i)$$

where the one-particle states are, in our case, given in Eq. 2.1. The corresponding many-particle energy $E_{\vec{n}}$ can be computed for this state simply as the sum of the corresponding one-particle energies

$$E_{\vec{n}}=E_{n_1,\ldots,n_N}=\sum_{i=1}^{N}\epsilon_{n_i}$$

where the one-particle energies are, for the particle in the box, given in Eq. 2.3.

It is important to note that the one-particle states have a large degeneracy. A many-particle wave function obtained by interchanging two one-particle states is again an eigenstate to the same energy. If all the particles would be distinguishable, this would indeed lead to a large degeneracy of the many-particle energies. However, if all the particles are identical, as required for the ideal gas, the degeneracy is largely reduced. This is important for evaluating the partition sum. We will discuss the role of identical particles in the following.

Indistinguishable particles

We required the particles to be identical, meaning indistinguishable. Thus, the Hamilton operator commutes with the exchange operator, which interchanges the arguments of two particles in the wave function (see ΦSX:Quantum Mechanics). The exchange operator $\hat{X}_{i,j}$ interchanges two particle coordinates as this is defined by

$$\hat{X}_{i,j}\Psi(\ldots,\vec{r}_i,\ldots,\vec{r}_j\ldots)=\Psi(\ldots,\vec{r}_j,\ldots,\vec{r}_i\ldots)$$

The eigenvalues of the exchange operator are $+1$ and -1[1]

The eigenstates of the Hamilton operator can be chosen as symmetrized (for bosons) or the antisymmetrized (for fermions) wave functions. The two possibilities correspond to the two eigenvalues of the exchange operator. A wave function is symmetrized with respect to exchange, if the interchange of the coordinates of two particles yields the same wave function. It is antisymmetric, if an interchange changes the sign of the wave function. For non-interacting particles we can construct the many-particle wave function from a product of one-particle wave functions $\phi_n(\vec{r})$. For the Boltzmann gas we obtain them from Eq. 2.1. The many-particle states have the form

$$\Psi_{n_1,\ldots,n_N}(\vec{r}_1,\ldots,\vec{r}_N)=\hat{P}_{\pm}\left[\phi_{n_1}(r_1)\phi_{n_2}(r_2)\ldots\phi_{n_N}(r_N)\right]$$

---

[1] If we exchange the same two particle twice, we obtain the original wave function. Thus, if x is some eigenvalue of X and $|\Psi_x\rangle$ is the corresponding eigenstate

$$\hat{X}_{i,j}|\Psi_x\rangle=|\Psi_x\rangle x$$

we obtain

$$\hat{X}^2_{i,j}|\Psi_x\rangle=|\Psi_x\rangle x^2\overset{\hat{X}_{i,j}=\hat{1}}{=}|\Psi_x\rangle\qquad\rightarrow\qquad x^2=1$$

where $\hat{P}_+$ is the symmetrization operator and $\hat{P}_-$ is the anti-symmetrization operator. $\hat{P}_\pm$ (anti)-symmetrizes the wave function and then normalizes the result. Let us illustrate the workings of the symmetrization operator for a two-particle system build up from orthonormal one-particle wave functions:

$$\hat{P}_\pm\left[\phi_i(\vec{r}_1)\phi_j(\vec{r}_2)\right] = \frac{1}{\sqrt{2}}\left(\phi_i(\vec{r}_1)\phi_j(\vec{r}_2) \pm \phi_j(\vec{r}_1)\phi_i(\vec{r}_2)\right)$$

The factor is required so that the resulting wave function remains normalized. The norm is the double integral $\int d^3r_1 \int d^3r_2 \Psi^*(\vec{r}_1,\vec{r}_2)\Psi(\vec{r}_1,\vec{r}_2)$.

The two eigenvalues of the exchange operator results in two fundamentally different classes of particles, namely Bosons and Fermions. Examples for Bosons are photons, phonons, Mesons. Electrons, protons, neutrons, quarks are fermions. Typically we attribute Fermions to real particles, while we attribute bosons to some interaction. (The photon is the particle responsible for the electromagnetic interaction between charged particles.) Fermions have all a half integer spin namely of $S_z = \frac{\hbar}{2}, \frac{3\hbar}{2}, \ldots$, while bosons have an integer spin such as $S_z = 0, \hbar, 2\hbar, \ldots$.

A consequence of the antisymmetry of Fermionic wave functions is the **Pauli principle**. Using the above example for the anti-symmetrized two-particle wave function shows that $\Psi(\vec{r},\vec{r}) = -\Psi(\vec{r},\vec{r}) = 0$.

Thus, the probability that two identical Fermions are in the same place vanishes. This is the first form of the Pauli principle, which also holds for interacting particles. It seems as if Fermions repel each other. Another consequence is that two identical fermions can not occupy the same one-particle state[2]: The wave function vanishes, if two one-particle quantum numbers are identical. Interchange of two one-particle quantum numbers in the product wave function is the same as the interchange of two particle coordinates and thus changes the sign of a wave function. Therefore the corresponding Fermionic wave function vanishes if two one-particle quantum numbers are identical. This is the reason why only two electrons, one for each spin, can be occupy a molecular orbital.

Avoid double counting in the partition sum

For evaluating the partition sum, it will be important that permuting the elements of the vector $\vec{n}$, which holds the quantum numbers[3], does not change the state for identical particles. The state $\Psi(\ldots, n_i, \ldots, n_j \ldots)$ is, up to an irrelevant phase factor, namely the sign, identical to $\Psi(\ldots, n_j, \ldots, n_i \ldots)$. Because the particles are identical, a permutation of the particle coordinates does not produce a new state. Thus, when we sum over all states, it will be necessary in the following to write the sum as

$$\sum_{\vec{n}} \approx \frac{1}{N!}\sum_{n_1}\cdots\sum_{n_N} \tag{2.4}$$

The factor $\frac{1}{N!}$ has been introduced[4] to avoid double counting of the same state. Because the quantum numbers $n_1, \ldots, n_N$ can be permuted $N!$ times[5] each state would be counted $N!$ times, if the factor $1/N!$ were not present.

The double-counting correction of Eq. 2.4 is not exact, because we have overcorrected those states with two or more equal quantum numbers. We ignore this problem for the moment. A few remarks may however be appropriate at this point. States with identical quantum numbers contribute appreciably to the partition sum only at sufficiently low temperatures, where all particles want to condense in the lowest one-particle state. At higher temperatures, states with identical quantum numbers have a negligible weight in the partition sum. States with identical quantum

---

[2] Note that this is statement is phrased in a colloquial way, since we cannot attribute an electron to one one-particle state.

[3] Remember that we have chosen $\vec{n}$, so that each component n_i uniquely identifies a one-particle state.

[4] The factorial is defined as $N! = 1 \cdot 2 \cdots N$

[5] It is a general statement that N elements of an ordered sequence can be permuted exactly $N!$ times, each state is counted $N!$ times. Because this result plays a central role in statistical physics, the reader is encouraged to convince himself of its validity.**Editor: this may be included in the appendix.**

numbers must also be treated differently for bosons and fermions: For Fermions those states must be excluded from the sum. For Bosons their contribution in the partition sum should be larger than in Eq. 2.4 , because identical indices only occur once in the multiple sum[6]. As a consequence, the low-temperature behavior of fermions and bosons differ radically. We will later develop a technique that will allow us to also treat also the low-temperature limit, the degenerate quantum gas, correctly. This will lead us the Bose[7] and the Fermi-statistics[8]. Note here, that the range of temperatures for which this high-temperature limit is applicable depends on the system. For electrons, this may require temperatures similar to those in the sun.

The factor $\frac{1}{N!}$ in the partition sum has important implications. If we consider a gas of several components, such as N_2, O_2, $CO_2, \ldots$ in our atmosphere, the treatment here only applies to a single component. The results for each component can then be combined in the the result of the more-component gas. Treating all components as identical would result in erroneous results!

2.2 Partition function of the ideal Boltzmann gas

2.2.1 Partition function for a fixed number N of particles

Now we can apply the formalism of the previous section. We always start with evaluating the partition function. All thermodynamic properties of the system can be evaluated from the partition function.

Before we evaluate the partition function we need to specify the **ensemble**

ENSEMBLE

An ensemble is a set of states with corresponding probabilities.
There are three types of named ensembles:

- the **micro-canonic ensemble** describes a closed system that can neither exchange energy nor particles with the environment. It is characterized by a fixed value for the energy E and the particle number N.
- The **canonical ensemble** describes a system that exchanges heat with the environment, but has a fixed particle number. It is characterized by a Temperature T and a fixed value for the particle number N.
- The **grand-canonical ensemble** describes a system that exchanges heat and particles with the environment. It is characterized by a temperature T and a chemical potential μ.

We specify the ensemble by deciding on the constraints that reflect the state of our knowledge. We start from a system which should have a defined average energy $U = \sum_{\vec{n}} P_{\vec{n}} E_{\vec{n}}$, a fixed number[9] of particles N and a defined volume $V = L_x L_y L_z$. (We will see later that the results will not depend on the shape of the box, as long as it is sufficiently large in each direction.) The average energy U is the internal energy.

---

[6] Consider for example a state where all quantum numbers are identical. It is only considered once in the multiple sum and needs not be corrected by the factor $1/N!$. Thus in order to correct Eq. 2.4 we would need to multiply its weight again with $N!$

[7] Satyendranath Bose. Indian Physicist 1894-1974. Invented Bose Einstein Statistics. Dirac introduced the name Boson in his honor. Professor of Physics at University of Dacca and later at Calcutta University.

[8] Enrico Fermi. Italian Physicist 1901-54. Nobel price in Physics 1938. Professor of Theoretical Physics at University of Rome 1927-1938. Emigrated to the USA in 1938 immediately after receiving the Nobel price to escape Mussolini's Dictatorship. Professor at Columbia University, USA 1939-1942. One of the leaders of the Manhattan project to develop nuclear energy and the atomic bomb. Professor at University of Chicago 1946-1954.

[9] Fixing the number of particles is different from fixing its average number. If we fix the average, there is a finite probability that the system has a particle number that differs form the average. That is, the particle number may fluctuate, which is excluded if we fix the number.

For independent particles, i.e. $E_{\vec{n}} = \sum_i \epsilon_{n_i}$, the partition function, Eq. 1.56, can be simplified[10]

$$
\begin{aligned}
Z(T,V,N) &= \sum_{\vec{n}} \mathrm{e}^{-\frac{1}{k_BT}E_{\vec{n}}} = \frac{1}{N!}\sum_{n_1}\sum_{n_2}\cdots\sum_{n_N}\mathrm{e}^{-\frac{1}{k_BT}\sum_i \epsilon_{n_i}} \\
&= \frac{1}{N!}\left(\sum_{n_1}\mathrm{e}^{-\frac{1}{k_BT}\epsilon_{n_1}}\right)\cdots\left(\sum_{n_N}\mathrm{e}^{-\frac{1}{k_BT}\epsilon_{n_N}}\right) \\
&= \frac{1}{N!}\left(\sum_{n}\mathrm{e}^{-\frac{1}{k_BT}\epsilon_{n}}\right)^N
\end{aligned}
\tag{2.5}
$$

Now we need to evaluate the sum

$$
\sum_n \mathrm{e}^{-\frac{1}{k_BT}\epsilon_n} \overset{Eq.\ 2.3}{=} \sum_{i,j,k=1}^{\infty}\mathrm{e}^{-\frac{\hbar^2\vec{k}^2_{i,j,k}}{2mk_BT}} = \sum_{i,j,k=1}^{\infty}\mathrm{e}^{-\frac{\hbar^2\pi^2}{2mk_BTL_x^2}i^2}\,\mathrm{e}^{-\frac{\hbar^2\pi^2}{2mk_BTL_y^2}j^2}\,\mathrm{e}^{-\frac{\hbar^2\pi^2}{2mk_BTL_z^2}k^2}
$$

Unfortunately this sum cannot be evaluated analytically. Therefore we resort to an approximation: If the pre factor $\frac{\hbar^2\pi^2}{2mk_BTL^2}$ is sufficiently small, we can convert the sum into an integral. This condition is met if the box is either sufficiently large and/or the temperature is sufficiently high. In this case, the individual terms in the sequence change only by small amounts[11]

$$
\sum_i \mathrm{e}^{-\frac{\hbar^2\pi^2}{2mk_BTL_i^2}i^2} \approx \int_0^\infty dx\ \mathrm{e}^{-\frac{\hbar^2\pi^2}{2mk_BTL_i^2}x^2} = \sqrt{\frac{2mk_BTL_i^2}{\hbar^2\pi^2}}\cdot\underbrace{\frac{1}{2}\int_{-\infty}^{\infty}dy\ \mathrm{e}^{-y^2}}_{=\sqrt{\pi}/2} = \sqrt{\frac{mk_BTL_i^2}{2\pi\hbar^2}}
$$

By combining the sum for the x,y, and z-directions, we obtain

$$
\sum_n \mathrm{e}^{-\frac{1}{k_BT}\epsilon_n} \approx \left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}} V
$$

where $V = L_xL_yL_z$ is the volume of the box.

By inserting this result into the expression Eq. 2.5 for the partition function, we obtain the partition function of the ideal Boltzmann gas

$$
Z^{BG}(T,V,N) = \frac{1}{N!}\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3N}{2}} V^N
\tag{2.6}
$$

---

[10] We write the partition function of volume instead of using the proper shape of the box, namely L_x, L_y and L_z. The reason is that the side-lengths of the box only enter in the form of the product, that is the volume.

[11] An error estimate can be obtained by approximating the integrand by step like functions that is either always larger than the integrand or always smaller than the integrand. The integral of the step functions can be exactly converted into a sum.

$$
\int_0^\infty dx\ \mathrm{e}^{-\alpha x^2} < \sum_{i=0}^{\infty}\mathrm{e}^{-\alpha i^2} = 1 + \sum_{i=1}^{\infty}\mathrm{e}^{-\alpha i^2} \qquad \text{and}
$$

$$
\int_0^\infty dx\ \mathrm{e}^{-\alpha x^2} > \sum_{i=0}^{\infty}\mathrm{e}^{-\alpha(i-1)^2} = \sum_{i=1}^{\infty}\mathrm{e}^{-\alpha i^2}
$$

$$
\Rightarrow \sqrt{\frac{\pi}{4\alpha}} - 1 = \left(\int_0^\infty dx\ \mathrm{e}^{-\alpha x^2}\right) - 1 < \sum_{i=1}^{\infty}\mathrm{e}^{-\alpha i^2} < \int_0^\infty dx\ \mathrm{e}^{-\alpha x^2} = \sqrt{\frac{\pi}{4\alpha}}
$$

If α tends to zero, the relative error vanishes.

Definition 2.1 *THERMAL DE BROGLIE WAVE-LENGTH*
The **thermal de-Broglie wavelength**[a] *is defined as*

$$\lambda_T \overset{def}{=} \sqrt{\frac{2\pi\hbar^2}{mk_BT}} \tag{2.7}$$

The de-Broglie wavelength is a length[b] *and it is temperature dependent.*

---

[a] Prince Louis-Victor De Broglie, 1892-1987 French theoretical physicist. Nobel price in Physics, 1929 for his discovery of the wave nature of electrons. De Broglie is the originator of the wave-particle dualism,

[b] $\hbar$ has the units of an angular momentum, and k_BT is an energy. Thus the units of λ_T are

$$\sqrt{\frac{(kg\frac{m^2}{s})^2}{kg\cdot(kg\frac{m^2}{s^2})}} = m$$

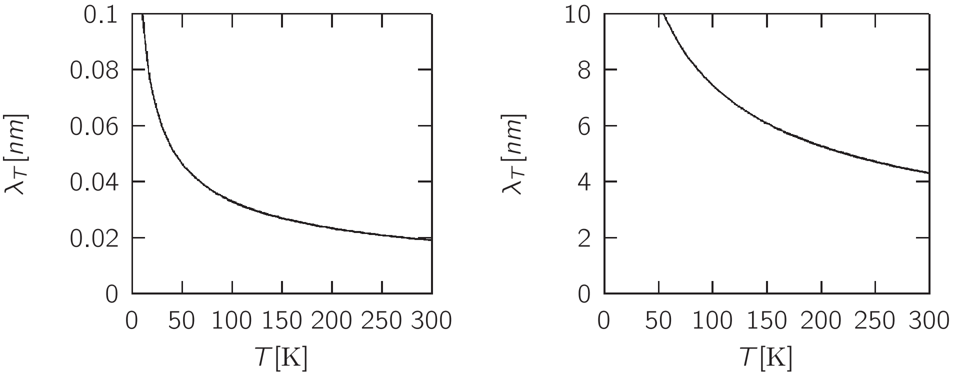


Fig. 2.2: Thermal de-Broglie wavelength as function of temperature. Left for nitrogen molecules, the main component of our atmosphere. Right for electrons. For nitrogen, the thermal de-Broglie wavelength becomes appreciable, that is proportional to intermolecular distances, only at very low temperatures. For electrons the thermal de-Broglie wavelength is substantially larger than atomic distances ($\approx$ 0.3 nm), while it is still smaller than the average distance of conduction electrons in a typically doped semiconductor. (Note that, in a semiconductor the electron mass must be replaced by its effective mass.)

The de-Broglie wavelength allows us to cast the requirements, that allowed us to convert the sum into an integral in more elegant form: The box dimensions must be much larger than the de-Broglie wavelength, that is

$$L_x >> \lambda_T \qquad L_y >> \lambda_T \qquad L_z >> \lambda_T$$

This assumption defines the ideal Boltzmann gas. At sufficiently low temperatures or at high density, even the ideal gas will exhibit phenomena, such as the Bose-Einstein condensation, which cannot be described in the present model.

2.2.2 Partition function at constant chemical potential

In the previous derivation we have used volume and particle number as parameters. Imagine now that we do not know the number N of particles exactly, but we still know the mean value of the number $\bar{N}$ of particles. Here, we will to derive the thermodynamic functions for this case.

We obtain a partition function that does not depend on the particle number, but on its Lagrange parameter, namely the **chemical potential** μ. We consider all quantum states, not only for a fixed number of particle numbers, but for all possible particle numbers. The quantum mechanical states, which cover also all particle numbers span the so-called **Fock space**.

We start constructing the partition function for two expectation values, namely the one for energy and the one for the particle number.

$$P_{\vec{n}}(T,V,\mu)=\frac{1}{Z(T,V,\mu)}\mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}-\mu N_{\vec{n}})}$$

The intensive variable conjugate to the particle number is called chemical potential μ.

We evaluate the partition function, which, now, does not depend on the particle number but on the chemical potential. In the last step we will use the Taylor-expansion of the exponential function $\mathrm{e}^x=\sum_{n=0}^{\infty}\frac{1}{n!}x^n$.

$$\begin{aligned}
Z(T,V,\mu) &= \sum_{\vec{n}}\mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}-\mu N_{\vec{n}})}=\sum_{N=0}^{\infty}\left(\sum_{\substack{\vec{n}\\ N_{\vec{n}}=N}}\mathrm{e}^{-\frac{1}{k_BT}E_{\vec{n}}}\right)\mathrm{e}^{\frac{1}{k_BT}\mu N}\\
&\overset{\text{Eq. 2.5}}{=} \sum_{N=0}^{\infty}Z^{BG}(T,V,N)\mathrm{e}^{\frac{1}{k_BT}\mu N}\overset{Eq.\ 2.6}{=}\sum_{N=0}^{\infty}\frac{1}{N!}\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3N}{2}}V^N\mathrm{e}^{\frac{1}{k_BT}\mu N}\\
&= \sum_{N=0}^{\infty}\frac{1}{N!}\left[\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}V\mathrm{e}^{\frac{1}{k_BT}\mu}\right]^N\\
&\overset{\mathrm{e}^x=\sum_{n=0}^{\infty}\frac{1}{n!}x^n}{=} \exp\left[\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}V\mathrm{e}^{\frac{1}{k_BT}\mu}\right]
\end{aligned}\tag{2.8}$$

2.3 Grand-canonical potential

The corresponding free energy for a variable particle number is called the **grand-canonical potential**. It is defined analogously to Eq. 1.61 as

Definition 2.2 *GRAND-CANONICAL POTENTIAL*

$$\Phi(T,V,\mu)=-k_BT\ln[Z(T,V,\mu)]\tag{2.9}$$

The partition function can be expressed by the grand-canonical potential as

$$Z(T,V,\mu)\overset{\text{Eq. 2.9}}{=}\mathrm{e}^{-\frac{1}{k_BT}\Phi(T,V,\mu)}$$

and the probabilities can be written as

$$P_{\vec{n}}(T,V,\mu)=\mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}-\mu N_{\vec{n}}-\Phi(T,V,\mu))}$$

For the Boltzmann gas, we obtain

$$\begin{aligned}
\Phi^{BG}(T,V,\mu) &\overset{Eq.\ 2.9}{=} -k_BT\ln[Z^{BG}(T,V,\mu)]\\
&\overset{Eq.\ 2.8}{=} -k_BT\ln\left\{\exp\left[\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}V\mathrm{e}^{\frac{1}{k_BT}\mu}\right]\right\}
\end{aligned}\tag{2.10}$$

From this, we obtain the final form of the grand-canonical potential of the ideal Boltzmann gas as

$$\boxed{\Phi^{BG}(T,V,\mu)=-k_BT\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}V\mathrm{e}^{\frac{\mu}{k_BT}}} \tag{2.11}$$

2.4 Equations of state

2.4.1 From the grand canonical potential to the entropy

Now we can work out the equations of state for the Boltzmann gas. The equations of state relate the intensive variables to the extensive variables.

In this special case it is easier to express the quantities N, S and U by the **grand canonical potential** $\Phi^{BG}(T,V,\mu)$ and to eliminate the grand-canonical potential in a second step.

The derivation of the following equations Eqs. 2.12, 2.13 and 2.14 can be simplified by a mathematical trick: If we need to form the derivative of products with many terms, it is often efficient to rewrite the product rule in the following way

$$\partial_x\Big(f(x)g(x)\Big)=\left(\frac{\partial_x f(x)}{f(x)}+\frac{\partial_x g(x)}{g(x)}\right)\Big(f(x)g(x)\Big)=\Big(\partial_x\ln[f(x)]+\partial_x\ln(g(x))\Big)\Big(f(x)g(x)\Big)$$

For long products, it avoids a lot of typing. It is useful to memorize the derivatives of the logarithm of a few elementary functions:

$$\partial_x\ln(x^n)=\frac{n}{x}$$
$$\partial_x\ln\Big(\mathrm{e}^{f(x)}\Big)=\partial_x f(x)$$

We start from Eq. 1.66, $\bar{X}_i=\frac{\partial F}{\partial f_i}$, with $f_i=\mu$ for $\bar{X}_i=N$ or with $f_i=T$ for $X_i=S$, as specified in table 1.1.

$$N^{BG}(T,V,\mu)\overset{Eq.\ 1.66}{=}-\frac{d\Phi^{BG}}{d\mu}\overset{\text{Eq. }2.11}{=}-\frac{\partial}{\partial\mu}\left[-k_BT\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}V\mathrm{e}^{\frac{1}{k_BT}\mu}\right]$$
$$=-\frac{1}{k_BT}\Phi^{BG} \tag{2.12}$$
$$S^{BG}(T,V,\mu)\overset{Eq.\ 1.65}{=}-\frac{d\Phi^{BG}}{dT}\overset{\text{Eq. }2.11}{=}-\frac{\partial}{\partial T}\left[-k_BT\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}V\mathrm{e}^{\frac{1}{k_BT}\mu}\right]$$
$$=-\left[\frac{5}{2T}-\frac{\mu}{k_BT^2}\right]\Phi^{BG} \tag{2.13}$$
$$U^{BG}(T,V,\mu)\overset{Eq.\ 1.67}{=}\Phi^{BG}+TS+\mu N$$
$$\overset{\text{Eqs. }2.13,2.12}{=}\left[1-\frac{5}{2}+\frac{\mu}{k_BT}-\frac{\mu}{k_BT}\right]\Phi^{BG}=-\frac{3}{2}\Phi^{BG} \tag{2.14}$$

We resolve the last equation for Φ^{BG} and obtain

$$\Phi^{BG}\overset{Eq.\ 2.14}{=}-\frac{2}{3}U \tag{2.15}$$

Now we insert this result into the expression Eq. 2.12 for the particle number and obtain

$$N\overset{Eqs.\ 2.12,2.15}{=}\frac{2U}{3k_BT} \tag{2.16}$$

from which we obtain the **caloric equation of state**:

CALORIC EQUATION OF STATE FOR THE IDEAL GAS

.

$$U \overset{\text{Eq. 2.16}}{=} \frac{3}{2} N k_B T \tag{2.17}$$

As we will see later, this result reflects that the average kinetic energy for a classical particle is $\frac{3}{2} k_B T$. For the ideal Boltzman gas, the kinetic energy is the only contribution to the internal energy U.

Now we take the expression Eq. 2.11 for Φ^{BG} and resolve it for μ

$$\begin{aligned} \mu &\overset{Eq.\ 2.11}{=} k_B T \ln\left[-\frac{\Phi^{BG}}{k_B T V} \left(\frac{2\pi\hbar^2}{m k_B T} \right)^{\frac{3}{2}} \right] \\ &\overset{\text{Eq. 2.12}}{=} k_B T \ln\left[\frac{N}{V} \left(\frac{2\pi\hbar^2}{m k_B T} \right)^{\frac{3}{2}} \right] \end{aligned} \tag{2.18}$$

Finally, we insert this expression Eq. 2.18 into the entropy Eq. 2.13

$$\begin{aligned} S^{BG} &\overset{\text{Eq. 2.13}}{=} -\left(\frac{5}{2} - \frac{\mu}{k_B T} \right) \frac{1}{T} \Phi^{BG} \\ &\overset{Eqs.\ 2.18,2.15}{=} \left(\frac{5}{2} - \ln\left[\frac{N}{V} \left(\frac{2\pi\hbar^2}{m k_B T} \right)^{\frac{3}{2}} \right] \right) \frac{2U}{3T} \\ &\overset{\text{Eq. 2.17}}{=} N k_B \left(\frac{5}{2} - \ln\left[\frac{N}{V} \left(\frac{3\pi\hbar^2}{m} \frac{N}{U} \right)^{\frac{3}{2}} \right] \right) \end{aligned}$$

From this we obtain the so-called **Sackur-Tetrode equation**[12].[13]

SACKUR-TETRODE EQUATION

$$\Rightarrow S^{BG}(U, V, N) = N k_B \left(\frac{5}{2} - \ln\left[\left(\frac{3\pi\hbar^2}{m} \right)^{\frac{3}{2}} \frac{N^{\frac{5}{2}}}{V U^{\frac{3}{2}}} \right] \right) . \tag{2.19}$$

The Sackur-Tetrode equation contains all information we will need to describe all the thermodynamic properties of the Boltzmann gas.

2.4.2 Internal energy and pressure

The Sackur-Tetrode equation also yields the internal energy as function of entropy, volume and number of particles.

---

[12] The Sackur-Tetrode equation is named for Hugo Martin Tetrode (1895-1931) and Otto Sackur (1880-1914) who independently derived the equation in 1912.

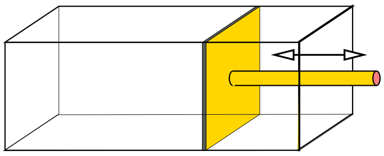
[13] The Sackur-Tetrode equation can also be derived directly from $Z(T, V, N)$ without introducing the the particle number N as known mean value. However, in that case one needs to employ Stirling's approximation. This tells two lessons: (1) using a fixed value of some quantity or its mean value makes a difference, and (2) for large N, where Stirling's formula becomes exact, the two results become identical. The limit of large N is also called the **thermodynamic limit**.

INTERNAL ENERGY OF THE IDEAL GAS

$$U^{BG}(S,V,N) \overset{Eq.\ 2.19}{=} \frac{3\pi\hbar^2}{m} N^{\frac{5}{3}} V^{-\frac{2}{3}} e^{\frac{2S}{3Nk_B}-\frac{5}{3}} \tag{2.20}$$

Like the Sackur-Tetrode equation also the internal energy expressed as function of S, V, N contains the entire thermodynamic information of the ideal gas.

Now we can evaluate the pressure. Earlier, in table 1.1, we have used the term **"pressure"** as the Lagrange multiplier related to the volume. It is not yet clear whether this pressure is identical to the quantity "pressure" used in mechanics. The pressure as defined in classical mechanics is force divided by area. Here we will show that the two are identical and that the definition used here is a generalization of the pressure from purely mechanic systems to thermodynamic systems.



Let us consider a thermally insulated container with one side-wall replaced by a movable piston. The container shall be filled with a gas. The piston allows us to change the volume of the container. When we compress the gas we have to perform work which is defined as $dW = Fds$. F is the force with which we push against the piston and ds is the distance by which the piston is moved. We want to move the piston only very slowly, which implies that the force we apply to the system is balanced by the force exerted by the gas on the piston. If the area of the piston is A, the work can be expressed by the volume

$$dW = Fds = \underbrace{\frac{F}{A}}_{p} \cdot \underbrace{Ads}_{-dV} = -pdV$$

because the volume changes by $dV = -Ads$. Because of overall energy conservation, the work done on the system is stored as internal energy of the system, that is $dU = dW$. Hence, if there are no other processes that add or remove energy from the system, we obtain the energy change as

$$dU = -pdV$$

We just required that there no other mechanisms that add or remove energy from the system. Thus we need to thermaly insulate the system so that no energy is lost from the system in the form of heat. Since the heat transfer to the environment $dQ = TdS$ vanishes, the entropy is conserved during this process. Thus, the **pressure** is defined more precisely as the derivative of the internal energy *at constant entropy*.

Note that this relates the pressure known from mechanics to a derivative of a thermodynamic potential, namely the internal energy. This derivative, on the other hand, is nothing but the definition of the thermodynamic pressure obtained from Eq. 1.76 on 43 with $f_i = -p$. Thus we have shown the te Lagrange multiplier named pressure is identical to the mechanical pressure.

Hence the pressure is defined as follows:

Definition 2.3 *PRESSURE*

$$p \overset{def}{=} -\frac{\partial U(S,V,N)}{\partial V} \tag{2.21}$$

The definition of the pressure is an equation of state. Among the equations of state, it plays a special role, because we cannot express the volume as an expectation value of an observable. (At least, I can not.)

If we evaluate the pressure of the Boltzmann gas,

$$p^{BG} \overset{Eqs.\ 2.21,2.20}{=} \frac{2}{3V}U \overset{Eq.\ 2.17}{=} \frac{Nk_BT}{V} \tag{2.22}$$

we obtain the third equation of state, namely the **ideal gas law**.

IDEAL GAS LAW

$$pV \overset{\text{Eq. 2.22}}{=} Nk_BT \tag{2.23}$$

2.4.3 Discussion of the Equations of state

Ideal gas law

The ideal gas law Eq. 2.23 is often written as $pV = nRT$. In this case the number of gas particles is counted in moles, so that $N = N_An$. One **mole (mol)** of particles contains $N_A = 6.0221367 \times 10^{23}$ mol^{-1} particles, where the proportionality factor N_A is the **Avogadro's constant**. The mol is the SI unit for amount of substance.[14] The **gas constant** $R = 8.31451$ J/(kmol) is defined as $R = N_Ak_B$.

The SI unit for the pressure is the **Pascal (Pa)**; 1 Pa=1 N/m^2=1 kg/(ms^2). Other units are in common use:

- The **Torr (torr)** is frequently used in meteorology and to measure the quality of a vacuum. The Torr is about the pressure which a 1mm high column of Mercury (Hg) exerts on the base under the influence of the gravitational field. A pressure below 10^{-8} Torr is called ultra-high vacuum (UHV).
- The **unit atmosphere (atm)** is defined by the relation 1 atm=760 Torr. One atmosphere is approximately ambient pressure.
- 1 **bar (bar)** is exactly 1 bar=10^5 Pa. A pressure of 1 bar corresponds to approximately ambient pressure.

1 bar=10^5 Pa	1 Torr≈133.322 Pa	1 atm=760Torr≈101325 Pa

The SI unit[**?**] for the temperature is the **Kelvin (K)**. The zero of the Kelvin temperature scale is defined as that temperature where the probabilities for all states but the ground state vanishes. The scale is defined such that 273.16 K is the temperature of the triple point of water. At the triple point of water, the solid phase, the liquid phase and the gas phase of water coexist. This requirement defines a unique point in the p-T phase diagram. The Boltzmann unit converts the temperature scale into an energy scale. Other units in common use are:

- **Celsius (°C)**[**?**] is another temperature scale. The conversion is defined by

  $$\frac{T}{^\circ\mathrm{C}} = \frac{T}{\mathrm{K}} - 273.15(\text{defined})$$

  The number is a definition and is not rounded. 0 C is approximately the freezing point of water at ambient pressure and 100 C is approximately the boiling point of water at ambient pressure.

---

[14] The reader may find such unit superfluous. This may also be true. However, without this constants the chemists would always deal with numbers in the 10^{23} range, which is very inconvenient.

- The temperature unit **Fahrenheit (F)**[?] is still in use in the United States. The conversion to the Celsius temperature scale is

$$\frac{T}{\mathrm{F}}=\frac{9}{5}\frac{T}{{}^\circ\mathrm{C}}+32$$

The equations of state are best represented graphically by keeping all but two of the independent variables fixed. This leads us to the following graphs:

- **Isotherm**: An isotherm is a line of constant temperature. It is represented in a diagram of pressure versus volume as shown in Fig. 2.3.

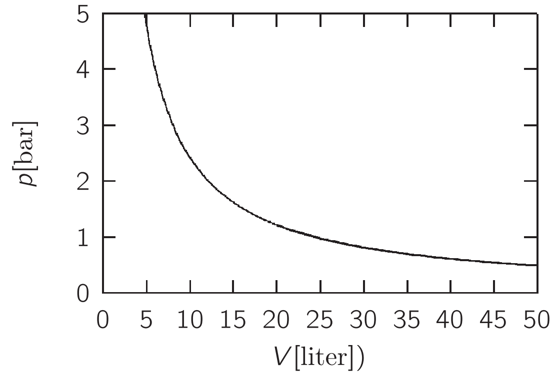


Fig. 2.3: Isotherm of a gas with 1 mol of particles at room temperature (293 K). 1 mol of N_2 weighs about 55 g.

- **Isobar**[15]: An isobar is a line of constant pressure. It is represented in a diagram of volume versus temperature as shown in Fig. 2.4

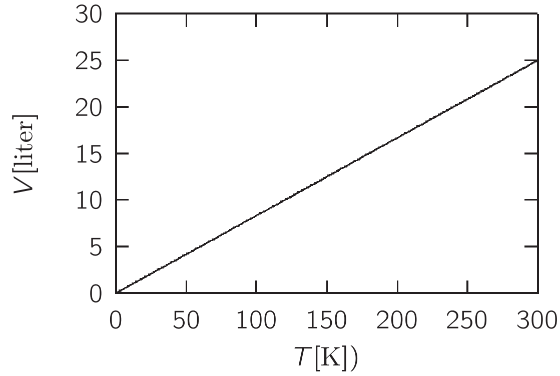


Fig. 2.4: Isobar for a gas with 1 mol particles at 1 bar of pressure. 1 mol corresponds to about 55 g of N_2.

- **Isochor**[16]: An isochor is a line of constant volume. an isochore for the ideal gas is given in Fig. 2.5.

---

[15] Isochor derives from the greek words "isos" meaning "equal" and "baros" meaning "weight"

[16] Isochor derives from the greek words "isos" meaning "equal" and "chora" meaning "space"

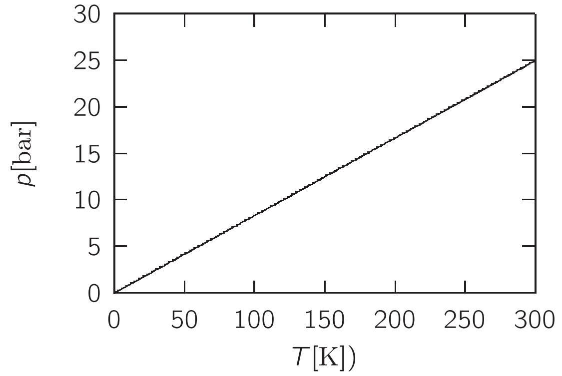


Fig. 2.5: Isochore for a gas with 1 mol particles in a volume of 1 liter. 1 mol corresponds to about 55 g of N_2.

Caloric equation of state

The caloric equation of state Eq. 2.17 of the Boltzmann gas has a very simple interpretation: On average, each molecule has a energy of $\frac{3}{2}k_BT$. Since the energy of free particles is only kinetic, we can use the internal energy to estimate the velocity of the particles. We use the classical expression for the kinetic energy $E_{kin} = \frac{1}{2}mv^2$

$$\frac{3}{2}k_BT = \frac{1}{2}m\langle v^2\rangle \qquad \Rightarrow \qquad \sqrt{\langle v^2\rangle} = \sqrt{\frac{3k_BT}{m}}$$

For a nitrogen molecule at room temperature, we obtain a velocity of 1800 km/h, which is about twice the speed of sound in air. The gas particles have ultrasonic speeds! The same velocity is 5 Å/ps, which amounts to a tenths of an atomic distance within a vibrational period of a bond vibration, which seems to be very slow. It is a general phenomenon that a very high speed in macroscopic dimensions corresponds to a very slow speed on microscopic dimensions.

Chemical potential

The last equation of state Eq. 2.18 does, to my knowledge, not have a name.

$$\begin{aligned}\mu^{BG} &= k_BT\ln\left[\frac{N}{V}\left(\frac{2\pi\hbar^2}{mk_BT}\right)^{\frac{3}{2}}\right] \\ &\overset{Eq.\ 2.23}{=} k_BT\ln\left[\left(\frac{2\pi\hbar^2}{m}\right)^{\frac{3}{2}} p\,(k_BT)^{-\frac{5}{2}}\right] \qquad (2.24)\end{aligned}$$

Editor: The formula above has been incorrect. Following errors have not been removed yet. Affects probably Fig 2.4 and 2.5 and the discussion. Using the fundamental equation we can express the chemical potential as the derivative of the internal energy at constant entropy and volume.

Thus,

CHEMICAL POTENTIAL

$$\mu = \frac{\partial U(S, V, N)}{\partial N} \tag{2.25}$$

The chemical potential can be identified with the energy required to remove one particle.[a]

---

[a] $dU \overset{\text{Eq. 2.25}}{=} \frac{\partial U}{\partial N} dN$. We choose $dN = -1$. 1 is a small number compared to N. Since the energy of the system changes by $-\mu$, we have to supply the energy μ to remove the particle, so that energy conservation is not violated.

The requirement that the entropy remains constant tells that no heat TdS is added or removed during this process. Thus the temperature may adjust in this process.

Note that the chemical potential is not the mean energy of a particle. The latter is defined by U/N and not by dU/dN.

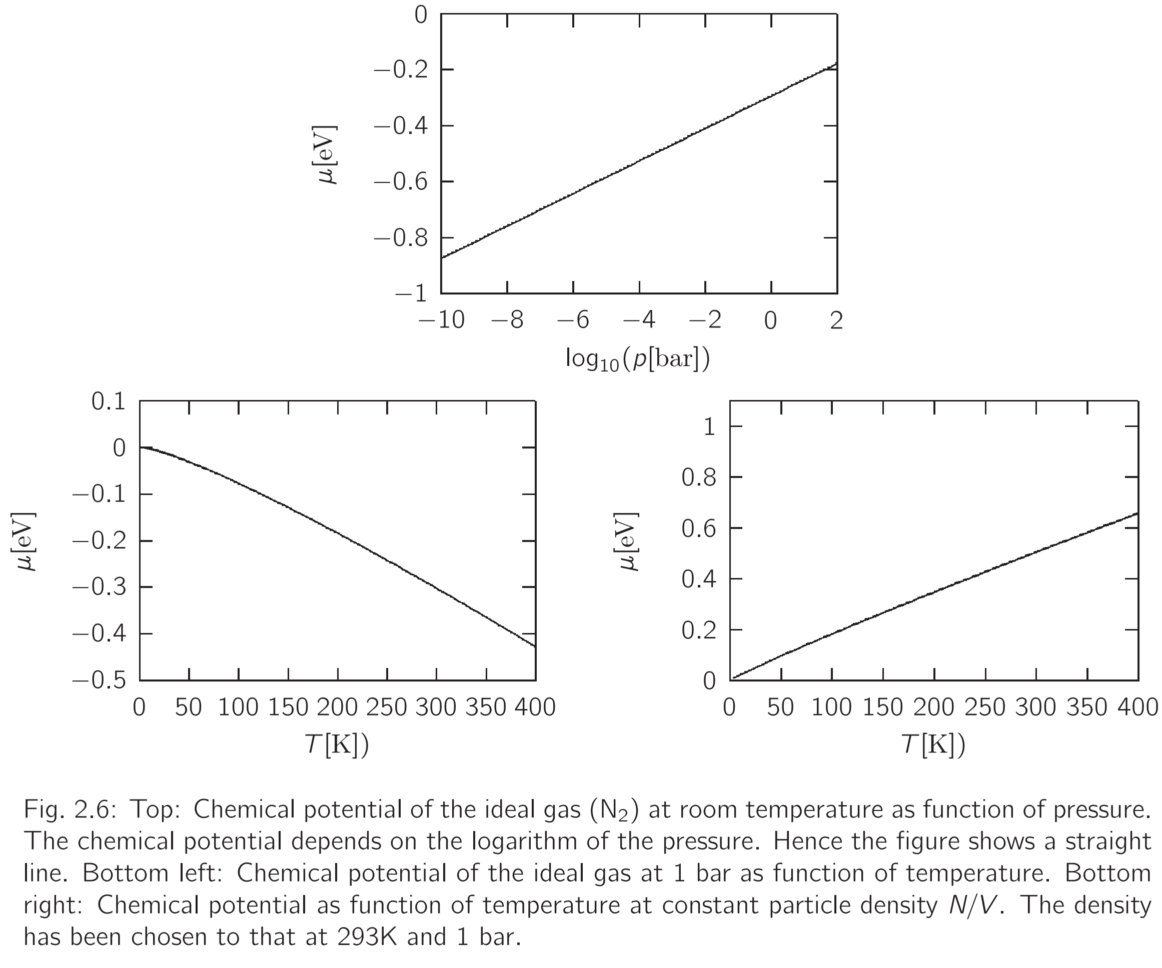


Fig. 2.6: Top: Chemical potential of the ideal gas (N_2) at room temperature as function of pressure. The chemical potential depends on the logarithm of the pressure. Hence the figure shows a straight line. Bottom left: Chemical potential of the ideal gas at 1 bar as function of temperature. Bottom right: Chemical potential as function of temperature at constant particle density N/V. The density has been chosen to that at 293K and 1 bar.

2.5 Response functions

We will discuss the response functions later in more detail. Here we list the definitions and show that they can be calculated from the equations of state.

2.5.1 Coefficient of thermal expansion

Definition 2.4 *COEFFICIENT OF THERMAL EXPANSION*

$$\alpha \overset{def}{=} \frac{1}{V}\left.\frac{dV}{dT}\right|_{p,N} \tag{2.26}$$

The coefficient of thermal expansion describes the relative expansion dV/V at constant pressure, when the temperature is increased[17].

$$\frac{dV}{V} = \alpha dT \qquad \text{at constant } p$$

To obtain the coefficient for thermal expansion for the Boltzmann gas, we use the ideal gas law

$$pV = Nk_BT \Rightarrow V(p,N,T) = \frac{Nk_BT}{p}$$
$$\Rightarrow \alpha^{BG} = \frac{Nk_B}{pV} \overset{Eq.\ 2.23}{=} \frac{1}{T}$$

2.5.2 Isothermal compressibility

Definition 2.5 *ISOTHERMAL COMPRESSIBILITY*
The isothermal compressibility is defined as

$$\kappa_T \overset{def}{=} -\frac{1}{V}\left.\frac{dV}{dp}\right|_{T,N} \tag{2.27}$$

The isothermal compressibility describes the relative contraction of the volume upon changing the pressure at constant temperature.

To obtain the isothermal compressibility for the Boltzmann gas, we use again the ideal gas law

$$pV = Nk_BT \Rightarrow V(p,N,T) = \frac{Nk_BT}{p}$$
$$\Rightarrow \kappa_T^{BG} = \frac{Nk_BT}{p^2V} \overset{Eq.\ 2.23}{=} \frac{1}{p}$$

2.5.3 Adiabatic compressibility

Definition 2.6 *ADIABATIC COMPRESSIBILITY*
The adiabatic compressibility is defined as

$$\kappa_S \overset{def}{=} -\frac{1}{V}\left.\frac{dV}{dp}\right|_{S,N} \tag{2.28}$$

The adiabatic compressibility describes the relative contraction of the volume upon changing the pressure at constant entropy.

---

[17] Multiply the equation for the coefficient of thermal expansion with dT.

2.5.4 Specific heat at constant volume

Definition 2.7 *SPECIFIC HEAT AT CONSTANT VOLUME*
The specific heat at constant volume is defined as

$$c_V \stackrel{def}{=} \left.\frac{dU}{dT}\right|_{V,N} \tag{2.29}$$

The specific heat specifies the amount of energy the system gains when its temperature is raised, while the volume is kept constant. Note that the specific heat is often given either per volume or per mole of substance. The molar specific heat is $\frac{N_A}{N}c_V$, *where* N_A *is Avogadro's constant. The specific heat per volume is* $\frac{1}{V}c_V$.

To obtain the specific heat at constant volume , we use the caloric equation of state

$$\begin{aligned} U &= \frac{3}{2}Nk_BT \\ \Rightarrow c_V^{BG} &= \frac{3Nk_B}{2} \end{aligned}$$

2.5.5 Specific heat at constant pressure

Definition 2.8 *SPECIFIC HEAT AT CONSTANT PRESSURE*
The specific heat at constant pressure is defined as

$$c_p \stackrel{def}{=} \left.\frac{dU}{dT}\right|_{p,N} \tag{2.30}$$

To obtain the specific heat at constant pressure , we use the caloric equation of state

$$\begin{aligned} U &= \frac{3}{2}Nk_BT \\ \Rightarrow c_p^{BG} &= \frac{3Nk_B}{2} \end{aligned}$$

Since the internal energy does not depend on pressure and volume, we obtain the same result as for the specific heat at constant volume.

2.6 Summary

Important formulas

$$\begin{aligned}
U^{BG} &= \frac{3}{2}Nk_BT \\
p^{BG}V &= Nk_BT \\
R &= N_Ak_B \\
p &= -\frac{\partial U(S,V,N)}{\partial V} \\
\lambda_T &= \sqrt{\frac{2\pi\hbar^2}{mk_BT}} \\
S^{BG}(U,V,N) &\overset{\text{Eq. 2.19}}{=} Nk_B\left(\frac{5}{2}-\ln\left[\left(\frac{3\pi\hbar^2}{m}\right)^{\frac{3}{2}}\frac{N^{\frac{5}{2}}}{VU^{\frac{3}{2}}}\right]\right)
\end{aligned}$$

2.7 Exercises

Turbocharger

Determine the amount of gasoline that can be burned per cycle of a car-engine.

- The car shall have a Hubraum of 1.2 liter. One liter equals 10^{-3} m^3
- a turbocharger creates a pressure of 2 bar. One bar is 1 bar=10^5 Pa and the Pascal Pa is the SI unit for pressure, that is 1 Pa=$1\frac{N}{m^2}=1\frac{kg}{ms^2}$. The pressure is the sum of the partial pressures of the individual gases.
- the reaction is

  $$2C_8H_{18}+25O_2 \rightarrow 16CO_2+18H_2O.$$

  It is assumed that all gasoline is burned.
- the engine is filled only every two turnovers.

Classical properties of the ideal gas

In the following we are interested to get some feeling for the ideal gas on a microscopical level. Therefore we perform simple estimates assuming that the gas molecules were classical point particles and would all have the same velocity (in value, not indirection!), which is equal to the mean velocity.

Similar quick and dirty estimates are often useful to get a first qualitative understanding of a system. Some caution and scepticism is always needed. With a some experience, however, one easily sees where the simplifying assumptions fail.

- Determine the mean distance between gas molecules at ambient pressures (1 bar) and temperatures $T = 293\ K = 20^oC$. Attribute an equal fraction of the volume to each gas molecule and determine the radius of a sphere with that volume. The mean distance is defined as twice the radius.
- Determine the thermal energy of one liter of an ideal gas at ambient pressure and temperature. The thermal energy is the energy, which is gained when the system is called to zero temperature.

- Consider that two gas molecules scatter when they come closer than 3 Å. Determine the mean free path, defined as the average distance between two scattering events. Determine the average time between two scattering events.
- Determine the pressure of an ideal gas as function of temperature and pressure from classical physics. Use the caloric equation of state to determine the classical kinetic energy of the particles. Determine then the total momentum change, when particles hit the wall of the container. The resulting force can be converted into a pressure.

Van der Waals gas

The van der Waals gas is an empirical generalization of the ideal gas in order to describe the interaction of the gas molecules. The van der Waals gas exhibits a phase transition where the gas becomes a liquid. The Helmholtz free energy of the van der Waals gas is

$$A(T,V,N) = Nk_BT\left[-1+\ln\left[\left(\frac{2\pi\hbar^2}{mk_BT}\right)^{\frac{3}{2}}\frac{N}{V-Nb}\right]\right]-a\frac{N^2}{V}$$

where a and b are parameters, that do not depend on T,V,N. When we choose the parameters $a=b=0$, we obtain the Boltzmann gas.

The Helmholtz free energy is defined as the free energy F for constant volume and number of particles.

$$A(T,V,N) = -k_BT\ln\left[\sum_{\vec{n}}\mathrm{e}^{-\frac{1}{k_BT}E_{\vec{n}}}\right]$$

For this system we do not know the states $|\vec{n}\rangle$. This equation only serves to make the definition of the Helmholtz potential explicit.

- Determine the equations of state for the entropy $S(T,V,N)$ and determine the internal energy $U(T,V,N)$
- Determine the entropy as function of energy, volume and number of particles.
- Determine the internal energy as function of entropy, volume and number of particles.
- Determine the pressure as function of T,V,N and compare the result to the ideal gas law. (Tip: Do not use $U(S,V,N)$ but show how the pressure is related to the Helmholtz potential and continue from there.)
- Draw the isotherms, isobars and isochores for this system.

Extra killer exercise: Try to rationalize the parameters of the Van der Waals gas from a microscopic model of classical gas molecules.

Chapter 3

Equilibrium

Now we can start playing. The idea is to induce changes of the thermodynamic state in order to heat a system, cool it or let it perform work. Shortly, we want to build some sort of engine.

Every engine is about transforming one form of energy into another. A car engine transforms chemical energy stored in the gasoline first into heat and then, at least partly, into mechanical energy. An electrical motor transforms electrical energy into mechanical energy. An oven transforms electrical energy into heat. A refrigerator uses electrical energy to transport heat into the environment. A plant transforms light into its growth, that is it uses the energy obtained from the sunlight to drive its chemical life processes. An animal uses the chemical energy stored in the food it consumes and the air it breathes into chemical and mechanical energy.

A common feature of all these processes is that they generate heat. Heat generation is a sign of entropy increase. All these processes are driven towards thermodynamic equilibrium. Said more simply, *a system approaches the most likely thermodynamic state, the state with maximum-entropy.*

This is in contrast to the common belief, that a system gets into equilibrium because it is driven towards lower energy. In reality the system simply approaches the most probable state. *"God does play dice. And she does so excessively!"* In order to increase entropy, other forms of energy must be converted into heat. As a consequence, the "useful" energy is minimized. We will show this aspect in more detail within this section.

The way of thinking in thermodynamics is as follows: In order to change the thermodynamic state of a system, we bring it into contact with another one, so that the two systems can exchange a certain variable, such as energy, volume etc. Once contact is established, the combined system can access more quantum mechanical states than before, simply because the two systems are now allowed to exchange the quantity, which was forbidden before contact. As a result, certain processes will start spontaneously. If we bring the systems together in a controlled sequence, we can build an engine. This is why we will first consider certain ways to bring systems into contact.

In the following we will discuss in some detail what happens as different systems are brought into contact, and how they approach thermodynamic equilibrium. Later we will learn how this process can be guided to make the most effective use of the energy consumed.

3.1 Equilibrium

In this section, we will learn the thermodynamic machinery, which can be used with small variations to describe nearly all the thermodynamic processes. Because the methodology appears abstract to the beginner, let us demonstrate it on a well-known mechanical example. The steps used here may appear unnecessarily complicated, but this is mainly due to the simple example chosen.

We consider a series of weights, that are connected to each other, and two fixed endpoints by springs. The energy of a spring has the form $U_i(x_i) = \frac{1}{2}c_i x_i^2$, where x_i is the length of the spring and

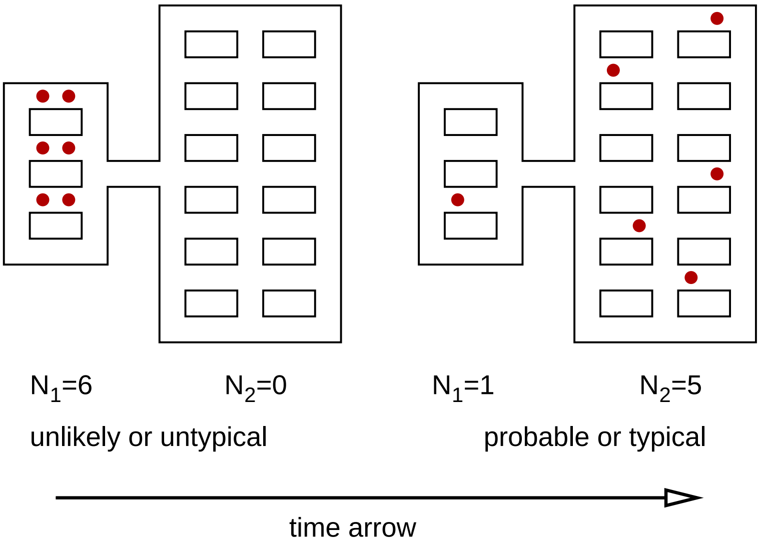
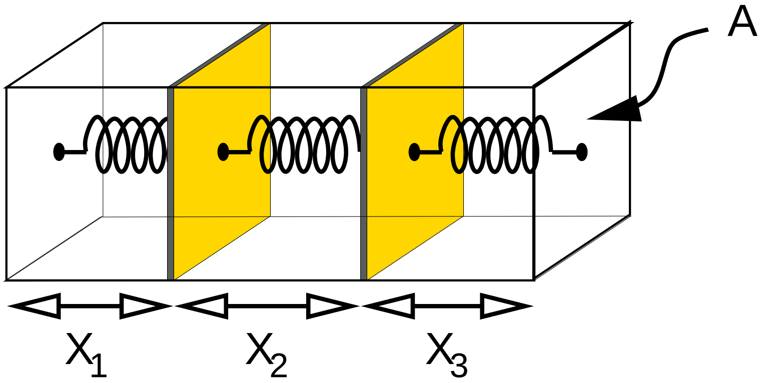


Fig. 3.1: Demonstration of the underlying principle of all thermodynamic machines: approach to disorder. Imagine a library with a small and a large room. Initially there is a meeting in the small room so that everybody is located there. After the meeting the students randomly distribute over all available places. It is immediately clear that, after the while, the students will be distributed as shown in the right figure. If we would videotape the distribution, we could immediately tell if the movie is running forward or backward. Why is that? There is nothing driving the students out of the small room. All processes are time reversible: Students select their places randomly. The particlular distribution of students on the right is just as likely as the one on the left. The answer is the following: We immediately identify characteristics of a distribution, such as the number of people in each room, namely (N_1, N_2). We simply do not distinguish very well between two distributions having the same number of students in each room. Because there is only one distribution of students with $(N_1, N_2) = (6, 0)$ and about 1.5 million with $(N_1, N_2) = (1, 5)$, we naturally expect that the system approaches the more likely situation. We say a configuration on the right is more typical, because it has the same characteristics as 1.5 million other configurations. This little model shows how an "arrow of time" develops spontaneously, even though all processes are time reversible. This arrow of time can be exploited to do useful work. Every thermodynamic machine is based on the same principle: approach to disorder or, in other words, apprach to equilibrium.

c_i is the force constant. When we ask for the equilibrium state, we would like to know the position of the weights (or the lengths of the springs), and of the total energy stored in the system.



If we let the weights loose from some initial positions, they will start moving corresponding to the net forces acting on them. The chain will start to oscillate. Because of energy conservation it will never come to rest, unless there is some friction, which dissipates some of the energy and generates heat. As the mechanical energy is reduced, the oscillations will die out and the weights will come to rest at their equilibrium positions. The equilibrium positions are characterized by the configuration with lowest energy.

Let us now determine the minimum-energy configuration: The constraint is taken care of by the method of Lagrange multipliers.

$$\mathcal{U}_{tot}(X_1, X_2, \dots, X_{tot}, f) = \sum_i U_i(X_i) + f(\sum_i X_i - X_{tot})$$

If the constraint is fulfilled, the function $\mathcal{U}_{tot}$ is nothing else than the total energy of the system.

We will have to fulfill

- the equilibrium conditions

$$\frac{\partial \mathcal{U}_{tot}}{\partial X_i} = 0$$

- and the constraint condition

$$\frac{\partial \mathcal{U}_{tot}}{\partial f} = 0$$

The order in which we will fulfill the conditions is irrelevant.

Before we continue, we regroup the terms in the total energy

$$\mathcal{U}_{tot}(X_1, X_2, \dots, X_{tot}, f) = \sum_i \left[U_i(X_i) + fX_i\right] - fX_{tot}$$

In order to fulfill the equilibrium condition, we minimize the total energy $\mathcal{U}_{tot}$ with respect to the variables X_i.

$$\mathcal{U}_{tot}(X_{tot}, f) = \sum_i \min_{X_i} \left[U_i(X_i) + fX_i\right] - fX_{tot} = \sum_i K_i(f) - fX_{tot}$$

where we have introduced the Legendre-transformed potential

$$K_i(f) \stackrel{\text{def}}{=} \min_X \Big[U_i(X) + fX\Big]$$

The Legendre transform is a common mathematical operation that we will discuss later in more detail.

In our example of mechanical springs, the function $K(f)$ has the form

$$K(f) = \min_X \Big[\frac{1}{2}cX^2 + fX\Big] \stackrel{cX+f=0}{=} \frac{1}{2}c\Big(-\frac{1}{c}f\Big)^2 + f\Big(-\frac{1}{c}f\Big) = -\frac{1}{2c}f^2$$

It is a little disturbing that the variables X_i, that we are ultimately interested in, disappeared from the picture altogether. We will come back to them later.

Now we perform the final step and fulfill the constraint equation, which determines the value of the Lagrange multiplier.

$$0 = \frac{\partial \mathcal{U}}{\partial f} = \sum_i \frac{dK_i}{df} - X_{tot}$$

This is a complex equation for f, and usually it is a non-trivial task to extract the Lagrange multiplier from it. At least numerically it is simple, and it amounts to the standard task of finding the zero of a function.

Now that we know the correct value for the Lagrange multiplier, we need to extract X_i. It can be shown that

$$\frac{\partial K_i(f)}{\partial f}=X_i$$

This identity is shown as follows:

$$\frac{\partial K(f)}{\partial f}=\frac{\partial}{\partial f}\min_X\Big[U(X)+fX\Big]=\frac{d}{df}\Big[U(X^{(0)}(f))+fX^{(0)}(f)\Big]$$

where $X^{(0)}(f)$ is the coordinate at the corresponding minimum.

$$\frac{dK}{\partial f}=\underbrace{\frac{\partial[U(X)+fX]}{\partial X}\bigg|_{X^{(0)}(f)}}_{=0}\frac{\partial X^{(0)}(f)}{\partial f}+\underbrace{\frac{\partial[U(X)+fX]}{\partial f}\bigg|_{X^{(0)}(f)}}_{=X}=X^{(0)}$$

Thus the coordinates are easily obtained as derivative of $K_i(f)$ once the Lagrange multiplier has been determined.

EQUILIBRIUM EVALUATED VIA LEGENDRE TRANSFORM

We summarize the procedure as follows:

1. determine

$$K_i(f)\stackrel{\text{def}}{=}\min_X\Big[U_i(X)+fX\Big] \tag{3.1}$$

   as a function of f for each subsystem

2. extract the Lagrange multiplier f from the constraint condition

$$\sum_i\frac{\partial K_i(f)}{\partial f}=X_{tot} \tag{3.2}$$

3. determine the coordinates

$$\frac{\partial K_i}{\partial f}=X_i \tag{3.3}$$

From a bird's eye view, we have replaced the minimum-energy principle by the principle of equalizing Lagrange multipliers, which are in our case, forces.

3.1.1 Reservoirs and the meaning of the Legendre-transformed potential

So far we have introduced the Legendre-transformed potential $K(f)$ as an abstract quantity used for a mathematical trick. Here we want to develop a physical understanding of this quantity.

Imagine that we bring our systems in contact with one system that is so large that it dominates the behavior of the other systems and that it is not affected in any appreciable way by the state of the other systems. We call such as system a **reservoir**.

RESERVOIR

A reservoir of an extensive quantity X is a system that has a linear energy dependence on X, i.e.

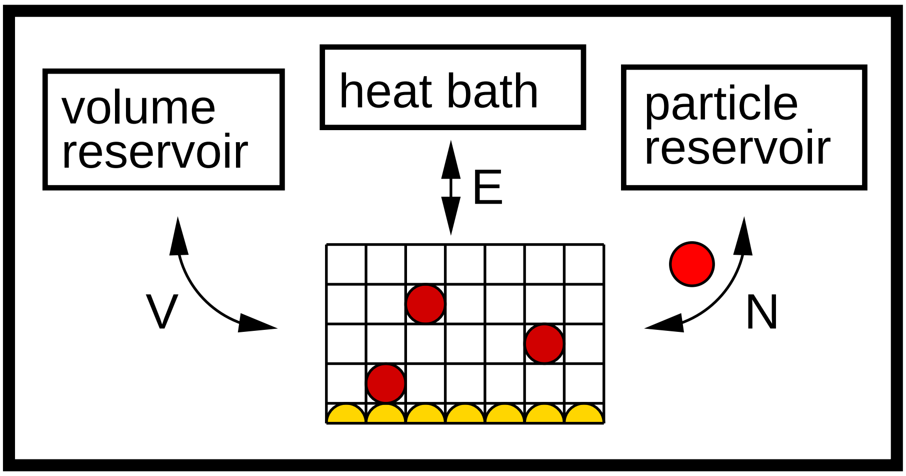
$$U_{XR}(X_{XR}) = -f_{XR}X_{XR} \tag{3.4}$$

The parameter f_{XR} is the single relevent parameter defining the reservoir. The reservoir imposes its value for the corresponding intrinsic parameter onto all systems that exchange the quantity X with it.

Note that, when we add a reservoir to the total energy, it enters with the opposite sign, that is as $U_{XR} = +f_{XR}X = -f_{XR}X_{XR}$. This is because we consider our system together with the reservoir as closed. Hence the total amount $X_{tot} = X_{XR} + X$ of the quantity X remains constant. As long as we compare situations where it is constant, the total amount X_{tot} is of no relevance. Therefore this constant can formally be set to zero.

No equilibrium is possible, if two reservoirs for the same quanity X but with different intrinsic parameters f_{XR} are in contact, directly or indirectly.

In our purely mechanical example the reservoir would be an ultra-long spring, that always exerts the same force on the other systems irrespective of its length. We would call such a spring "length-reservoir", because it can exchange length with the other system without changing its properties.



There can be a reservoir for every conserved quantity that can be exchanged. For us, the most important ones are the

- the heat bath $E(S) = T_{HB}S$ exchanges energy and is characterized by its temperature.
- the volume reservoir $E(V) = -p_{VR}V$ exchanges volume and is characterized by its pressure.
- the particle reservoir $E(N) = \mu_{PR}N$ exchanges particles and is characterized by its chamical potential.

The cautious reader will notice that the heat bath does not properly fit, because it exchanges entropy, which is not a conserved quantity. This is an artefact of the energy representation. Originally we have a maximum-entropy principle and the energy is one of the conserved quantities, that are exchanged between the systems. In this framework the exchanged quantity is energy. After changing to the energy representation these roles got somewhat intermixed.

System in contact with a reservoir

In order to arrive at the physical meaning of the Legendre-transformed potential let us consider a system in contact with a reservoir and determine its equilibrium state:

Let us consider a laboratory system in contact with a reservoir. The total energy is

$$U_{tot}(X,X')=U(X)+U_{XR}(X')-f(X+X'-X_{tot})$$

We directly satisfy the constraint condition by $X'=X_{tot}-X$.

$$U_{tot}(X)=U(X)+U_{XR}(X_{tot}-X)=U(X)-f_{XR}(X_{tot}-X)=U(X)+f_{XR}X-f_{XR}X_{tot}$$

Now we minimize the internal energy and obtain X_0

$$\left.\frac{\partial U}{\partial X}\right|_{X_0}+f_{XR}=0 \qquad \rightsquigarrow \qquad X_0$$

If we insert this result we obtain

$$U_{tot}(X_0(f_{XR}))=U(X_0)+f_{XR}X_0-f_{XR}X_{tot}=K(f_{XR})-f_{XR}X_{tot}$$

Because of the complete amount X_{tot} is not relevant in this context, we may choose it equal to zero. With this convention we can express the Legendre transformed potential as the minimum value of the internal energy of the Laboratory system and its environment, the reservoir.

Thus, up to a constant, the Legendre transform $K(f)=\min_X\left[U(X)+fX\right]$ of the internal energy $U(X)$ of a laboratory system is the internal energy of the laboratory system combined with its environment, with which it can exchange the variable X. The environment is characterized by the variable f and it is described by a reservoir with internal energy $U_{XR}(X')=-fX'$.

Processes in contact with a reservoir

In order to arrive at a deeper understanding of the Legendre transformed potential, let us introduce another parameter y, so that the internal energy $U(X,y)$ depends on two variables. The parameter y is like a handle, that can be adjusted at will, and that allows me to drive the system through a thermodynamic process.

As we change the parameter y, the system exchanges an amount of the quantity X with the resevoir in order to reestablish equilibrium. The equilibrium values of X define a path $X^{(0)}(y)$ as function of y.

Let us now follow the energy changes as the variable y is changed from one value, y_1, to another one, y_2.

The energy change of our system is

$$\Delta U=U(X^{(0)}(y_2),y_2)-U(X^{(0)}(y_1),y_1)$$

Because the total amount of X remains fixed, the amount of X in the reservoir changes by

$$\Delta X_{XR}=-\Delta X=-\Big[X^{(0)}(y_2)-X^{(0)}(y_1)\Big]$$

The energy change of the reservoir is

$$\Delta U_{XR}=-f\Delta X_{XR}=f\Delta X=fX^{(0)}(y_2)-fX^{(0)}(y_1)$$

Hence the change of the total energy is

$$\Delta U_{tot}=\Delta U+\Delta U_{XR}=\Big[U(X^{(0)}(y_2),y_2)+fX^{(0)}(y_2)\Big]-\Big[U(X^{(0)}(y_1),y_1)+fX^{(0)}(y_1)\Big]$$

With

$$K(f,y)=\min_X\Big[U(X,y)+fX\Big]=U(X^{(0)}(y),y)+fX^{(0)}(y)$$

we obtain the total change in energy ΔY_{tot} as the parameter changes from y_1 to y_2 as

$$\Delta U_{tot}=K(f,y_2)-K(f,y_1)$$

Thus we can relate the change of the Legendre-tranformed potential with the total-energy changes, when the system is in contact with a reservoir.

MEANING OF THE LEGENDRE-TRANSFORMED POTENTIAL

- A change of the Legendre-transformed potential is the sum of the energy changes of the system *and* the reservoir.
- The change of the Legendre-transformed potential describes the energy change of the system and the work done on the reservoir, respectively the heat transferred to the heat bath.
- The heat TdS generated by a spontaneous process that proceeds in contact with a reservoir, is opposite to a change of the Legendre transformed potential.
- A process in contact with a reservoir will approach the minimum value of the Legendre-transformed potential. (re-formulation of the minimum-energy principle)

3.2 Legendre Transforms

An important step in the derivations given above was a Legendre transform from $U(X)$ to $K(f)$. Legendre transforms are crucial to the mathematical formalism of thermodynamics and the important Quantities in thermodynamics can only be understood in these terms.

Therefore let us summarize the main aspects:

LEGENDRE TRANSFORM

The **Legendre transform** from $U(X)$ to $K(f)$ and its back-transform can be defined as

$$K(f)=\min_X\Big[U(X)+fX\Big]$$
$$U(X)=\min_f\Big[K(f)-fX\Big]$$

The back-transform leads back to the original function $U(X)$ if $U(X)$ is purely convex[a]. If $U(X)$ is not convex, the back-transform is the lower straight-line envelope of $U(X)$. This envelope is concave. The definition is adapted to the applications in thermodynamics. The most general definition, there the minimum condition can be also a maximum condition. One may define the signs of U, f, K in an arbitrary manner, and instead of the Legendre transform is bijective, i.e. one-to-one, if the functions are either purely convex or purely concave.

---

[a] A function $f(x)$ is **convex** if the secant through two points $(x_1, f(x_1))$ and $(x_2, f(x_2))$ lies on or above the function within the interval $[x_1, x_2]$, that is

$$f(\lambda x_1+(1-\lambda)x_2)\leq \lambda f(x_1)+(1-\lambda)f(x_2)\qquad\text{for}\qquad 0\leq\lambda\leq 1$$

. To memorize it: A convex lens is shaped like an "X". If one lies the lens flat down, the upper surface of is convex.

The Legendre transform is performed in two steps:

- starting from $U(X)$ one determines the value of the minimum $X^{(0)}(f)$ of $U(X)+fX$ as function of f. Thus we need to resolve $f=-\frac{\partial U(X)}{\partial x}$ for X.

- the Legendre-transformed potential is obtained by insertion, i.e. $K(f)=U(X^{(0)}(f)+fX^{(0)}(f)$

3.2.1 Geometric interpretation

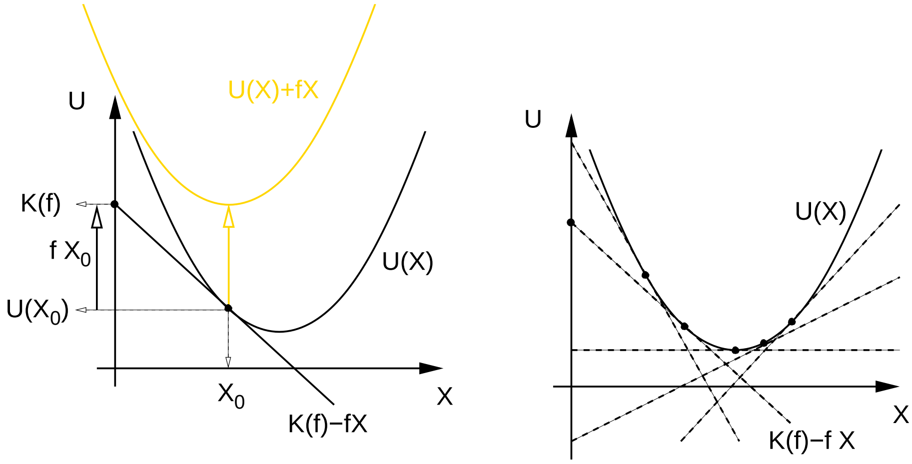


Fig. 3.2: Graphical representation of the Legendre transform. One starts out from the original function $U(V)$. The ordinate of a tangent $t_f(X)=K(f)-fX$ with a given slope $-f$ is the value of the corresponding Legendre-transformed potential $K(f)$. The enthalpy is at the same time the minimum of the sum $U(X)+fX$ of the internal energy and the energy $E_{XR}=fX$ of the reservoir with a given force, as illustrated in yellow. Once the Legendre transformed function $K(f)$ is known, the original function, the internal energy $U(X)$ can be reconstructed from the crossing points of the tangents $t_p(X)=K(f)-fX$ with infinitesimally different slope.

The Legendre transform $K(f)$ is constructed from a function $U(X)$ as the value of the tangent to the function $U(X)$ with slope $-f$ at $X=0$.

The tangent $T_f(X)$ to $U(X)$ at point X_0 is

$$T_f(X)=U(X_0)+\underbrace{\left.\frac{dU}{dX}\right|_{X_0}}_{-f}(X-X_0)$$

The value of the tangent with slope $-f$ at $X=0$ defines the Legendre transform

$$K(f)\overset{\text{def}}{=}T_f(0)=U(X_0)-\underbrace{\left.\frac{dU}{dX}\right|_{X_0}}_{-f}X_0$$

Thus **the Legendre transformed function $K(f)$ is the value at crossing between the vertical axis with the tangential line to the function $U(X)$ with slope $-f$.**

The definition can be written in two equations, which define the forward Legendre transform

$$K(f)=U(X)+fX \qquad \text{where} \qquad f=-\frac{\partial U(X)}{\partial X}$$

The original function $U(X)$ can be reconstructed from $K(f)$ as the curve, that is enveloped by the set of tangential lines. A point on this function can be constructed as the crossing of two tangents with infinitesimally different slopes. This provides us with a rule for the back transform: Consider two tangential lines

$$
\begin{aligned}
T_f(X) &= K(f) - fX \\
T_{f+\Delta f}(X) &= K(f+\Delta f) - (f+\Delta f)X
\end{aligned}
$$

Now we determine the crossing X_0 of two tangents with nearly identical slope from

$$
\begin{aligned}
& T_f(X_0) = T_{f+\Delta_f}(X_0) \\
\Rightarrow\ & K(f) - fX_0 = K(f+\Delta f) - (f+\Delta f)X_0 \overset{Taylor}{=} K(f) + \frac{dK}{df}\Delta f + O(\Delta f)^2 - fX_0 - \Delta f X_0 \\
\Rightarrow\ & 0 = (\frac{dK}{df} - X_0)\Delta f + O(\Delta f)^2 \\
\overset{\Delta f \to 0}{\Rightarrow}\ & X_0 = \frac{dK(f)}{df}
\end{aligned}
$$

The value of the tangential line at X_0 yields the original function

$$
U(X^0) = T_f(X^0) = K(f) + \underbrace{\frac{dK(f)}{df}}_{X_0} f
$$

Thus we can write the back transform in two equations

$$
U(X) = K(f) - fX \qquad \text{where} \qquad X = \frac{dK(f)}{df}
$$

The Legendre transform is unique if the original function is either concave $\frac{\partial U}{\partial X} > 0$ or convex $\frac{\partial U}{\partial X} < 0$. The Legendre transformed function is defined on the interval between minimum and maximum slope. Interestingly the function does not need to be continuous nor differentiable.

3.2.2 Examples:

One minimum

The most simple example for a Legendre transform is to start from a parabolic internal energy.

$$
U(X) = U_0 + \frac{1}{2}c(X - X_0)^2
$$

The minimum U_0 of the internal energy $U(X)$ is at X_0.

The Legendre transformed thermodynamic potential $K(f)$ is obtained (1) by adding the internal energy fX of an X-reservoir, and then (2) by determining the minimum value of this total energy with respect to X:

$$
K(f) = \min_X \Big[U_0 + \frac{1}{2}c(X - X_0)^2 + fX\Big] = U_0 + \frac{1}{2}c(X^{(0)} - X_0)^2 + fX^{(0)}\Big] \tag{3.5}
$$

The position of the minimum $X^{(0)}$ is

$$
c(X^{(0)} - X_0) + f = 0 \qquad \Rightarrow \qquad X^{(0)}(f) = X_0 - \frac{1}{c}f \tag{3.6}
$$

Thus we obtain

$$
\begin{aligned}
K(f) &\overset{\text{Eq. 3.5}}{=} U_0 + \frac{1}{2}c(X^{(0)} - X_0)^2 + fX^{(0)} \\
&\overset{\text{Eq. 3.6}}{=} U_0 + \frac{1}{2}c(X_0 - \frac{1}{c}f - X_0)^2 + f(X_0 - \frac{1}{c}f) \\
&= U_0 + \frac{1}{2c}f^2 + fX_0 - \frac{1}{c}f^2 \\
&= U_0 + fX_0 - \frac{1}{2c}f^2 \\
&= U_0 + \frac{1}{2}cX_0^2 - \frac{1}{2c}(f - cX_0)^2 \\
&= U(0) - \frac{1}{2c}(f - cX_0)^2
\end{aligned}
$$

The Legendre transform $K(f)$ has negative curvature, with a maximum at $f = cX_0$.

Phase transition of first order:

One requirement for a bijective[1] Legendre transform is that the function is either concave or convex. This requirement is violated if the internal energy has several minima, as it is the case near phase transitions. One can learn a lot about Legendre transforms by investigating these cases. Because the qualitative features of the internal energy near a phase transition are very similar, it is helpful to make oneself familiar with the resulting patterns.

Let us start with an internal energy $U(X)$ that has a double well feature.

$$U(X) = (X^2 - 1)^2 + X$$

The thermodynamic potentials and equations of state for this model are shown in Fig. 3.3 for $\alpha = 1$.

The system has one global minimum and one metastable minimum. If the thermodynamic force changes, the system will find one or the other minimum more favorable. The thermodynamic force f is defined by the reservoir, the X-reservoir, coupled to our system. Including the energy fX of the reservoir, the total energy $U(X)+fX$, the global and the metastable minimum interchange their role with the sign of $f+1$. As a result the physically relevant state drops into the new global minimum, which results in an abrupt change of the actually realized state variable X. Such an abrupt change of a state variable is characteristic for a first-order phase transition.

Before we continue, let us evaluate the "force" $f(X)$ and the Legendre transformed potential $K(f)$. Because the relation $X \rightarrow f$ cannot be inverted, the Legendre transform $K(f)$ cannot be given in a closed form as function of f.

$$
\begin{aligned}
f(X) = -\frac{dU}{dX} &= -4X(X^2 - 1) - 1 \\
K(f(X)) = U(X) + fX &= -(X^2 - 1)(3X^2 + 1)
\end{aligned}
$$

The Legendre forward transform from $U(X)$ to $K(f)$ is unique and leads to the two black segments of $K(f)$ in Fig. 3.3. $K(f)$ is defined as the lowest value of a multivalued function. Characteristic is that the Legendre transformed function is not differentiable, but that it has a kink at the intersection of line segments (1) and (5). At this point, tangents with a number of different slopes can be placed at $K(f)$. Each slope corresponds to one possible value for $X = \frac{dK}{df}$. These values correspond to the dashed line in the figure for $X(f)$.

The back transform leads to $U(X)$, but the non-concave segments (2), (3) and (4) are replaced by a dashed straight line segment. For this new function the Legendre transform is bijective.

---

[1] A mapping between two sets is bijective if it establishes a one-to-one correspondence between the elements of the two sets. The Legendre transform is bijective, if the Legendre backtransform leads back to the original function.

If the function $U(X)$ is concave, the slope increases with increasing X. Hence the force f decreases with increasing X. Thus as we follow $U(X)$ from left to right, the function $K(f)$ goes from right to left, as in the black and red line segments. If the function $U(X)$ is convex, as in the blue line segment, the Legendre transform $K(f)$ goes from left to right, while we follow the corresponding points of $U(X)$ in the same directionfrom left to right.

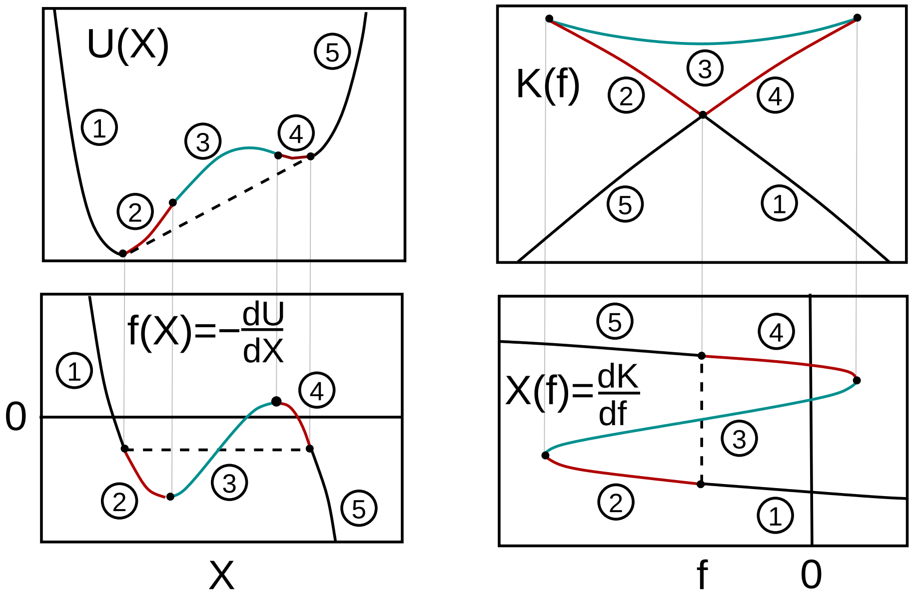


Fig. 3.3: Legendre transform for an internal energy with several minima

3.2.3 Legendre transform in classical mechanics

We have encountered Legendre transformations before, namely, when we transformed from the Lagrange function $\mathcal{L}(x, v, t)$ to the Hamilton function $H(p, x, t)$

$$H(p_i, x_i, t) = \sum_i p_i v_i - \mathcal{L}(x_i, v_i, t)$$

$$p_i = \frac{\partial \mathcal{L}}{\partial v_i}$$

where the momentum p_i is the conjugate quantity to the velocity v_i. The back transformation is

$$\mathcal{L}(x_i, v_i, t) = \sum_i p_i v_i - H(p_i, x_i, t)$$

$$v_i = \frac{\partial H}{\partial p_i}$$

Both yield the same equations of motion: Starting with the Hamilton function we use Hamilton's equations

$$\dot{x}_i = \frac{\partial H}{\partial p_i} \qquad \wedge \qquad \dot{p}_i = -\frac{\partial H}{\partial x_i}$$

whereas we use the Euler-Lagrange equations, if the Lagrange function is known.

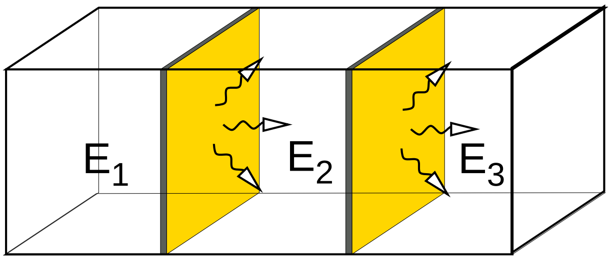
$$\dot{x}_i = v_i \qquad \wedge \qquad \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial v_i}\right) = \frac{\partial \mathcal{L}}{\partial x_i}$$

3.3 From the maximum-entropy principle to the minimum-energy principle

We have motivated the minimum energy principle as result of the maximum entropy principle, by saying entropy growth is related to heat creation. Heat creation in the context of a mechanical or other non-thermal system is accompanied by dissipation of energy. Thus the energy is minimized.

Here we want to do things in more detail without assuming the minimum energy principle directly. Rather, we modify the approach described for the exchange of mechanical energy, and adapt it to our problem.

Imagine a set of systems in thermal contact, that is, they are able to exchange heat.



The total entropy of the system is maximized, while the total energy is conserved.

$$\mathcal{S}_{tot}(U_1,U_2,\ldots,U_{tot},T)=\sum_i S_i(U_i)-\frac{1}{T}\left(\sum_i U_i-U_{tot}\right) \tag{3.7}$$

$$=-\frac{1}{T}\Big(\sum_i\left[U_i-TS_i(U_i)\right]-U_{tot}\Big) \tag{3.8}$$

Now we maximize the expression with respect to the energies U_i, and obtain

$$\begin{aligned}S_{tot}(U_{tot},T)&=\max_{U_1,U_2,\ldots}\mathcal{S}(U_1,U_2,\ldots,U_{tot},T)\\&=-\frac{1}{T}\Big(\sum_i\min_{U_i}\left[U_i-TS_i(U_i)\right]-U_{tot}\Big)\\&=-\frac{1}{T}\Big(\sum_i F_i(T)-U_{tot}\Big)\end{aligned} \tag{3.9}$$

where we introduced the free energy

$$F_i(T)=\min_U\Big[U-TS_i(U)\Big] \tag{3.10}$$

From Eq. 3.8, we can read that the entropy is maximized, when the free energy $\mathcal{F}$ defined by

$$\mathcal{F}(U_1,U_2,\ldots,U_{tot},T)\stackrel{\text{def}}{=}U_{tot}-T\mathcal{S}_{tot}(U_1,U_2,\ldots,U_{tot},T)$$

is minimized with respect to the U_i.

From the minimum condition in the definition of F_i we obtain

$$\frac{\partial F_i}{\partial T}=-S(U_i^{(0)}(T))=-S_i(T)$$

From the constraint condition, we obtain the temperature.

$$0 \overset{\text{Eq. 3.7}}{=} \frac{\partial \mathcal{S}_{tot}}{\partial(1/T)} = -\Big(\sum_i F_i(T) - U_{tot}\Big) - \frac{1}{T}\Big(\sum_i \underbrace{\frac{\partial F_i}{\partial T}}_{S_i}\underbrace{\frac{\partial T}{\partial(1/T)}}_{-T^2}\Big)$$

$$= \quad -\Big(\sum_i \underbrace{F_i(T) + T\overbrace{\frac{\partial F_i}{\partial T}}^{S_i(U_i^{(0)})}}_{U_i^{(0)}} - U_{tot}\Big)$$

Finally we can perform a back transform of the free energy to obtain

$$U_i(S) = \max_T\Big[F_i(T) + TS\Big] \tag{3.11}$$

3.4 Thermodynamic potentials

In the following we will investigate varous common arrangements of contacts, which consist of attaching one or more reservoirs to a system. The $U(S, V, N)$ and its Legendre-transformed potentials are called **thermodynamic potentials**. The thermodynamic potentials contain all the information on thermodynamic equilibrium states. The same physical quantity expressed by the “wrong” set of variables is not a thermodynamic potential, but an equation of state. An example would be the internal energy expressed by the temperature instead of the entropy.

Most of the thermodynamic potential for the variable set (T, S), (p, V) and (μ, N) have names. This is because their importance for gases and chemical reactions.

NAMED THERMODYNAMIC POTENTIALS

(S, V, N)	Internal energy	$U(S, V, N)$
(T, V, N)	Helmholtz potential	$A(T, V, N) = \min_S U(S, V, N) - TS$
(S, p, N)	Enthalpy	$H(S, p, N) = \min_V U(S, V, N) + pV$
(T, p, N)	Gibbs potential	$G(T, p, N) = \min_{S,V} U(S, V, N) - TS + pV$
(T, V, μ)	Grand canonical potential	$\Phi(T, V, \mu) = \min_{S,N} U(S, V, N) - TS - \mu N$

The naming is, however, not uniform:

- The thermodynamic potentials that contain the temperature are called **free energies**.
- The Helmholtz potential, in particular, is the prototypical free energy. It is also named simply free energy and it is given the symbol F instead of A.
- The Gibbs potential[2] is also named Gibbs enthalpy, free enthalpy, Gibbs free energy, or Gibbs energy or Gibbs function. According to IUPAC the official name is Gibbs energy or Gibbs function.
- The grand-canonical potential is also named **Landau function**.

---

[2] Josiah Willard Gibbs: Feb. 11, 1839-Apr. 28, 1903. Spent nearly his entire career at Yale university in Connecticut, USA. Visited Paris, Berlin and Heidelberg in the years from 1866-1869. Was Influenced in Heidelberg by Kirchhoff and Helmholtz.

3.4.1 The fundamental relation

The fundamental relation is the key to the equations of state, because it shows the relation of a thermodynamic potential to the dependent variables.

We start from the fundamental relation for the internal energy

$$dU = TdS - pdV + \mu dN$$

The content of this equation are the three equations of state

$$T(S,V,N) = \frac{\partial U}{\partial S} \quad ; \quad p(S,V,N) = -\frac{\partial U}{\partial V} \quad ; \quad \mu(S,V,N) = \frac{\partial U}{\partial N}$$

The fundamental relation can be expressed equally in terms of the other thermodynamic potentials

$$\begin{aligned}
dA &\overset{A=U-TS}{=} \underbrace{TdS - pdV + \mu dN}_{dU} \underbrace{-SdT - TdS}_{-d(TS)} = -SdT - pdV + \mu dN \\
dH &\overset{H=U+pV}{=} \underbrace{TdS - pdV + \mu dN}_{dU} \underbrace{+pdV + Vdp}_{d(pV)} = TdS + Vdp + \mu dN \\
dG &\overset{G=U-TS+pV}{=} \underbrace{TdS - pdV + \mu dN}_{dU} \underbrace{-SdT - TdS}_{-d(TS)} \underbrace{+pdV + Vdp}_{d(pV)} = -SdT + Vdp + \mu dN \\
d\Phi &\overset{\Phi=U-TS-\mu N}{=} \underbrace{TdS - pdV + \mu dN}_{dU} \underbrace{-SdT - TdS}_{-d(TS)} \underbrace{-\mu dN - Nd\mu}_{-d(\mu N)} = -SdT - pdV - Nd\mu
\end{aligned}$$

3.4.2 Conversion of themodynamic potentials

A typical problem in thermodynamics is that one has an experiment where one set of extensive of intensive variables is controlled while the thermodynamic information is available in another set of variables.

- If we have some process where we suppress exchange of an extensive variable with the environment or with another system, it is convenient to formulate all expressions by the extensive variable.
- If the thermodynamic system is in contact with a reservoir for this quantity, the reservoir will determine the value of the intensive variable, so that we would formulate all expressions in terms of the intensive variable.

One way is to perform a variable transform of the equations of state. However this may be cumbersome. Therefore one prefers to transform the thermodynamic potential, and derive the equations of state from the transformed thermodynamic potential. Thus we obtain the information directly in the desired variable set.

FUNDAMENTAL RELATIONS FOR NAMED THERMODYNAMIC POTENTIALS

Internal energy	$U(S,V,N)$	$dU = TdS - pdV + \mu dN$
Helmholtz potential	$A(T,V,N) = U - TS$	$dA = -SdT - pdV + \mu dN$
Enthalpy	$H(S,p,N) = U + pV$	$dH = TdS + Vdp + \mu dN$
Gibbs potential	$G(T,p,N) = U - TS + pV$	$dG = -SdT + Vdp + \mu dN$
Grand-canonical Potential	$\Phi(T,V,\mu) = U - TS - \mu N$	$d\Phi = -SdT - pdV - Nd\mu$

COOKING RECIPE FOR THE CONVERSION OF THERMODYNAMIC POTENTIALS

The goal will be to obtain the thermodynamic potential with one variable set, say the Gibbs potential $G(T, p, N)$, from a another, known thermodynamic potential for another variable set, say $U(S, V, N)$.

1. Have one thermodynamic potential such as $U(S, V, N)$ and its fundamental relation $dU = TdS - pdV + \mu dN$ ready. The thermodynamic potential encodes the complete thermodynamic information of the system. The fundamental relation defines the corresponding conjugate variables $(T, -p, N)$ and thus contains the equations of state. [a]

2. Identify the variables to be replaced by their conjugate variables: In this case we need to replace S by T and V by p.

3. For each variable add the variation of the product of old and new variable, i.e. $d(TS)$ and $-d(pV)$, with a sign opposite to the corresponding terms, i.e. TdS and $-pdV$, in the original fundamental relation. The signs must be opposite to remove the dependence of the final expression on the variables, i.e. S and V, that should be replaced by their conjugate variables T and $-p$.

   $$\begin{aligned}\underbrace{dU - d(TS) - d(pV)}_{dG} &= \underbrace{TdS - pdV + \mu dN}_{dU} \underbrace{-TdS - SdT}_{-d(TS)} \underbrace{+pdV + Vdp}_{d(pV)} \\ &= -SdT + Vdp - \mu dN\end{aligned}$$

   Thus we have defined the Legendre transformation required to obtain the new thermodynamic potential, namely

   $$G(T, p, N) = \min_{S,V} \left[U(S, V, N) - TS + pV\right]$$

4. Now we form the derivative of the term in the parenthesis with respect to the two variables S, V that shall be replaced. These are the variables which are varied to find the minimum. This provides us with two equations of state, namely

   $$\frac{\partial U(S, V, N)}{\partial S} = T \qquad \text{and} \qquad \frac{\partial U(S, V, N)}{\partial V} = -p$$

   This system of equations is solved for S and V to obtain these quantities as function of the new variable set. We obtain a result of the form $S = S_0(T, p, N)$ and $V = V_0(T, p, N)$.

5. Now we obtain the new thermodynamic potential as

   $$G(T, p, N) = U\Big(S_0(T, p, N), V_0(T, p, \mu), N\Big) - TS_0(T, p, N) + pV_0(T, p, N)$$

---

[a] The fundamental relation can be expressed in any thermodynamic potential, but it is economic to memorize the one for the internal energy and the definition of the other thermodynamic potentials. The other fundamental relations can then easily be constructed.

3.5 Thermodynamic potentials of the Boltzmann gas

We start out from the internal energy that we determined earlier from the Sackur-Tetrode equation

$$U^{BG}(S, V, N) \overset{\text{Eq. 2.20}}{=} \frac{3\pi\hbar^2}{m} N^{\frac{5}{3}} V^{-\frac{2}{3}} \mathrm{e}^{\frac{2S}{3Nk_B} - \frac{5}{3}} \tag{3.12}$$

and determine the thermodynamic potentials of the Boltzmann gas for the other sets of variables.

3.5.1 Thermal contact: Helmholtz potential

The Helmholtz potential is the thermodynamic potential for a system in thermal contact, that is it can exchange heat with other systems or a heat bath.

In order to determine the Helmholtz potential for the Boltzmann gas, we start from the definition

$$A(T,V,N) = \min_S \Big[U(S,V,N) - TS\Big] \tag{3.13}$$

which specified the Helmholtz potential as the appropriate Legendre transform of the internal energy. The internal energy of the Boltzmann gas is known from Eq. 3.12.

In order to eliminate the entropy, we need todetermine the minimum of the above expression with respect to the entropy.

$$\begin{aligned}
0 &= \frac{\partial}{\partial S}\Big[U(S,V,N) - TS\Big] \qquad \Rightarrow \qquad \frac{\partial U}{\partial S} = T \\
\Rightarrow \qquad T &= \frac{\partial U^{BG}}{\partial S} \overset{\text{Eq. 3.12}}{=} \frac{\partial}{\partial S}\left(\frac{3\pi\hbar^2}{m} N^{\frac{5}{3}} V^{-\frac{2}{3}} \mathrm{e}^{\frac{2S}{3Nk_B}-\frac{5}{3}}\right) \\
&= \left[\frac{2}{3Nk_B}\right] \underbrace{\left(\frac{3\pi\hbar^2}{m} N^{\frac{5}{3}} V^{-\frac{2}{3}} \mathrm{e}^{\frac{2S}{3Nk_B}-\frac{5}{3}}\right)}_{U^{BG}}
\end{aligned} \tag{3.14}$$

As a small side remark, we have just obtaine the caloric equation of state, Eq. 2.17, for the Boltzmann gas

$$\overset{\text{Eq. 3.14}}{\Rightarrow} \qquad U^{BG} = \frac{3}{2} N k_B T \tag{3.15}$$

When we resolve Eq. 3.14 for the entropy, we obtain the entropy that minimizes the expression in Eq. 3.13 as function of temperature.

$$\begin{aligned}
\overset{\text{Eq. 3.14}}{\Rightarrow} \qquad S &= \frac{3Nk_B}{2}\left(\frac{5}{3} + \ln\left[T \cdot \frac{3Nk_B}{2} \frac{m}{3\pi\hbar^2} N^{-\frac{5}{3}} V^{\frac{2}{3}}\right]\right) \\
&= Nk_B\left(\frac{5}{2} + \ln\left[\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}} VN^{-1}\right]\right)
\end{aligned} \tag{3.16}$$

Insertion of the optimum entropy Eq. 3.16 into Eq. 3.13 leads to to the final expression for the Helmholtz potential.

$$\begin{aligned}
A(T,V,N) &= U(T,V,N) - TS(T,V,N) \\
&\overset{\text{Eqs. 3.15,3.16}}{=} \underbrace{\frac{3}{2}Nk_BT}_{U(T,V,N)} - \underbrace{Nk_BT\left(\frac{5}{2} + \ln\left[\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}} VN^{-1}\right]\right)}_{TS(T,V,N)} \\
&= Nk_BT\left(-1 - \ln\left[\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}} VN^{-1}\right]\right) \\
&= Nk_BT\left(+\ln\left[\left(\frac{2\pi\hbar^2}{mk_BT}\right)^{\frac{3}{2}} \frac{N}{V}\right] - 1\right)
\end{aligned}$$

This is our final result for the Helmholtz potential $A(T,V,N)$.

HELMHOLTZ POTENTIAL OF THE BOLTZMANN GAS

$$A^{BG}(T,V,N)=Nk_BT\left(\ln\left[\left(\frac{2\pi\hbar^2}{mk_BT}\right)^{\frac{3}{2}}\frac{N}{V}\right]-1\right) \tag{3.17}$$

3.5.2 Thermal contact and Volume exchange: Gibbs potential

$$G(T,p,N)=\min_{S,V}\Big[U(S,V,N)-TS+pV\Big]=\min_{V}\Big[A(T,V,N)+pV\Big]$$

$$\begin{aligned}\frac{\partial A^{BG}}{\partial V}&=-p\\ \frac{\partial}{\partial V}\left\{Nk_BT\left(\ln\left[\left(\frac{2\pi\hbar^2}{mk_BT}\right)^{\frac{3}{2}}\frac{N}{V}\right]-1\right)\right\}&=-Nk_BT\frac{1}{V}=-p\\ V&=\frac{Nk_BT}{p}\end{aligned}$$

$$\begin{aligned}G(T,p,N)&=A^{BG}(T,V^{(0)},N)+pV^{(0)}\\ &=Nk_BT\left(\ln\left[\left(\frac{2\pi\hbar^2}{mk_BT}\right)^{\frac{3}{2}}\frac{p}{k_BT}\right]-1\right)+Nk_BT\\ &=Nk_BT\ln\left[\left(\frac{2\pi\hbar^2}{mk_BT}\right)^{\frac{3}{2}}\frac{p}{k_BT}\right]\end{aligned}$$

GIBBS POTENTIAL OF THE BOLTZMANN GAS

$$G(T,P,N)=Nk_BT\ln\left[\left(\frac{2\pi\hbar^2}{mk_BT}\right)^{\frac{3}{2}}\frac{p}{k_BT}\right] \tag{3.18}$$

3.5.3 Enthalpy

Let us now determine the enthalpy for the Boltzmann gas from the expression for the internal energy given in Eq. 2.20 on p. 57.

The enthalpy is defined as

$$H(S,p,N)=\min_{V}\Big[U(S,V,N)+pV\Big]$$

$$p(S,V,N)=-\frac{\partial U}{\partial V}\overset{\text{Eq. 3.12}}{=}\frac{2U}{3V}$$

We use it first to replace $U=\frac{3}{2}pV$, so that $H=U+pV=\frac{5}{2}pV$.

Then we come back to our original goal, namely to express the volume by (S, p, N)

$$p(S,V,N) = \frac{2U}{3V} \overset{\text{Eq. 3.12}}{=} \frac{2}{3V}\frac{3\pi\hbar^2}{m}N^{\frac{5}{3}}V^{-\frac{2}{3}}\mathrm{e}^{\frac{2S}{3Nk_B}-\frac{5}{3}}$$
$$= \frac{2\pi\hbar^2}{m}N^{\frac{5}{3}}V^{-\frac{5}{3}}\mathrm{e}^{\frac{2S}{3Nk_B}-\frac{5}{3}}$$

Next we resolve this expression for the volume

$$\Rightarrow \qquad V(S,p,N) = \left[\frac{2\pi\hbar^2}{m}N^{\frac{5}{3}}p^{-1}\mathrm{e}^{\frac{2S}{3Nk_B}-\frac{5}{3}}\right]^{\frac{3}{5}}$$
$$= Np^{-\frac{3}{5}}\left[\frac{2\pi\hbar^2}{m}\right]^{\frac{3}{5}}\mathrm{e}^{\frac{2S}{5Nk_B}-1}$$

Now we insert this expression for the volume into the epression for the enthalpy.

$$H(S,p,N) = U(S,V^{(0)},N) + pV^{(0)} = \frac{5}{2}pV^{(0)}$$
$$= \frac{5}{2}p\,\underbrace{Np^{-\frac{3}{5}}\left[\frac{2\pi\hbar^2}{m}\right]^{\frac{3}{5}}\mathrm{e}^{\frac{2S}{5Nk_B}-1}}_{V}$$
$$= \frac{5}{2}Np^{\frac{2}{5}}\left[\frac{2\pi\hbar^2}{m}\right]^{\frac{3}{5}}\mathrm{e}^{\frac{2S}{5Nk_B}-1}$$

ENTHALPY OF THE BOLTZMANN GAS

$$H(S,p,N) = \frac{5}{2}Np^{\frac{2}{5}}\left[\frac{2\pi\hbar^2}{m}\right]^{\frac{3}{5}}\mathrm{e}^{\frac{2S}{5Nk_B}-1} \tag{3.19}$$

3.5.4 Thermal contact and particle exchange: Grand canonical potential

$$\Phi(T,V,\mu) = \min_{S,N}\Big[U(S,V,N) - TS - \mu N\Big]$$

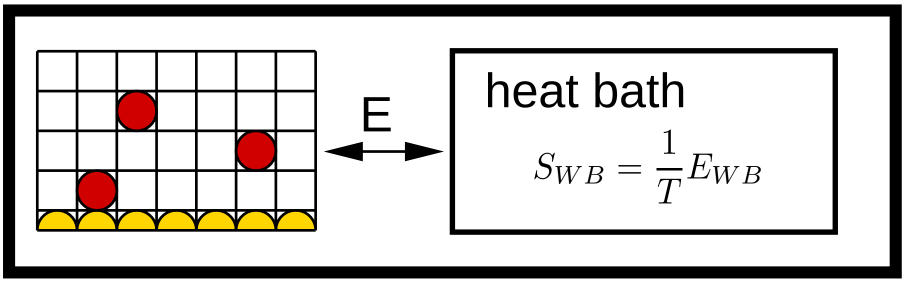
The grand-canonical potential of the Boltzmann gas has already been derived earlier in Eq. 2.11 on p. 55.

GRAND-CANONICAL POTENTIAL OF THE BOLTZMANN GAS

$$\Phi(T,V,\mu) \overset{\text{Eq. 2.11}}{=} -k_BT\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}V\mathrm{e}^{\frac{\mu}{k_BT}} \tag{3.20}$$

3.6 Thermal contact

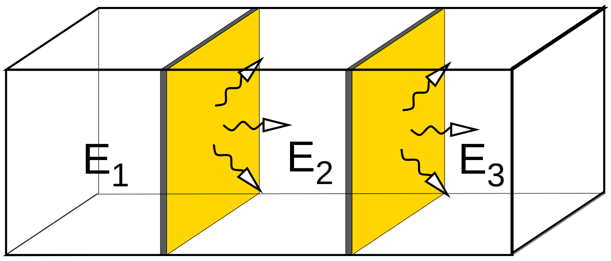
The first thermodynamic process, we will investigate is to control the temperature of a system. We can change the internal energy of a system by heating it up, that is by bringing it into contact with another, hotter system. We will now investigate this process in some detail.



The thermodynamic potential to be used with thermal contact is the Helmholtz potential

$$A(T, V, N) = \min_S U(S, V, N) - TS$$

The Helmholtz potential is the energy of the system and that of the environment. If we apply work on a system, it may warm up. However, because it is in contact with a heat bath, some of the energy is transferred into the heat bath. This imples that we have to do more work on the system than if it were isolated. The change of the Helmholtz potential describes the amount of work I have to do to change the state, considering that some of the energy is lost to the environment.



3.6.1 Thermal contact microscopically

Up to now we have used the formalism of thermodynamics, which is a macroscopic description. We may have lost sight of the underlying microscopic picture.

Microscopic system

Let me illustrate the principle on a very simple example, namely two harmonic oscillators which can exchange energy in a random manner. In each transition the total energy is conserved. We describe the initial state of the two harmonic oscillators by the two quantum numbers (n_1, n_2). The quantum numbers can assume any non-negative integer value, so that the ground state is characterized by $n = 0$. The energy of the "hot" harmonic oscillator is $E_1 = \hbar\omega(n_1 + \frac{1}{2})$ and that of the "cold" harmonic oscillator $E_2 = \hbar\omega(n_2 + \frac{1}{2})$.

The two systems can now exchange energy. The possible transitions lead to the states

Transition	$\Delta E(n_1, n_2)$
$(n_1, n_2) \rightarrow (0, n_1 + n_2)$	$n_1\hbar\omega$
$(n_1, n_2) \rightarrow (1, n_1 + n_2 - 1)$	$(n_1 - 1)\hbar\omega$
$(n_1, n_2) \rightarrow (2, n_1 + n_2 - 2)$	$(n_1 - 2)\hbar\omega$
$\vdots$	$\vdots$
$(n_1, n_2) \rightarrow (n_1 + n_2 - 1, 1)$	$-(n_2 - 1)\hbar\omega$
$(n_1, n_2) \rightarrow (n_1 + n_2, 0)$	$-n_2\hbar\omega$

h

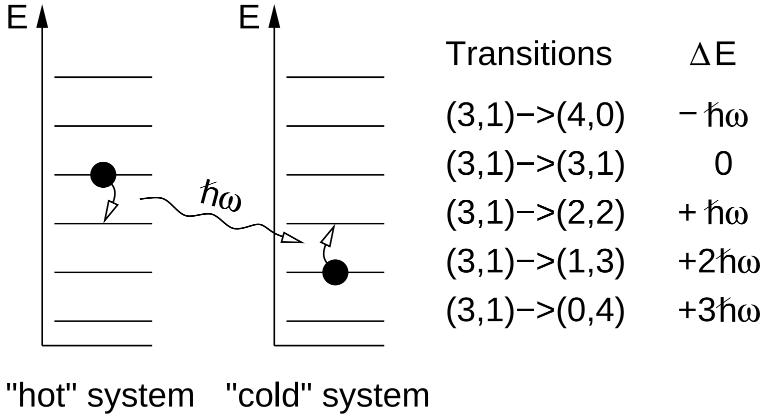


Fig. 3.4: Schematic diagram of two harmonic oscillators in thermal contact. It is shown that on average a random energy exchange will transfer energy from the hot to the cold system. Initially the first (hot) harmonic oscillator is in the 3rd excited state and the second (cold) harmonic oscillator is in the first excited state. The initial state of the combined system is identified by the quantum numbers $(3,1)$. Transitions must conserve the energy and hence the sum of excitations in the final state must be again equal to $3+1=4$. Averaging over all possible transitions we find that, on average, the energy of $1\hbar\omega$ is transferred from the "hot" to the "cold" system.

We see immediately that the mean energy transfer $\langle\Delta E\rangle_{n_1,n_2}$, for all possible transitions starting from the same initial state $(n_1, n_2$, is [3]

$$\langle\Delta E\rangle_{n_1,n_2}=\frac{n_1-n_2}{2}\hbar\omega\ , \tag{3.21}$$

where n_1 characterizes the hot system and n_2 with $n_2<n_1$ characterizes the cold system. The mean energy transfer is always from the higher excited system to the less energetic system. Thus the energy transfer has a well defined direction. The energy is transferred from the hot to the cold system. The only assumption was that all transitions are equally likely!

The terms "hot" and "cold" should not be applied to individual states, so-called **microstates**, but only to ensembles for which we can define a temperature. In order to obtain the ensemble average we need to average the mean energy transfer Eq. 3.21 with probabilities for the initial states, that are derived for two separate systems of which one is coupled to a heat bath at a higher temperature T_h and the other is coupled to a heat bath at a lower temperature T_c.

After contact with a heat bath at temperature T_h, the probabilties of the "hot" subsystem is

$$P^{(1)}_{n_1}(T_h)=\frac{1}{Z(T_h)}\mathrm{e}^{-\frac{1}{k_BT_h}\hbar\omega(n_1+\frac{1}{2})}$$

and those for the cold system, which has been in contact with a heat bath at temperature T_c, the

---

[3] For a given initial state characterized by the two excitation numbers n_1, n_2, the final states are $(0, n_1+n_2), (1, n_1+n_2-1), \dots (n_1+n_2, 0)$ or $\{(j, n_1+n_2-j)\quad \text{for}\quad j=0,\dots,n_1+n_2\}$. For a given final state, i.e. for a given value of j, the energy transfer from the left to the right system is equal to the energy loss of the left system. i.e $-\Big(\hbar\omega(j+\frac{1}{2})-\hbar\omega(n_1+\frac{1}{2})\Big)=\hbar\omega(n_1-j)$. Thus the mean energy transfer is

$$\langle\Delta E\rangle_{n_1,n_2}=\frac{1}{n_1+n_2+1}\sum_{j=0}^{n_1+n_2}\hbar\omega(n_1-j)=\frac{1}{2}\Big(\hbar\omega n_1-\hbar\omega n_2\Big)$$

In the last step, which brought us to the final result Eq. 3.21, I simply exploited that the summands are equispaced which allows me to take the mean value of the first and the last summand.

probabilities for the cold system are

$$P^{(2)}_{n_2}(T_c) = \frac{1}{Z(T_c)} \mathrm{e}^{-\frac{1}{k_B T_c}\hbar\omega(n_2+\frac{1}{2})}$$

Because the two systems are initially fully separated, they are uncorrelated so that the probabilities for the entire system can be written as a product of the probabilities of the two subsystems.

$$P^{(1,2)}_{n_1,n_2}(T_h,T_c) = P^{(1)}_{n_1}(T_h)P^{(2)}_{n_2}(T_c) = \frac{1}{Z(T_h)Z(T_c)} \mathrm{e}^{-\frac{1}{k_B T_h}\hbar\omega(n_1+\frac{1}{2})}\mathrm{e}^{-\frac{1}{k_B T_c}\hbar\omega(n_2+\frac{1}{2})} \tag{3.22}$$

The mean energy transfer in one transition is just obtained by weighting the mean energy transfer for one configuration (n_1, n_2) over all initial states. The result can be expressed as one-half of the difference of the internal energy.

$$\begin{aligned}
\langle \Delta E\rangle_{T_h,T_c} &= \sum_{n_1,n_2=0}^{\infty} P^{(1,2)}_{n_1,n_2}(T_h,T_c)\langle\Delta E\rangle_{n_1,n_2} \\
&\overset{\text{Eqs. 3.22 ,3.21}}{=} \sum_{n_1,n_2=0}^{\infty} P^{(1)}_{n_1}(T_h)P^{(2)}_{n_2}(T_c)\frac{n_1-n_2}{2}\hbar\omega \\
&= \sum_{n_1,n_2=0}^{\infty} P^{(1)}_{n_1}(T_h)P^{(2)}_{n_2}(T_c)\cdot\frac{1}{2}\Big(\underbrace{\hbar\omega(n_1+\frac{1}{2})}_{E^{(1)}_{n_1}} - \underbrace{\hbar\omega(n_2+\frac{1}{2})}_{E^{(2)}_{n_2}}\Big) \\
&= \underbrace{\Big(\sum_{n_1=0}^{\infty} P^{(1)}_{n_1}(T_h)E^{(1)}_{n_1}\Big)}_{U_1}\underbrace{\Big(\sum_{n_2=0}^{\infty} P^{(2)}_{n_2}(T_c)\Big)}_{=1} - \underbrace{\Big(\sum_{n_1=0}^{\infty} P^{(1)}_{n_1}(T_h)\Big)}_{=1}\underbrace{\Big(\sum_{n_2=0}^{\infty} P^{(2)}_{n_2}(T_c)E^{(2)}_{n_2}\Big)}_{U_2} \\
&= \frac{U_1-U_2}{2}
\end{aligned}$$

Because the internal energy of a harmonic oscillator increases with temperature, the energy is transferred from the hot to the cold system and the energy transfer grows with the temperature difference of the two systems.

The important aspect that is to be learned that there is no mysterious intrinsic driving force forces the energy from the hot to the cold system. In constrast, the heat transfer is a result of pure chance.

Macroscopic system

Editor: This may be better together with H-Theorem and irreversibility

Therefore, let us dig even a little deeper and investigate the processes on a quantum mechanical level. For this discussion we limit ourselves to a container with only two compartments, denoted by A and B. Let us consider that the gas in each compartment is in an well defined energy eigenstate characterized by their quantum numbers $\vec{n}_A$ and $\vec{n}_B$. Once we bring the containers into contact, the wave functions of the different containers have a small overlap, because each wave function extends a small distance into into the wall.[4] Hence the gas is not in an eigenstate of the combined system. The wave function can still be expressed by a superposition of the energy eigenstates of the isolated systems as

$$|\Psi^{AB}(t)\rangle = \sum_{\vec{n}_A,\vec{n}_B} |\vec{n}_A\rangle|\vec{n}_B\rangle c_{\vec{n}_A,\vec{n}_B}(t)$$

---

[4] A more realistic description would be to involve the states of the wall as a separate system. Then the gas molecules of each container interact with the states of the wall, which results in an effective coupling of the gas in the two containers.

Here $|\vec{n}_A\rangle$ are the eigenstates of the first compartment $|\vec{n}_B\rangle$ are those of the second compartment. In real space, the wave function of the combined system would depend on the coordinates of all particles in A and in B. The basis functions are product wave functions of the individual systems.

$$|\vec{n}_A\rangle|\vec{n}_B\rangle\hat{=}\Phi_{\vec{n}_A,\vec{n}_B}(\vec{r}^A_1,\dots,\vec{r}^A_{N_A},\vec{r}^B_1,\dots,\vec{r}^B_{N_B})=\phi_{\vec{n}_A}(\vec{r}^A_1,\dots,\vec{r}^A_{N_A})\psi_{\vec{n}_B}(\vec{r}^B_1,\dots,\vec{r}^B_{N_B})$$

Before the systems come into contact, these states are energy-eigenstates of the combined system with eigenvalues

$$E_{\vec{n}_A,\vec{n}_B}=E^A_{\vec{n}_A}+E^B_{\vec{n}_B}$$

Once the systems come into contact, however, the states are no more energy-eigenstates and their weights will change over time. Still, the product states still serve as basis functions.

The coefficients $c_{\vec{n}_A,\vec{n}_B}$ are time dependent and their dynamics is determined by the time-dependent Schrödinger equation. The changes of the coefficients can be described by the transitions between the eigenstates of the isolated compartments. We can evaluate the probability to find the gas of the first container[5] in a particular state $\vec{n}_A$ of subsystem A. We project onto all states for which system A is in state $\vec{n}_A$, which implies that we "trace out" the system B,i.e. we form the trace over all states in the sub-Hilberspace of subsystem B. Here we use the notation $|\vec{n}_A,\vec{n}_B\rangle=|\vec{n}_A\rangle|\vec{n}_B\rangle$.

$$\begin{aligned}P_{A,\vec{n}_A}(t)&=\langle\Psi^{AB}(t)|\Big(\sum_{\vec{n}_B}\vec{n}_A,\vec{n}_B\rangle\langle\vec{n}_A,\vec{n}_B|\Big)|\Psi^{AB}(t)\rangle=\sum_{\vec{n}_B}c^*_{\vec{n}_A,\vec{n}_B}(t)c_{\vec{n}_A,\vec{n}_B}(t)\\&=\int d^{3N_A}r_A\int d^{3N_A}r'_A\int d^{3N_B}r_B\;\Psi^{AB*}(\vec{r}_A,\vec{r}_B,t)\phi_{\vec{n}_A}(\vec{r}_A)\phi^*_{\vec{n}_A}(\vec{r'}_A)\Psi^{AB}(\vec{r'}_A,\vec{r}_B,t)\end{aligned}$$

So far we considered the evolution of a single energy state. In order to obtain the desired result, we need to average about the initial conditions $c_{\vec{n}_A,\vec{n}_B}(t=0)$ with the corresponding probabilities. Note that the probabilities do not determine the phases of the wave functions. Thus one needs to **average over all relative phases** of the wave functions in the two containers.

As the containers are brought into contact, the gas in one container undergoes transitions from one eigenstate to another, therefore changing the energy in this container. Because the energy of both containers must be conserved, each transition must be accompanied by a transition in the same or in the other container, which balances the energy. If one container is in a state of high energy, implying that it is hot, it will more likely undergo a transition to a state of lower energy, that is it will cool down, while raising a state in the other container towards higher energy, thus heating the other container.

The dynamics of the coefficients may be very complicated and without an accurate description of the coupling, we cannot work it out. The best assumption[6] is that the system ends up with equal probability in all states that are compatible with the conservation laws. This brings us back to the maximum entropy principle with the constraint that the energy of the system is known.

This is an important concept! The argument that we do not know any better how the system behaves does not mean that the real dynamics shows this behavior. Thus the connection of the subjective argument, the maximum-entropy principle, and the objective basis, which can be tested experimentally is still open. The connection is the **ergodic principle**.

Let us now consider the change of the entropy in the first container: What changes as the two systems come into contact? The energy levels of the two subsystems remain the same. Only the probabilities for the individual states change. If the probabilities change, also the entropy $S=\sum_{\vec{n}}P_{\vec{n}}\ln[P_{\vec{n}}]$ changes. We assume that each system remains in thermal equilibrium as the contact is

---

[5] Note that the in the following expression one needs to simply integrate out all the degrees of freedom of the second system.

[6] This argument implies that we are guessing and that our guesses have nothing to do with reality and the precision of the prediction of statistical physics. There is an argument lying behind, which only becomes only evident, if one considers really large numbers of states. In that case nearly all conceivable initial states would result in probability distributions which are **very** similar to a completely random distribution. Here we do not want to go into that detail.

established. This in turn implies that the transitions between eigenstates of one container are more frequent when they exchange energy with transitions in the same container than if they exchange energy with the other container. As a result of this assumption the probabilities are consistent with the instantaneous energy of the system, that is $P_{\vec{n}}(t) = \mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}-A(T(t)))}$.

The differential change of the entropy is

$$\begin{aligned}
S &= -k_B \sum_{\vec{n}} P_{\vec{n}} \ln[P_{\vec{n}}] && (3.23)\\
dS &= -k_B \sum_{\vec{n}} \left(dP_{\vec{n}} \ln[P_{\vec{n}}] + P_{\vec{n}} \frac{dP_{\vec{n}}}{P_{\vec{n}}} \right) = -k_B \sum_{\vec{n}} \left(\ln[P_{\vec{n}}] + 1 \right) dP_{\vec{n}} \\
&\overset{Eq.\ 1.55}{=} -k_B \sum_{\vec{n}} \left(\ln[\mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}-A(T))}] + 1 \right) dP_{\vec{n}} \\
&= \frac{1}{T} \sum_{\vec{n}} \left(E_{\vec{n}} - A(T) - k_BT \right) dP_{\vec{n}} \\
&= \underbrace{\frac{1}{T} \left(\sum_{\vec{n}} E_{\vec{n}} dP_{\vec{n}} \right)}_{=dU} - \frac{A(T) + k_BT}{T} \underbrace{\left(\sum_{\vec{n}} dP_{\vec{n}} \right)}_{=0} = \frac{1}{T} dU \\
\Rightarrow dU &= T dS && (3.24)
\end{aligned}$$

Eq. 3.24, which is identical to Eq. 1.73 in the absence of work, describes how the energy of the system changes as it is heated up. The change of the energy can be written as

$$dU = T dS$$

The change of energy dU that flows from one system to the other due to a change in temperature is called **Heat**[7]. It is often denoted by the symbol Q and the differential change is $dQ = TdS$. When a system heats up, its entropy grows. If the temperature is low, only the low-energy states occur with large probability. If the entropy increases, the probability of the low-energy states decreases and that of the high energy states increases. Thus the internal energy of the system increases when it is heated up.

Example: We can now evaluate the heat transfer for the Boltzmann gas using its Helmholtz potential from Eq. 3.17. We consider here the heat transfer ΔQ into one compartment, which raises the temperature from its initial value T_i to its final value T_f.

$$\begin{aligned}
\Delta Q &= \int_{S_i}^{S_f} T dS = \int_{T_i}^{T_f} T \frac{dS^{BG}}{dT} dT \overset{S=-\frac{\partial A}{\partial T}}{=} -\int_{T_i}^{T_f} T \frac{d^2A^{BG}}{dT^2} dT \\
\frac{d^2A^{BG}}{dT^2} &\overset{Eq.\ 3.17}{=} \frac{\partial^2}{\partial T^2} \left\{ Nk_BT \left(\ln\left[\frac{\lambda_T^3 N}{V} \right] - 1 \right) \right\} \\
&= \frac{\partial}{\partial T} \left\{ Nk_B \left(\ln\left[\frac{\lambda_T^3 N}{V} \right] - 1 \right) + Nk_BT \left(-\frac{3}{2T} \right) \right\} \\
&= -\frac{3Nk_B}{2T} \\
\Delta Q &= \int_{T_i}^{T_f} T \cdot \frac{3Nk_B}{2T} dT = \frac{3}{2} Nk_BT_f - \frac{3}{2} Nk_BT_i \overset{Eq.\ 2.17}{=} U_f - U_i
\end{aligned}$$

It is just the difference of the internal energy as anticipated.

---

[7] Heat is called Wärme in German

Minimum principle for the Helmholtz free energy

We have found that systems in thermal equilibrium approach a state of minimum free energy. We have, however required that the subsystems themselves are internally in equilibrium. Here we show that the minimum-free-energy principle is actually equivalent with the maximum-entropy principle in its most general form.

We consider one system in contact with a heat bath characterized by T_{hb}. The entropy is defined for arbitrary probabilities as

$$\begin{aligned}\mathcal{S}(P_{\vec{n}},T_{hb}) &= -k_B\sum_{\vec{n}} P_{\vec{n}}\ln[P_{\vec{n}}] + \underbrace{\frac{1}{T_{hb}}U_{hb}}_{S_{hb}}\\ &\quad -\lambda\left(\sum_{\vec{n}} P_{\vec{n}} - 1\right) - \frac{1}{T}\left(\sum_{\vec{n}} P_{\vec{n}}E_{\vec{n}} + U_{hb} - U_{tot}\right)\end{aligned}$$

where we introduced the constraints for normalization and energy conservation. Note that T is only a Lagrange multiplier. We can satisfy the energy conservation constraint already by resolving the constraint for U_{hb}, namely $U_{hb} = U_{tot} - \sum_{\vec{n}} P_{\vec{n}}E_{\vec{n}}$, and by inserting the corresponding expression.

$$\begin{aligned}\mathcal{S}(P_{\vec{n}},T_{hb}) &= -k_B\sum_{\vec{n}} P_{\vec{n}}\ln[P_{\vec{n}}] - \frac{1}{T_{hb}}\sum_{\vec{n}} P_{\vec{n}}E_{\vec{n}} + \frac{1}{T_{hb}}U_{tot} - \lambda\left(\sum_{\vec{n}} P_{\vec{n}} - 1\right)\\ &= -\frac{1}{T_{hb}}\left[\underbrace{\sum_{\vec{n}} P_{\vec{n}}E_{\vec{n}} + k_BT_{hb}\sum_{\vec{n}} P_{\vec{n}}\ln[P_{\vec{n}}]}_{\mathcal{A}(P_{\vec{n}},T)} - U_{tot} + \lambda T_{hb}\left(\sum_{\vec{n}} P_{\vec{n}} - 1\right)\right]\end{aligned}$$

Thus we define a Helmholtz free energy as function of the probabilities

$$\mathcal{A}(P_{\vec{n}},T) \stackrel{\text{def}}{=} \underbrace{\sum_{\vec{n}} P_{\vec{n}}E_{\vec{n}}}_{U} + \underbrace{k_BT\sum_{\vec{n}} P_{\vec{n}}\ln[P_{\vec{n}}]}_{-TS} = \sum_{\vec{n}} P_{\vec{n}}\left(E_{\vec{n}} + k_BT\ln[P_{\vec{n}}]\right)$$

Since T_{hb} and U_{tot} are constants, the distribution that maximizes the entropy is the same that minimizes the free energy under the constraint that the probabilities are normalized.

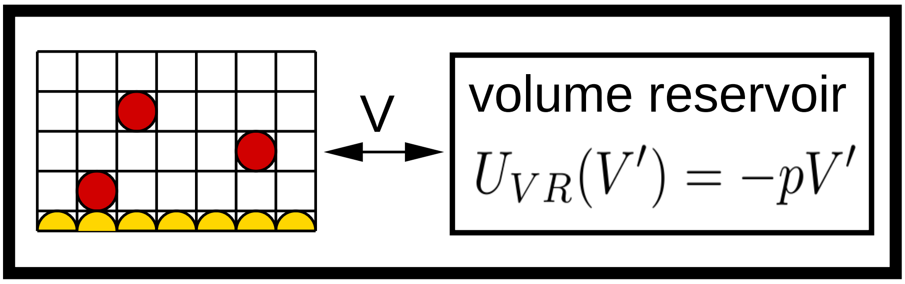
Thus we obtain the minimum entropy principle, which says that **the probabilities of a system in contact with a heat bath, which is characterized by a temperature T_{hb}, minimizes the Helmholtz free energy $\mathcal{A}(P_{\vec{n}},T_{hb})$ under the constraint that the probabilities add up to one.** The reader can convince himself that the probabilities are

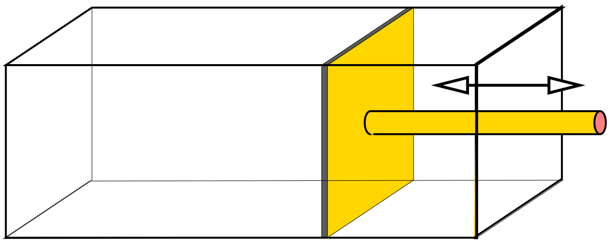
$$P_{\vec{n}} = \frac{1}{Z(T_{hb})}\mathrm{e}^{-\frac{1}{k_BT_{hb}}E_{\vec{n}}}$$

3.7 Mechanical contact

We have seen in the last section that we can change the energy of a system by heating it up. If we want to build an engine, we need to understand how to put this energy into use: We need to extract mechanical energy.

Let us consider here a gas-container with a side-wall replaced by a movable piston. The gas container shall be thermally isolated, so that no heat can flow in and out of the system. Thus $dQ = TdS = 0$. This implies that the entropy of the container is constant. So far, the driving force that brings a system into thermal equilibrium has been the entropy increase. Exactly that driving force shall vanish. Indeed this implies that we cannot apply the maximum entropy principle. Therefore we discuss here a quasi-static process, where the state of the system is dictated externally by moving the piston.





If the gas exerts[8] a force onto the piston, we can let it do work. **Work** is the mechanical energy that is transferred from the gas to a mechanical system, that is

$$\Delta E_{mech} = -dU = +pdV$$

The mechanical system can be a spring, for example, but it may as well a system of gears, that directs the work to the wheels of a vehicle.

Mechanical contact microscopically

Let us consider the change of internal energy in thermal equilibrium.

$$U = \sum_{\vec{n}} P_{\vec{n}} E_{\vec{n}}(V)$$

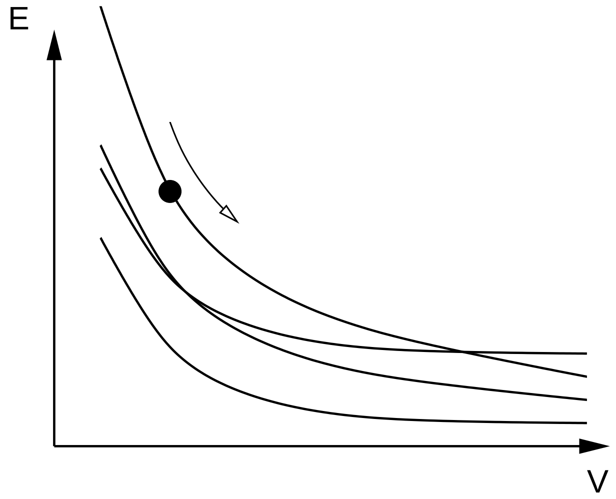
A change of the state may change the probabilities and the energy levels.

$$dU = \underbrace{\sum_{\vec{n}} dP_{\vec{n}} E_{\vec{n}}(V)}_{TdS} + \underbrace{\sum_{\vec{n}} P_{\vec{n}} dE_{\vec{n}}}_{dW}$$

The first term TdS is the heat transported into the system and the second term is the work done on the system. If we require that no heat is transferred, the first term vanishes. Note, that heat is a form of energy, that obeys a conservation law. Unlike the entropy, energy cannot be created or destroyed. Excluding the probabilities does not necessarily mean that the probabilities do not change.

---

[8] The German word for "exert" is "ausüben".



If we consider a process where the the volume of the container changes, also the energy levels shift $\frac{dE_{\vec{n}}}{dt}=\frac{dE_{\vec{n}}}{dV}\frac{dV}{dt}$. Let us consider that the gas is in a well-defined eigenstate of the Hamiltonian, with quantum numbers $\vec{n}$. As the volume changes, also the energy level shift. The reader may confirm that the one-particle levels explicitly depend on the size and shape of the box by comparing with Eq. 2.3. The total energy of the gas is $E_{\vec{n}}(L_x)$. Thus the force, which the gas exerts on the container, is $F_x=-\frac{dE_{\vec{n}}}{dL_x}$. Now we average the result using the probabilities and obtain

$$F_x=-\sum_{\vec{n}}P_{\vec{n}}\frac{dE_{\vec{n}}}{dL_x}=-A_{yz}\sum_{\vec{n}}P_{\vec{n}}\frac{dE_{\vec{n}}}{dV}$$
$$\Rightarrow p=\frac{F_x}{A_{yz}}=-\sum_{\vec{n}}P_{\vec{n}}\frac{dE_{\vec{n}}}{dV}$$

where A_{yz} is the area of the piston. For the new position of the piston, the probabilities do not fulfill the maximum-entropy principle any more. Transitions among the eigenstates of the system result in new probabilities. Hence the probabilities change with volume. We can divide the infinitesimal process in two separated ones.

1. The first step preserves the probabilities, but changes the energy levels in response to the change in the volume. As the probabilities remain unaffected, this step does not contribute to the entropy.

2. The second step, the energy-levels remain fixed, but the probabilities change from

   $$P_{\vec{n}}(T,V)=\mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}(V)-A(T,V))}$$

   where $A(T,V)$ is the Helmholtz potential, to

   $$P_{\vec{n}}(T+dT,V+dV)=\mathrm{e}^{-\frac{1}{k_B(T+dT)}(E_{\vec{n}}(V+dV)-A(T+dT,V+dV))}$$

   so that

   $$dP_{\vec{n}}=\frac{\partial P_{\vec{n}}(T,V)}{\partial T}dT+\frac{\partial P_{\vec{n}}(T,V)}{\partial V}dV$$

   The changes of the probabilities correspond to transitions between the states. Since the system is thermally isolated, the energy during these transition must be conserved. As a result, $TdS=$

0.

$$
\begin{aligned}
TdS &= \sum_{\vec{n}} E_{\vec{n}} dP_{\vec{n}} = 0 \\
\Rightarrow 0 &= \underbrace{\left(\sum_{\vec{n}} E_{\vec{n}} \frac{\partial P_{\vec{n}}(T,V)}{\partial T}\right)}_{\frac{dU}{dT}|_{V,N}} dT + \left(\sum_{\vec{n}} E_{\vec{n}} \frac{\partial P_{\vec{n}}}{\partial V}\right) dV \\
dT &= -\frac{\sum_{\vec{n}} E_{\vec{n}} \frac{\partial P_{\vec{n}}}{\partial V}}{\sum_{\vec{n}} E_{\vec{n}} \frac{\partial P_{\vec{n}}(T)}{\partial T}} dV
\end{aligned}
$$

Thus the temperature changes in an adiabatic process.

Application to the ideal gas

Now we can apply our findings to the Boltzmann gas. Since we consider an isentropic process, we can use the expression of the internal energy Eq. 2.20 as function of S, V, N. This expression is repeated here:

$$
U^{BG}(S,V,N) \overset{\text{Eq. } 2.20}{=} \frac{N^{\frac{5}{3}}}{V^{\frac{2}{3}}} \left(\frac{3\pi\hbar^2}{m}\right) \mathrm{e}^{\frac{2S}{3Nk_B}-\frac{5}{3}}
$$

Thus we obtain the pressure as function of the volume for an adiabatic process as

$$
\begin{aligned}
p^{BG}(S,V,N) &= -\left.\frac{dU^{BG}}{dV}\right|_{S,N} = -\frac{2}{3V} U^{BG}(S,V,N) \\
&= -\frac{2}{3}\left(\frac{N}{V}\right)^{\frac{5}{3}} \left(\frac{3\pi\hbar^2}{m}\right) \mathrm{e}^{\frac{2S}{3Nk_B}-\frac{5}{3}}
\end{aligned}
$$

As the volume changes the temperature behaves as

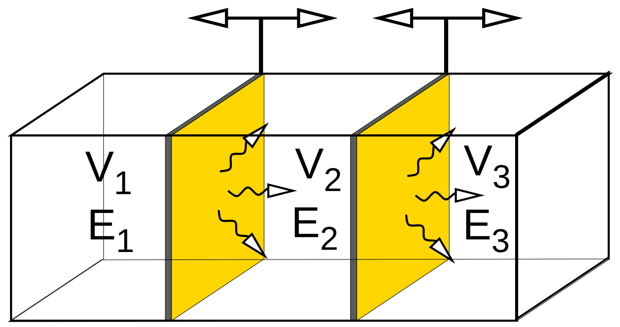
$$
\begin{aligned}
T^{BG}(S,V,N) &= +\left.\frac{dU^{BG}}{dS}\right|_{V,N} = \frac{2}{3Nk_B} U^{BG}(S,V,N) \\
&= \frac{2}{3k_B}\left(\frac{N}{V}\right)^{\frac{2}{3}} \left(\frac{3\pi\hbar^2}{m}\right) \mathrm{e}^{\frac{2S}{3Nk_B}-\frac{5}{3}}
\end{aligned}
$$

We see that as we compress a gas, it warms up, while it cools, when we expand the gas. This principle us to build thermal engines or build a refrigerator. It can be easily verified using a bycicle pump by compressing it, while closing the outlet. The pump will become warm.

The behavior

$$
\begin{aligned}
T^{BG}(V) &\sim V^{-\frac{2}{3}} \\
p^{BG}(V) &\sim V^{-\frac{5}{3}}
\end{aligned}
$$

obtained here, can be tested experimentally.



3.7.1 Particle exchange microscopically

Let us consider two systems that can exchange particles. The eigenstates can be composed out of the eigenstates of the individual compartments, but now an N particle state of the combined system can be set up by states that have M particles in one compartment and $N-M$ particles in the other compartment. The coupling allows transitions between states with different M. Such a transition describes the motion of one particle through the permeable wall. If we describe only one compartment, we need to allow states with all possible number of particles.

3.8 Chemical equilibrium

Let us consider a chemical reaction at constant temperature and pressure. In this case we can make use of the minimum principle for the Gibbs free energy $G(T, p, N_1, \ldots, N_m)$.

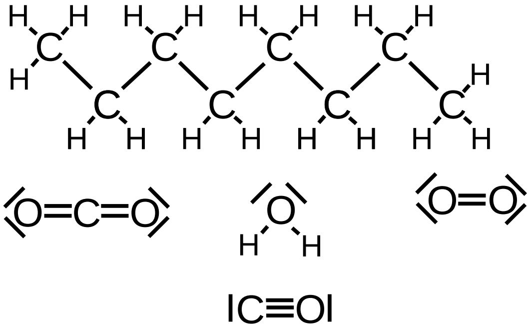


Fig. 3.5: Reactants and products of the gasoline burning reaction. Octane C_8H_{18}, oxygen (O_2), carbon dioxide (CO_2), water (H_2O) and carbon monoxide (CO).

Now we consider a chemical reaction such as

$$\begin{aligned}2C_8H_{18} + 25O_2 &\leftrightarrow 16CO_2 + 18H_2O\\ 2CO + O_2 &\leftrightarrow 2CO_2\end{aligned}$$

The pre factors are called **stoichiometric coefficients** . One can check the consistency of a reaction equation by ensuring that the same number of atoms for each element occur on both sides of the reaction.

The state of the system is characterized by the number N_j of molecules that are present in the reaction, respectively their concentrations. We can combine the particle numbers into a vector $\vec{N}$ of length m. In the above example the vector would have the following form.

$$\vec{N} = (N_{C_8H_{18}}, N_{O_2}, N_{CO_2}, N_{H_2O}, N_{CO})$$

The dimension of the concentration space is $m = 5$.

Each reaction equation determines a direction $\vec{\nu}_k$ in the m-dimensional concentration space. These vectors describe how the particle numbers change if the reaction proceeds once towards the right-hand side of the reaction equation. If there are t reaction equations, one will obtain t such vectors.

In our example the vectors $\vec{\nu}_k$ are

$$\begin{aligned}\vec{\nu}_1 &= (-2, -25, 16, 18, 0)\\ \vec{\nu}_2 &= (0, -1, 2, 0, -2)\end{aligned}$$

These directions $\vec{\nu}_k$ are composed of the stoichiometric coefficients $\nu_{j,k}$ of the respective reaction, which are inserted in the correct position of the configuration vector. The numbers on the left-hand side of the reaction are multiplied by -1 because those molecules are consumed if the reaction goes to the right.

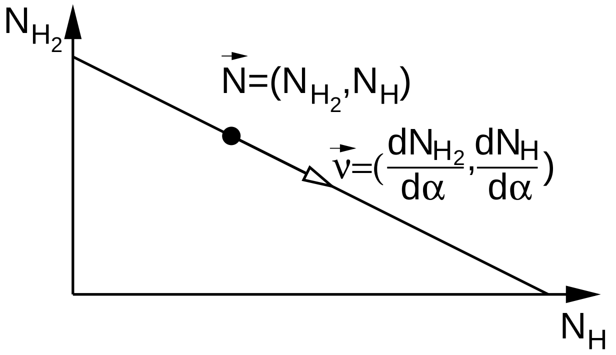


Fig. 3.6: illustration of the stoichiometric coefficients for a simple two component reaction $H_2 \leftrightarrow 2H$.

2

As the reaction proceeds, the configuration vector can always be written as

$$N(\{\alpha_m\}) = \vec{N}_0 + \sum_{k=1}^{t} \vec{\nu}_k \alpha_k$$

The coefficients α_k describe how far the reaction proceeds. We will call them **reaction parameters**.

3.8.1 Partial pressure

We consider a reaction at constant pressure and constant temperature. For an ensemble at constant pressure and temperature the appropriate thermodynamic potential should be the Gibbs energy.

However, it turns out to be more convenient to start out with the Helmholtz potential and convert to the Gibbs energy later on. The reason is the following: The Gibbs energy has been chosen for a situation where several subsystems exchange volume, and where the total volume is the sum of the volumes occupied by the subsystems. That is, each component was enclosed in its own compartment. This does not apply to the problem at hand.

However, here, each component occupies the complete volume and there is no volume exchange. Thus the proper thermodynamic potential is the Helmholtz potential. Only the total volume is related to the overall pressure of the volume reservoir. This can be taken into account by a Legendre transform of the final (total) Helmholtz potential.

Since the components of the gas do not interact with each other, except that they can be converted into each other during a chemical reaction, the total Helmholtz potential is the sum of the Helmholtz potentials of the individual components.

$$A_{tot}(T, V, N_1, N_2, \ldots) = \sum_j A_j(T, V, N_j)$$

We can now determine the total pressure using the equation of state

$$p_{tot} = -\left.\frac{dA_{tot}}{dV}\right|_{T,\vec{N}} = \sum_j \underbrace{-\left.\frac{dA_j}{dV}\right|_{T,N_j}}_{p_j} = \sum_j p_j$$

Thus the total pressure is the sum of the partial pressures of the individual components of the gas mixture.

$$p_{tot}=\sum_j p_j \tag{3.25}$$

where the **partial pressures** of the components are defined as

$$p_j=-\left.\frac{dA_j}{dV}\right|_{T,N_j} \tag{3.26}$$

It may be instructive to see how the Gibbs energy is composed of the Gibbs energies G_j of its components.

$$\begin{aligned}
G(T,p_{tot},\vec{N}) &= \min_V A(T,V,\vec{N})+p_{tot}V \\
&\overset{Eq.\ 3.25}{=} \min_V \sum_j \left(A_j(T,V,N_j)+p_jV\right) \\
&= \sum_j \min_V \left(A_j(T,V,N_j)+p_jV\right) \\
&= \sum_j G_j(T,p_j,N_j)
\end{aligned} \tag{3.27}$$

The fact that each component has to be treated at its own partial pressure is why it is inconvenient to work with Gibbs energies in this case.

3.8.2 Chemical equilibrium

Chemical equilibrium at constant volume

The main goal of this section is to determine the equilibrium concentrations of a chemical reaction. This is obtained by minimizing the total Helmholtz potential with respect to the reaction parameters α_k. If we consider a reaction at fixed volume and constant temperatur we select the Helmholtz potential $A(T,V,\vec{N})$ as the appropriate thermodynamic potential.

$$\begin{aligned}
&\left.\frac{d}{d\alpha_k}\right|_{T,V} A(T,V,\vec{N}(\vec{\alpha}))=0 \\
&\Rightarrow \sum_j \underbrace{\left.\frac{dA_j}{\partial N_j}\right|_{T,V}}_{\mu_j(T,V,N_j)} \underbrace{\frac{dN_j}{d\alpha_k}}_{\nu_{j,k}}=0
\end{aligned}$$

Thus we obtain the equilibrium condition

$$\sum_j \mu_j(T,V,N_j)\nu_{j,k}=0$$

Thus,for each reaction equation, we obtain one equation for the particle numbers N_j. This allows us to determine all reaction parameters, as there is again one for each reaction.

Chemical equilibrium at constant pressure

If we consider a situation at constant pressure, we should minimize the Gibbs energy $G(T,p,\vec{N})$.

$$\frac{d}{\partial\alpha_k}\bigg|_{T,V} G(T,p_{tot},\vec{N}(\vec{\alpha}))=0$$

$$\overset{\text{Eq. 3.27}}{\Rightarrow}\sum_j \underbrace{\frac{dG_j}{dN_j}\bigg|_{T,p_j}}_{\mu_j(T,p_j,N_j)}\underbrace{\frac{dN_j}{d\alpha_k}}_{\nu_{j,k}}+\sum_j\underbrace{\frac{dG_j}{dp_j}\bigg|_{T,p_j}}_{V}\underbrace{\sum_p\frac{\partial p_j(T,p_{tot},\vec{N})}{\partial N_p}\frac{dN_p}{d\alpha_k}}_{\frac{dp_{tot}}{d\alpha_k}}=0$$

$$\Rightarrow\sum_j\mu_j(T,p_j,N_j)\nu_{j,k}+V\underbrace{\frac{\partial p_{tot}}{d\alpha_k}}_{=0}=0$$

Because we require that the total pressure is constant during the reaction in this experiment, we obtain

$$\sum_j\mu_j(T,p_j,N_j)\nu_{j,k}=0 \qquad (3.28)$$

Nevertheless we need to determine the individual partial pressures. For ideal, that is non-interacting, systems there is a unique relation between the concentration of a species and its partial pressure.

General chemical equilibrium

We can generalize the result Eq. 3.28 above analogously to the more general case

CHEMICAL EQUILIBRIUM

In chemical equilibrium the following equations are fulfilled

$$\sum_{j=1}^{m}\mu_j\nu_{j,k}=0 \qquad \text{for} \qquad k=1,\dots,t \qquad (3.29)$$

where m is the number of components (reactands, products,etc.) and t is the number of reaction equations. The constants $\nu_{j,k}$ are the stoichiometric coefficients. The chemical potentials μ_j are derived from the appropriate chemical potential, i.e. $\mu(T,p,N)=\frac{\partial G(T,p,N)}{\partial N}$, $\mu(T,V,N)=\frac{\partial A(T,V,N)}{\partial N}$, $\mu(S,p,N)=\frac{\partial H(S,p,N)}{\partial N}$, $\mu(S,V,N)=\frac{\partial U(S,V,N)}{\partial N}$. The volume is always taken to be the total volume, while the pressure is the appropriate partial pressure of that component.

3.8.3 Internal molecular degrees of freedom of a gas

Before we bring the above equilibrium equations into the form of the law of mass action, let us investigate the form of the Helmholtz potential for real gases.

The law of mass action applies to all non-interacting gases. If gases are dilute, their interaction can be ignored. Without showing this here it can even be generalized to dilute solutions.

Let us consider here the free energy of a gas: The free energy of real molecules can be divided into individual energy contributions, just like the Hamiltonian. [9]

---

[9] This decomposition is valid only for sufficient low temperatures. For example, an-harmonic terms in the vibrations lead to a coupling between vibrational states which are relevant at high temperatures.

- **Ground state energy** E_0 without zero-point vibrations.
- **Translational energy**, E_{trans} which is simply the kinetic energy related to the center of gravity of the molecule. The contribution to the Hamiltonian is $\frac{\vec{P}^2}{2M}$, where $\vec{P}$ is the total momentum of a molecule and M is the total mass. Note that this term is exactly the one treated in the Boltzmann gas.
- **Rotational energy**. The energy of the rotations of an almost spherical molecule can be written as $E_{rot}(\ell, m) = \frac{\hbar^2\ell(\ell+1)}{I}$ where ℓ is the main angular-momentum quantum number. and I is the moment of inertia[10]. In the sum over states it is important to consider the $2\ell + 1$ degeneracy of the rotational levels. Note that the energy spectrum for a non-isotropic moment of inertia is quite complex. For molecules without even an axial symmetry the eigenstates can no more be labeled by ℓ, m. Atoms do not have a rotational energy. For linear molecules there is only the main quantum number ℓ.
- **Vibrational energy**. The vibrations can be decomposed into independent vibrational modes, where each mode is represented by a harmonic oscillator. $E_{vib}(n_j) = \sum_{j=1}^{3q-6} \hbar\omega_j(n_j + \frac{1}{2})$. With q we denote the number of atoms in the molecule. ω_j are the frequencies of the vibrational modes. We subtracted 6 degrees of freedom from the sum, which correspond to the three translational and the three rotational degrees of freedom. For atoms we must only substract the 3 translational degrees of freedom instead of 6 and for linear molecules we need to subtract three translational and two rotational degree of freedom.
- **Electronic excitations** E_e. In most cases electronic excitations play a minor role: If the excitation energy is larger than kT, excitations are so rare that they can be ignored. The situation is of course different if the highest occupied and the lowest unoccupied states are degenerate or nearly degenerate.

The total energy is the sum of these terms. We call all terms except the translational term the internal degrees of freedom. In the partition sum for the gas we obtain

$$A(T, V, N) = -k_B T \ln \left[\sum_{\vec{n}} \mathrm{e}^{-\frac{1}{k_B T} \sum_{j=1}^{N} \left[E_{0,j}+E_{trans,j,\vec{n}}+E_{rot,j,\vec{n}}+E_{vib,j,\vec{n}}+E_{e,j,\vec{n}}\right]} \right]$$

where the set of quantum numbers $\vec{n}$ includes the translational quantum numbers of the particle in the box, the rotational quantum numbers of all molecules, vibrational and electronic excitations of all molecules. Each energy contributions depends on the set of quantum numbers

We can divide the exponential into individual factors.

$$A(T, V, N) = -k_B T \ln \left[\sum_{\vec{n}} \prod_{j=1}^{N} \mathrm{e}^{-\frac{E_{0,j}}{k_B T}} \mathrm{e}^{-\frac{E_{trans,j,\vec{n}}}{k_B T}} \mathrm{e}^{-\frac{E_{rot,j,\vec{n}}}{k_B T}} \mathrm{e}^{-\frac{E_{vib,j,\vec{n}}}{k_B T}} \mathrm{e}^{-\frac{E_{e,j,\vec{n}}}{k_B T}} \right]$$

Since the quantum numbers can be divided into sets, where each belongs to a particular atom, we can perform the sum for each atom independently from the other. Since furthermore each individual contribution of one atom has its own set of quantum numbers, we can perform the sums even for each term independently from the other. Important is however, to include the factorial of N to avoid double counting of permutations of identical molecules. Since the individual sums are now identical for all atoms, I drop the atom index j

$$\begin{aligned} A(T, V, N) = -k_B T \ln\Big[\frac{1}{N!} \prod_{j=1}^{N} \mathrm{e}^{-\frac{E_0}{k_B T}} \sum_{n,\dots} \mathrm{e}^{-\frac{E_{trans,n}}{k_B T}} \sum_{\ell,m} \mathrm{e}^{-\frac{E_{rot,\ell,m}}{k_B T}} \\ \cdot \sum_{n_1,\dots,n_{3q-6}} \mathrm{e}^{-\frac{E_{vib,n_1,\dots,n_{3q-6}}}{k_B T}} \sum_{\vec{n}_e} \mathrm{e}^{-\frac{E_{e,\vec{n}_e}}{k_B T}} \Big] \end{aligned}$$

---

[10] german: Trägheitstensor

The factorial is now considered together with the translational term, because we obtain then simply the term for the Boltzmann gas.

$$A(T,V,N) = NE_0 \underbrace{-k_BT\ln\left[\frac{1}{N!}\sum_{n,\ldots} \mathrm{e}^{-\frac{E_{trans,n}}{k_BT}}\right]}_{A^{BG}} - Nk_BT\ln\left[\sum_{\ell,m}\mathrm{e}^{-\frac{E_{rot,\ell,m}}{k_BT}}\right]$$
$$-Nk_BT\ln\left[\sum_{n_1,\ldots,n_{3q-6}}\mathrm{e}^{-\frac{E_{vib,n_1,\ldots,n_{3q-6}}}{k_BT}}\right] - Nk_BT\ln\left[\sum_{\vec{n}_e}\mathrm{e}^{-\frac{E_{e,\vec{n}_e}}{k_BT}}\right]$$

Note that all terms except the translational terms are proportional to the particle number and only the translational term contains the volume. The translational term is, however, simply the ideal gas contribution A^{BG} We define the Helmholtz-potential of the internal degrees of freedom A_{int} as the Helmholtz-potential of a single molecule whose center of gravity is fixed in space. Thus we can write the free energy as

$$A(T,V,N) = NA_{int}(T) + A^{BG}(T,V,N) \qquad (3.30)$$

Legendre transform with respect to volume provides the Gibbs potential[11]

$$\begin{aligned} G(T,p,N) &= NA_{int}(T) + G^{BG}(T,p,N) \\ &\overset{Eq.\ 3.18}{=} N\left(A_{int}(T) + k_BT\ln\left[\frac{\lambda_T^3 p}{k_BT}\right]\right) \end{aligned} \qquad (3.31)$$

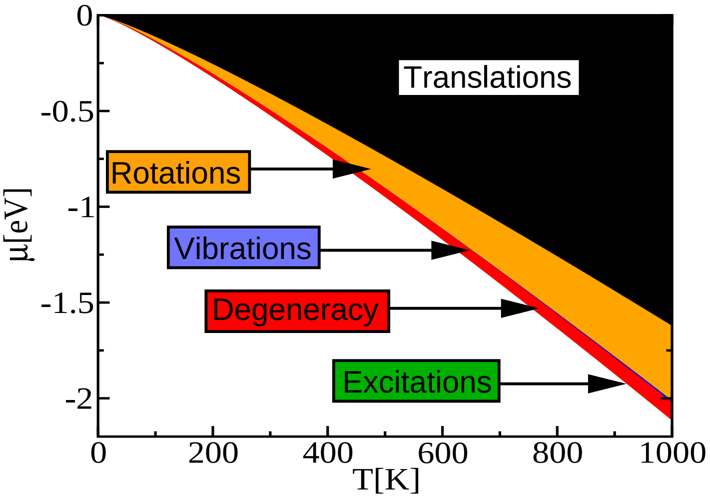


Fig. 3.7: Chemical potential of a gas of oxygen molecules at a partial pressure of 1 bar as function of temperature relative to its value at zero temperture. The dominant contribution is due the translational degrees of freedom, which are described by the ideal gas. Degeneracy describes the entropy due to the three-fold degenerate electronic ground state, which is a triplet state. That is: The oxygen molecule has two unpaired electrons. Therefore the total spin can assume the values $S_z = -1, S_z = 0, S_z = 1$. The three states are energetically degenerate and contribute $S = Nk_B\ln 3$ to the entropy. Excitations are, as it is usual the case at this low temperatures, negligible. Note that the chemical potential can be changed by 2 eV by changing the temperature.

---

[11] Note that $A_{int}(T) = G_{int}(T)$

3.8.4 Law of mass action

Now we can use the equilibrium conditions and the definition of the chemical potential $\mu(T,p,N)=\frac{\partial G(T,p,N)}{\partial N}$

$$
\begin{aligned}
0 &\overset{\text{Eq. 3.29}}{=} \sum_{j=1}^{m} \mu_j(T,p_j,N_j)\nu_{j,k} \\
&\overset{Eq.\ 3.31}{=} \sum_{j=1}^{m} \left(A_{int,j}(T) + k_BT \ln\left[\frac{\lambda_T^3 p_j}{V}\right]\right)\nu_{j,k} \\
&= \sum_{j=1}^{m} \left(\underbrace{A_{int,j}(T) + k_BT \ln\left[\frac{\lambda_T^3 p_{tot}}{V}\right]}_{\mu(T,p_{tot},N)} + k_BT \ln\left[\frac{p_j}{p_{tot}}\right]\right)\nu_{j,k} \\
&= \sum_{j=1}^{m} \mu_j(T,p_{tot})\nu_{j,k} + k_BT \sum_j \ln\left[\frac{p_j}{p_{tot}}\right]\nu_{j,k} \\
&= \sum_{j=1}^{m} \mu_j(T,p_{tot})\nu_{j,k} + k_BT \ln\left[\prod_j \left(\frac{p_j}{p_{tot}}\right)^{\nu_{j,k}}\right]
\end{aligned}
\tag{3.32}
$$

Note that the chemical potential, expressed by temperature and pressure does no more depend on the particle number, which is why we dropped the argument N_j. It can be shown that this is a general property of ideal systems.[12]

Thus we obtain our first version of the law of mass action.

$$
\prod_j \left(\frac{p_j}{p_{tot}}\right)^{\nu_{j,k}} \overset{\text{Eq. 3.32}}{=} \mathrm{e}^{-\frac{1}{k_BT}\sum_j \mu_j(T,p_{tot},N_j)\nu_{j,k}} = \prod_j \left(\mathrm{e}^{-\frac{\mu_j(T,p_{tot},N)}{k_BT}}\right)^{\nu_{j,k}}
\tag{3.33}
$$

The partial pressures can be related to the relative particle numbers[13] using the ideal gas law $p_jV = N_jk_BT$, which holds similarly for the total pressure $p_jV = N_jk_BT$, because $p_{tot} = \sum_j p_j$ and $N_{tot} = \sum_j N_j$. Thus we obtain

$$
\frac{p_j}{p_{tot}} = \frac{N_j}{N_{tot}}
\tag{3.34}
$$

Combining this result with the above form of the mass action law, we obtain its final form

---

[12] The proof is analogous to that of the Euler equation derived later. The idea behind is that an the chemical potential is an intrinsic variable and therefore does not change if we increase the size of the system by increasing all extrinsic variables proportionally. Since it has the particle number as its sole extrinsic argument, it cannot depend on this argument.

[13] In order to make sure that the ideal gas law is applicable to this case, one can go back to the definition of the partial pressure and insert the Helmholtz potential of an ideal gas.

LAW OF MASS ACTION OF THE IDEAL GAS

[a].

$$\prod_j \left(\frac{N_j}{N_{tot}}\right)^{\nu_{k,j}} \overset{Eqs.\ 3.33,3.34}{=} \underbrace{\prod_j \left(\mathrm{e}^{-\frac{\mu_j(T,p_{tot})}{k_BT}}\right)^{\nu_{j,k}}}_{K}$$

The constant on the right hand side is called **equilibrium constant**

$$K \overset{\mathrm{def}}{=} \prod_j \left(\mathrm{e}^{-\frac{\mu_j(T,p_{tot})}{k_BT}}\right)^{\nu_{j,k}}$$

---

[a] German: Massenwirkungsgesetz

Note that the law of mass action has been derived here only for ideal gases. Nevertheless it is applicable for a wide range of other ideal systems. Currently, I do not understand the underlying reason for its wider applicability and therefore I cannot precisely state its full range of applicability. However, I caution the reader regarding its limitations.

The left hand side of the law of mass action is expressed by the relative concentrations of the molecules. The important point about the law of mass action is that the equilibrium constant does *not* depend on the amount of molecules. Note that the right-hand side is independent of the number of particles since the Gibbs free energy of a gas of non-interacting molecules is linear in N as seen from Eq. 3.17 and Eq. 3.31.Irrespective of the amounts available before the reaction starts, the mass action law works with the same equilibrium constant.

To evaluate the concentrations it is best to take the logarithm of the mass action law, solve the resulting linear system of equations for the logarithm of the relative concentrations and then determine the concentration from the result.

The interpretation of the law of mass action is as follows: Consider a reaction $A+B \leftrightarrow C$. If there is a constant supply of A, and if the back-reaction is suppressed by removing the product, the amount of C produced will be proportional to B. Similarly if there is a constant supply of B the amount of products will be proportional to A. Hence the amount of products is proportional to the product of N_AN_B. Let us write $\partial_t N_C = c_1 N_A N_B$. The back reaction is simply proportional to the amount of C. Hence if we suppress the forward reaction we obtain $\partial_t N_c = -c_2 N_C$. Equilibrium is reached if the amount of C produced is equal to the amount of C that reacts, that is if $c_1 N_A N_B = c_2 N_C$, which leads us to $N_A N_B / N_C = c_2/c_1$.

This simple example actually shows the the mass action law is not at all restricted to non-interacting gases. For the ideal gas, however, we can directly work out the equilibrium constant.

3.9 Stability

Sofar we have only exploited that the entropy is stationary. For a stable state the entropy must be a maximum.

Let us now consider a homogeneous system, that can be divided into many identical subsystems. Each subsystem is in thermal and mechanical contact with the other subsystems. Hence the other subsystems act as a heat bath and volume reservoir to the selected subsystem. Hence the Gibbs potential of the subsystem must have a minimum with respect to energy and volume exchange with the corresponding reservoirs.

Let us illustrate this on an example, where we consider a system in contact with a heat bath.

Hence the Gibbs potential

$$\tilde{G}(T_{hb}, p_{vr}, U, V, N) = U - T_{hb}S(U, V, N) + p_{vr}V$$

must be a minimum.

The minimum condition says that all eigenvalues of the matrix

$$\begin{pmatrix} \frac{\partial^2\tilde{G}}{\partial U^2} & \frac{\partial^2\tilde{G}}{\partial U\partial V} \\ \frac{\partial^2\tilde{G}}{\partial V\partial U} & \frac{\partial^2\tilde{G}}{\partial V^2} \end{pmatrix}$$

must be positive.

For a symmetric 2×2 matrix this implies that the diagonal elements and the determinant is positive.

$$\frac{\partial^2\tilde{G}}{\partial U^2} \geq 0; \qquad \frac{\partial^2\tilde{G}}{\partial V^2} \geq 0; \qquad \det\begin{pmatrix} \frac{\partial^2\tilde{G}}{\partial U^2} & \frac{\partial^2\tilde{G}}{\partial U\partial V} \\ \frac{\partial^2\tilde{G}}{\partial V\partial U} & \frac{\partial^2\tilde{G}}{\partial V^2} \end{pmatrix} \geq 0$$

- The first condition says that the specific heat at constant volume must be positive

$$\begin{aligned} \frac{\partial^2\tilde{G}}{\partial U^2} &= \frac{\partial}{\partial U}\left[1 - T_{hb}\underbrace{\frac{\partial S}{\partial U}\bigg|_{V,N}}_{=1/T}\right] = -T_{hb}\frac{\partial}{\partial U}\frac{1}{T} = -T_{hb}\left(-\frac{1}{T_{hb}^2}\right)\frac{\partial T}{\partial U}\bigg|_{V,N} \\ &\overset{c_V=\frac{\partial U}{\partial T}|_{V,N}}{=} \frac{1}{T_{hb}c_V} \\ &\overset{T\geq 0}{\Rightarrow} c_V \geq 0 \end{aligned}$$

This shows that a state with negative specific heat is not stable. Such a state would have a maximum of the free energy, and would develop away from the stationary state, until it finds a minimum, where the specific heat is positive.

- **Editor: this is not finished**

$$\begin{aligned} \frac{\partial^2\tilde{G}}{\partial V^2}\bigg|_{U,N} &= \frac{\partial}{\partial V}\left[-T_{hb}\underbrace{\frac{\partial S}{\partial V}\bigg|_{U,N}}_{=-p/T} + p_{vr}\right] = T_{hb}\frac{\partial}{\partial V}\bigg|_{U,N}\left(\frac{p}{T}\right) \\ &= T_{hb}\left[\frac{1}{T_{hb}}\frac{\partial p}{\partial V}\bigg|_{U,N} - \frac{p_{vr}}{T_{hb}^2}\frac{\partial T}{\partial V}\bigg|_{U,N}\right] \\ &= \frac{\partial p}{\partial V}\bigg|_{U,N} - \frac{p_{vr}}{T_{hb}}\frac{\partial T}{\partial V}\bigg|_{U,N} \end{aligned}$$

- The third equation

3.10 Summary

3.10.1 Important formulas

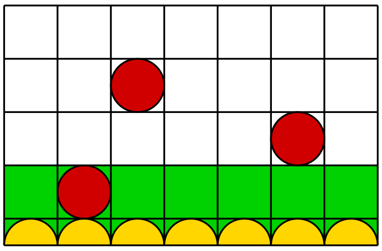
$$
\begin{aligned}
dS &= \frac{1}{T}dU + \frac{p}{T}dV - \frac{\mu}{T}dN \\
dU &= TdS - pdV + \mu dN \\
p &= -\left.\frac{dU}{dV}\right|_{S,N} \\
\mu &= -\left.\frac{dU}{dN}\right|_{S,V} \\
T &= -\left.\frac{dU}{dS}\right|_{V,N} \\
A(T,V,N) &= U - TS(U,V,N) \\
dA &= -SdT - pdV + \mu dN \\
H(S,V,N) &= U + pV \\
dH &= TdS + Vdp + \mu dN \\
G(T,p,N) &= U - TS + pV = A + pV \\
dG &= -SdT + Vdp + \mu dN \\
U_{hb}(S) &= T_{hb}S \\
U_{vr}(V) &= -p_{vr}V \\
U_{pr}(N) &= \mu_{pr}N
\end{aligned}
$$

3.10.2 Exercises

Adsorption

When a gas comes in contact with a surface some of the gas molecules The adsorption of gas molecules on a surface is important for many different fields such as film growth for transistors, heterogeneous catalysis and corrosion. In this exercise we will investigate the concentration of adsorbed molecules depending on partial pressure of the gas and temperature.

Let us first consider a simple discrete model which will provide us an insight into the underlying principles. The model consists of a square array of $n, m+1$ boxes. The bottom layer of boxes represents the surface, above which there are m layer of states in the gas. A gas molecule in the surface layer has an energy which is lower by the adsorption energy E_{ad} compared to the boxes in the boxes representing the gas. The energy of a molecule is entirely determined by the box. It is E_0 in a gas-box and $E_0 - E_{ad}$ in a surface box. There can be at most one particle in a box. The gas molecules shall be identical particles.



- determine the entropy of N particles in the total system when M particles are adsorbed.

- determine the concentration of gas molecules M/n as function of temperature.
- discuss the limes $T \to 0$ and $T \to \infty$.

Now we leave the discrete model: We consider the gas as ideal gas, while the surface still has a discrete number of adsorption sites with an adsorption energy E_{ad}.

- use the expression for the chemical potential to determine the concentration of adsorbed molecules as function of pressure and temperature.

(The following note is not required for the present exercise. For practical applications, however, note the following: Rotational and vibrational levels do not contribute much to the energy but can have an enormous impact on the concentration. It is important to introduce as many vibrational levels on the surface as there are rotational and vibrational degrees of freedom in the molecule.)

Negative temperature

Consider a system of spins that can align parallel or antiparallel to the field. The spins are related to a magnetic moment, which can have z-components values $+\mu$ and $-\mu$. The total energy of the system is

$$E(\vec{m}) = \sum_i -m_i B_z$$

1. Determine the entropy as function of the internal energy.
2. Determine the temperature as function of internal energy. Draw a graph. Discuss the solutions with negative temperature.
3. Place the spins in contact with a heat bath and investigate the stability of the resulting points of stationary free energy.

Chemical reaction

Nitrogen is the second most frequent element in the dry mass of biological matter. It is also abundant in our atmosphere. Nevertheless plant growth is severely limited by the availability of nitrogen. This is the reason why we add fertilizer in agriculture to increase crop productivity. The reason that there is a shortage for nitrogen is that nitrogen molecules are extremely inert, as they are bound by the probably strongest bond in nature. before plants can access nitrogen it must first be converted into ammonia $\mathrm{NH_3}$.

In this exercise we will calculate the concentration of ammonia in the atmosphere. Consider the reactions

$$\begin{aligned} 2H_2O &\leftrightarrow O_2 + 2H_2 \\ N_2 + 3H_2 &\leftrightarrow 2NH_3 \\ 2O_2 + N_2 &\leftrightarrow 2NO_2 \end{aligned}$$

The total pressure shall be 1 bar.

The molar Gibbs free energies in kJ/mol of the participating molecules are

O_2	0.00
NO_2	51.31
NH_3	-26.50
H_2O	-228.57
H_2	0.00

Determine the concentration of NH_3 and NO_2 relative to that of N_2.

Needs to more equations.

Chapter 4

Thermodynamic Processes

4.1 Classification of processes

Editor: include graphs for the isotherms,isobars etc of the ideal gas

We can distinguish different idealized thermodynamic processes.

isobar	p=const
isochor	V=const
isotherm	T=const
adiabatic	thermally insulated
isentropic	S=const

Fig. 4.1: Sadi Carnot, 1796-1832

We can represent this process using diagrams containing the variables T, S, p, V. The T, S diagram allows us to obtain the heat added or removed from the system, because the heat is $dQ = TdS$. The p, V diagram shows the work because $dW = pdV$.

If we consider iso-choric heating or cooling, we can use $S(U, V, N)$ and evaluate

$$\begin{aligned}
T &= \left(\left.\frac{dS}{dU}\right|_{V,N}\right)^{-1} = \left.\frac{dU}{dS}\right|_{V,N} \\
T^{BG} &\overset{Eq.\ 2.20}{=} \frac{\partial}{\partial S}\left[\frac{3\pi\hbar^2}{m}N^{\frac{5}{3}}V^{-\frac{2}{3}}\mathrm{e}^{\frac{2S}{3Nk_B}-\frac{5}{3}}\right] \\
&= \frac{1}{k_B}\frac{2\pi\hbar^2}{m}\left(\frac{N}{V}\right)^{\frac{2}{3}}\mathrm{e}^{\frac{2S}{3Nk_B}-\frac{5}{3}}
\end{aligned}$$

Similarly we can construct the pressure as function of the volume

$$\begin{aligned}
p &= -\left.\frac{dU}{dV}\right|_{S,N} \\
p^{BG} &\overset{Eq.\ 2.20}{=} -\frac{\partial}{\partial V}\left[\frac{3\pi\hbar^2}{m}N^{\frac{5}{3}}V^{-\frac{2}{3}}\mathrm{e}^{\frac{2S}{3Nk_B}-\frac{5}{3}}\right] \\
&= \frac{2\pi\hbar^2}{m}\left(\frac{N}{V}\right)^{\frac{5}{3}}\mathrm{e}^{\frac{2S}{3Nk_B}-\frac{5}{3}}
\end{aligned}$$

In a cyclic process we can evaluate the heat that has been added to the system by the enclosed area in the T-S diagram, while the work done corresponds to the area in the p-V diagram.

A thermodynamic process can be constructed from the following four idealized steps.

- isobar: An isobaric process runs at constant pressure.

$$
\begin{aligned}
T^{BG}(S) &= \left(\frac{2\pi\hbar^2}{m}\right)^{\frac{3}{5}} p^{\frac{2}{5}} \mathrm{e}^{\frac{2S}{5Nk_B}-1} \\
p^{BG}(V) &= const
\end{aligned}
$$

- iso-choric: An iso-choric process runs at constant volume

$$
\begin{aligned}
T^{BG}(S) &= \frac{1}{k_B}\frac{2\pi\hbar^2}{m}\left(\frac{N}{V}\right)^{\frac{2}{3}} \mathrm{e}^{\frac{2S}{3Nk_B}-\frac{5}{3}} \\
V(p^{BG}) &= const
\end{aligned}
$$

- adiabatic: In an adiabatic process no heat dQ is added or removed from the system

$$
\begin{aligned}
S(T^{BG}) &= const \\
p^{BG}(V) &= \frac{2\pi\hbar^2}{m}\left(\frac{N}{V}\right)^{\frac{5}{3}} \mathrm{e}^{\frac{2S}{3Nk_B}-\frac{5}{3}}
\end{aligned}
$$

- isentropic: A special case of an adiabatic process, in which the entropy is conserved. Because the heat is $dQ = TdS$ an isentropic process is always adiabatic.
- isotherm: An isothermic process runs at constant temperature.

$$
\begin{aligned}
T^{BG}(S) &= const \\
p^{BG}(V) &= Nk_BT\frac{1}{V}
\end{aligned}
$$

4.2 Reversible and irreversible processes

We have seen that systems approach a state that maximizes the overall entropy. What drives the approach to equilibrium is the increase in entropy. What appears like a driving force is actually a completely random process. Since a state with larger entropy is more probable than those with lower entropy, the system is more likely to end up in a state with larger entropy. In a macroscopic system it is so much more likely, that the chance to find the system in a state with low entropy is negligible.

Since the entropy can only grow, a process that leads to an entropy production is called **irreversible**. An irreversible process can not be brought back to its initial state, because that would require an entropy reduction. Nevertheless the entropy of a subsystem can decrease at the expense of an entropy increase of the environment, which at least offsets the entropy reduction of the subsystem.

On the other hand, we can imagine idealized processes, that do not lead to an entropy production. Such a process is called **reversible**. However, because the driving force for such a process, namely the entropy production, vanishes, such a process is also arbitrarily slow. Reversible processes are quasi-static.

A process, which proceeds so slowly, that the system is at all times in equilibrium is called **quasistatic**. What means slow? Typical systems relax into thermal equilibrium on a nanosecond time scale. Thus quasi-static process can indeed be quite fast. While all reversible processes are quasi-static, there are also quasi-static processes that are irreversible.

4.2.1 The experiment of Gay-Lussac

In order to demonstrate the difference between irreversible and reversible processes, let us discuss the experiment of Gay-Lussac. The original goal of the experiment of Gay-Lussac was to relate the specific heats at constant volume and at constant pressure. This will, however, not be our focus.

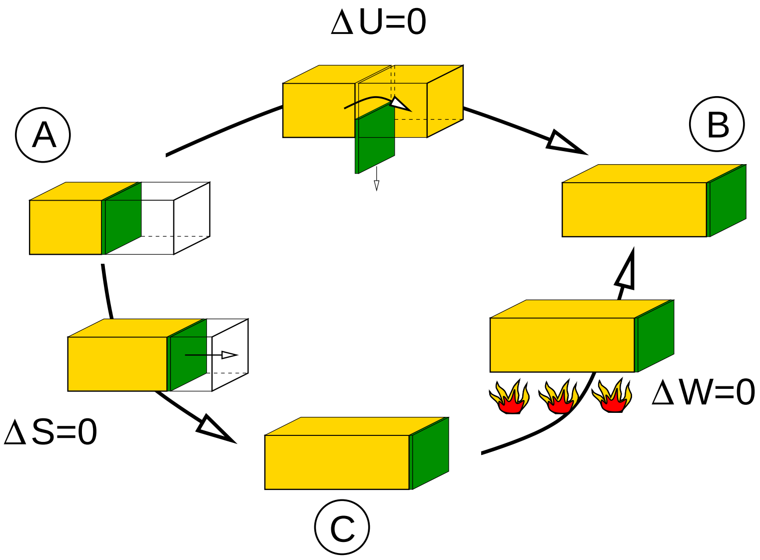


Fig. 4.2: Schematic illustration of the experiment of Gay Lussac.

The experiment of Gay-Lussac uses a container with two compartments with Volumes V_1 and V_2. The container shall be thermally insulated towards the outside, so that no energy is exchanged with the environment. Initially, one of the containers is filled with N gas molecules, while the other is empty.

1. Now the wall separating the compartments is removed, leading to a spontaneous, that is irreversible, expansion of the gas into the second compartment
2. In the second experiment the wall is moved slowly, and the force that the gas exerts against the wall is used to generate work. The energy, which is removed from the system in the form of work, is later added by heating it up.

Both processes lead to the same thermodynamical state. Pressure, temperature, energy etc. are identical. The difference is that in the first process entropy increases spontaneously, while in the seccond process, the extensive quantities are changed reversibly, and the entropy is supplied in a second step, by heat transfer.

In order to gain some experience with working out thermodynamic processes we will investigate the processes of the Gay-Lussac experiment in some detail. The goal is to understand the difference of reversible and irreversible processes. It is also demonstrated that we can replace every irreversible process by an reversible process and one where the system is heated up.

Adiabatic expansion

The wall separating the two compartments in the Gay-Lussac experiment is removed, so that no work is performed. Since the container is thermally insulated, the process is adiabatic and no heat is transported out of the system. Because $dU = TdS - pdV$ and both terms vanish, the energy is conserved, that is $\Delta U = 0$. However the process is not isentropic. The entropy increases. The expansion of the gas into the second compartment is irreversible and hence spontaneous.

Let us evaluate the entropy production for the Boltzmann gas. From the Sackur-Tetrode equation

Eq. 2.19, we can calculate the entropy difference.

$$S^{BG}(U,V,N) = Nk_B\left\{\frac{5}{2} - \ln\left[\left(\frac{3\pi\hbar^2}{m}\right)^{\frac{3}{2}}\frac{N^{\frac{5}{2}}}{VU^{\frac{3}{2}}}\right]\right\}$$
$$\Delta S_{AB} = Nk_B\ln\left(\frac{V_B}{V_A}\right) = Nk_B\ln\left(\frac{V_1+V_2}{V_1}\right) \geq 0 \qquad (4.1)$$

For the ideal gas the energy the temperature remains constant: Since the internal energy is directly connected to the temperature via the caloric equation of state, the energy conservation leads to a constant temperature.

This is opposite to common experience. If there is an attractive interaction between the molecules, this "binding energy" is partially lost during the expansion. The energy to break this cohesion is supplied by lowering the kinetic energy of the gas molecules. Hence the temperature is reduced. An example can be exploited on desert trips: One way to keep drinks cool is to cover them with a moist cloth. (Even though much less stylish, it also works well with your head. This technique is even more useful, because with a cool head you may not waste your remaining water to cool your drinks.) Since there is no such cohesion in an ideal gas, the temperature remains constant in our example.

Isentropic expansion

Now we move the wall slowly into the right compartment. During this process we let the system perform work. The energy of the system decreases by the amount of work that is performed, that is $dU = -dW = -pdV_1$. The entropy on the other hand remains the same because this process is quasi-static and no heat is lost to the environment.

Let us calculate the work in two different ways:

- We can use the two relevant equations of state, which are the ideal gas law and the caloric equation of state.

$$pV = Nk_BT \quad \overset{U=\frac{3}{2}Nk_BT}{\Rightarrow} \quad p = \frac{2U}{3V}$$
$$dU = -pdV = -\frac{2U}{3V}dV \quad \Rightarrow \quad \underbrace{\frac{1}{U}dU}_{d\ln(U)} = -\frac{2}{3V}dV$$
$$\frac{d\ln(U)}{dV} = -\frac{2}{3V} \quad \Rightarrow \quad \ln(U) = C - \frac{2}{3}\ln(V) = C + \ln(V^{-\frac{2}{3}})$$
$$U_C = U_A\left(\frac{V_C}{V_A}\right)^{-\frac{2}{3}}$$

We immediately see one major disadvantage of this technique, namely that it requires to solve differential equations. This can easily become very cumbersome for more complicated systems.

Now we can calculate temperature and pressure for the final state

$$U = \frac{3}{2}Nk_BT \quad \Rightarrow \quad T_C = \frac{2U_C}{3Nk_B} = T_A\left(\frac{V_C}{V_A}\right)^{-\frac{2}{3}}$$
$$pV = Nk_BT \quad \Rightarrow \quad p_C = \frac{Nk_BT}{V} = p_A\left(\frac{V_C}{V_A}\right)^{-\frac{5}{3}}$$

We see that the gas cools as energy is removed from the system. This effect can easily be observed with a bicycle pump. If we close the opening with a finger and compress the gas, the pump heats up. This is the reverse effect to the one seen here, the isentropic compression.

- Now we determine the same results using the thermodynamic potentials:

  We start with the internal energy as function of entropy and volume.

$$U^{BG} \stackrel{Eq. 2.20}{=} \frac{3\pi\hbar^2}{m} N^{\frac{5}{3}} V^{-\frac{2}{3}} \mathrm{e}^{\frac{2S}{3Nk_B}-\frac{5}{3}}$$

  Since entropy and particle number remain constant, we can immediately read of the change of the internal energy when the volume is changed

$$\begin{aligned} U_C - U_A &= \frac{3\pi\hbar^2}{m} N^{\frac{5}{3}} \left(V_C^{-\frac{2}{3}} - V_A^{-\frac{2}{3}}\right) \mathrm{e}^{\frac{2S_A}{3Nk_B}-\frac{5}{3}} = U_C \left[\left(\frac{V_C}{V_A}\right)^{-\frac{2}{3}} - 1\right] \\ U_C &= U_A \left(\frac{V_C}{V_A}\right)^{-\frac{2}{3}} \qquad (4.2) \end{aligned}$$

  In this case the result has been obtained analytically without differential equations. Of course a lot of work had to be done to obtain the thermodynamic potential in the first hand.

  Now, we obtain the temperature and pressure as derivatives of the thermodynamic potentials, even though we could as well use the equations of state at this point.

$$\begin{aligned} T_C &= \left.\frac{dU}{dS}\right|_{V_C,N} = \frac{2U_C}{3Nk_B} = \frac{2U_A}{3Nk_B}\left(\frac{V_C}{V_A}\right)^{-\frac{2}{3}} \\ p_C &= -\left.\frac{dU}{dV}\right|_{S_C,N} = \frac{2}{3V_C}U_V = -\frac{2U_A}{3V_A}\left(\frac{V_C}{V_A}\right)^{-\frac{5}{3}} \end{aligned}$$

Isochoric heating

Now we put back the energy, we obtained in the form of work, by heating the gas up. While the expansion was reversible this process is irreversible.

We use again the expression for the entropy

$$\begin{aligned} S^{BG} &= Nk_B\left\{\frac{5}{2} - \ln\left[\left(\frac{3\pi\hbar^2}{m}\right)^{\frac{3}{2}} \frac{N^{\frac{5}{2}}}{VU^{\frac{3}{2}}}\right]\right\} \\ \Delta S_{CB} &= \frac{3}{2}Nk_B \ln\left(\frac{U_B}{U_C}\right) \stackrel{U_B=U_A}{=} \frac{3}{2}Nk_B \ln\left(\frac{U_A}{U_C}\right) \\ &\stackrel{Eq. 4.2}{=} \frac{3}{2}Nk_B \ln\left[\left(\frac{V_C}{V_A}\right)^{\frac{2}{3}}\right] = Nk_B \ln\left[\frac{V_C}{V_A}\right] \\ &= Nk_B \ln\left[\frac{V_1+V_2}{V_1}\right] \end{aligned}$$

Thus the entropy increased by exactly the amount as during the adiabatic expansion, given in Eq. 4.1.

Discussion

We can draw the following lesson from the above experiment. **To each irreversible process we can attribute a sequence of an isentropic process and an irreversible heating.** Both processes can be quasi-static.

There is a small puzzle that we wish to resolve. Imagine the process, where the wall is moved slowly, but we do not use the work to collect the energy. The process seems similar to the isentropic expansion, but it is still adiabatic, that is no energy is removed from the system. The answer to this problem is that this process cannot be isentropic. In order to move the wall slowly there must be some force hindering its motion. This force can only come from an external system that can collect work, and thus change its internal energy, or from friction. The friction dissipates the energy as the wall moves. Since the system is adiabatic, the dissipated energy is returned to the system in the form of heat. Thus the corresponding process is quasi-static but irreversible.

For every irreversible process we can find a corresponding quasi-static process that leads to the same final state. In the quasi-static process we can assume the system always in internal equilibrium. Thus the intensive variables are defined at each time during the process, which is not possible in an abrupt process.

4.3 Thermal engines

4.3.1 The principles

A thermal engine is one that uses thermal energy into useful work. It should be cyclic so that the conversion does not come to halt. The prototypical thermal engine is a car engine. Another type of engines are those that use work to cool a system. Such an engine is obtained if we run the former backwards. The typical example would be a refrigerator.

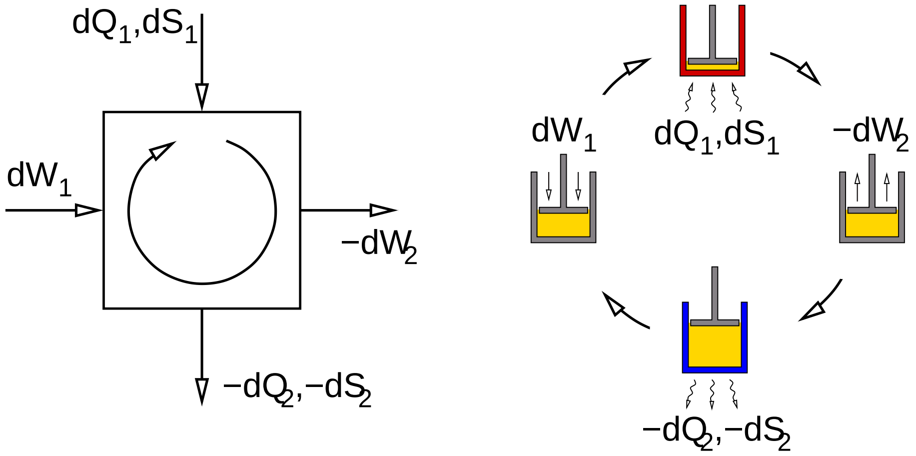


Fig. 4.3: The principle of a thermal engine is a cycle which converts heat into work. In the first step (left), a small amount of work is applied to the engine, for example to compress a gas. Then (top) the engine is heated, which brings the gas into a state of higher energy. This energy is (right) put to use, by extracting a large amount of work from the heat engine. The energy of the gas is now much smaller. A small mount of heat must be still released (bottom) before the container is compressed again. Overall a small amount of work and a large amount of heat is converted into a large amount of work and a small amount of heat. The energy conservation requires that the sum of heat and work added to the engine is the same as the one extracted from it. The net entropy however increases, and this entropy increase is what drives the engine.

The prototypical system consists of four steps.

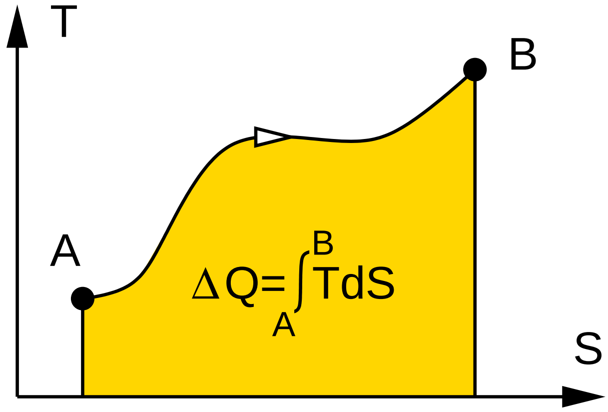
1. heating: The gas is in the compressed state, and its temperature is raised. The energy of the of the engine increases by $dQ_1 > 0$.
2. expansion: The gas is allowed to expand and work $|\Delta W_2| = -\Delta W_2$ is performed by the system. The energy of the engine changes by $\Delta W_2 < 0$.
3. cooling: As the system is expanded we cool the system. Thermal energy $|\Delta Q_2| = -dQ_2$ is lost from the engine in the form of heat. The energy of the engine changes by $dQ_2 < 0$

4. compression: We need to perform work to bring the system back into its initial state. The energy of the engine increases by $\Delta W_1 > 0$.

4.3.2 Graphical representation of a cyclic process

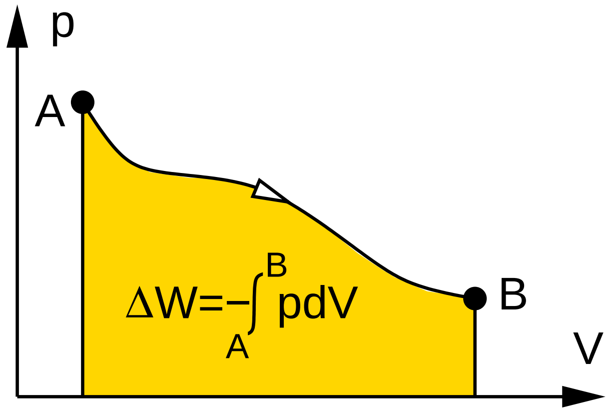
Heat equation

The heat added to a system or removed from it can directly be read from a diagram which shows the temperature versus entropy. Since the heat is $\Delta Q = \int_{S_1}^{S_2} T dS$, the heat added to the system is directly related to the area below $T(S)$ in the interval from S_1 to S_2. If the direction of the process points to the right, that is towards increasing entropies, heat is added to the system. If the process goes to the left heat is removed from the system.



Work equation

The work done on the system performed by it can be read from a diagram which shows the pressure versus volume. Since the work done by the system is $\Delta W = -\int_{V_1}^{V_2} pdV$, the work performed by the system is directly related to the area below $p(V)$ in the interval from V_1 to V_2. If the direction of the process points to the right, that is towards increasing volume, work is done by the system. If the process goes to the left work is done on the system.



Cyclic processes

If we consider a cyclic process the lines in the $T-S$ diagram and the $p-V$ diagram are closed. The area enclosed by the line in the $T-S$ diagram is the heat converted into work, if the process runs clockwise. If it runs counterclockwise heat is generated by the engine. The work done by the system can be read from the $p-V$ diagram.

The areas of the curve in the $T-S$ and the $p-V$ diagram are identical due to energy conservation.

The efficiency can be read from the T_S diagram. The lost heat is the area below the lower curve.

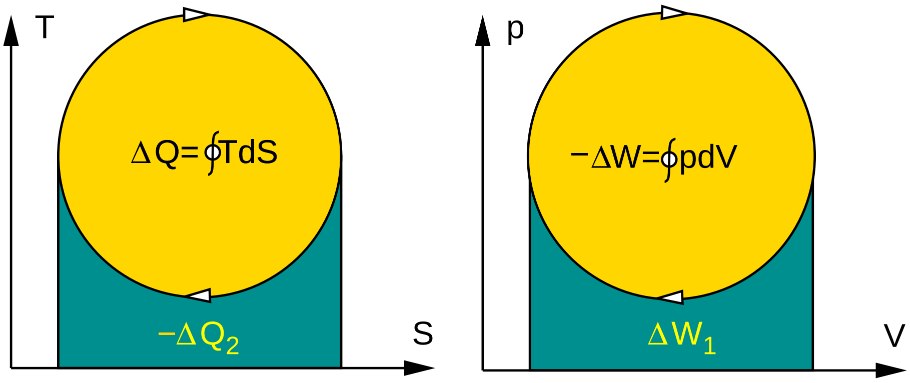


Fig. 4.4: The overall heat that is converted into work per turnover can be seen from both the p-V or the T-S diagram as the area inside the cyclic path of the engine. $|\Delta Q_2|$ is the thermal energy that is not converted into work. An engine is efficient, if the area related to $|\Delta Q_2|$ is small compared to ΔQ.

4.3.3 Reversible heat source

In order to build an efficient engine it is important to identify the sources for heat production. Let us consider consider the process where one system is heated by bringing it into contact with another one, so that energy can be exchanged.

During thermal contact the energy of the two subsystems change by

$$dU_1 = T_1 dS_1 \qquad dU_2 = T_2 dS_2$$

During the thermal coupling the total energy of both systems must be conserved. Hence

$$0 = dU_1 + dU_2 = T_1 dS_1 + T_2 dS_2 \qquad \Rightarrow \qquad dS_2 = -\frac{T_1}{T_2} dS_1$$

Thus the total entropy production is

$$\begin{aligned} dS_{tot} &= dS_1 + dS_2 = (1 - \frac{T_1}{T_2}) dS_1 = \frac{T_2 - T_1}{T_2} dS_1 \\ &= \frac{T_2 - T_1}{T_1 T_2} \underbrace{T_1 dS_1}_{dQ_1} = \left(\frac{1}{T_1} - \frac{1}{T_2}\right) dQ_1 \end{aligned}$$

Thus we see that the heat transfer is reversible, if the two systems are at the same temperature.

Like with all reversible processes, there is no driving force for the heat transfer, and the heat transfer would be infinitely slow. Thus an engine, that runs fast, necessarily also produces entropy. It is the entropy production that actually makes the engine to turn.

4.3.4 Reversible work source

Mechanical systems generate heat only due to friction. For electrical systems driving a current through a wire with a finite resistance also creates heat. This problem can also be described by a kind of friction acting on the electrons.

Thus in order to avoid entropy production during work transfer we need to avoid friction.

4.3.5 Efficiency of a thermal engine

We can estimate the efficiency of the engine. The **efficiency** is defined as the fraction of total work performed by the heat provided.

$$\eta = \frac{-\Delta W_2 - \Delta W_1}{\Delta Q_1} \tag{4.3}$$

where $-\Delta W_2$ is the work done by the engine and ΔW_1 is the work consumed by the engine, as defined in Eq. 1.75 on p. 42. Thus $-\Delta W_2 - \Delta W_1$ is the net work performed by the engine in each cycle. ΔQ_1 is the amount of heat introduced into the engine, while ΔQ_2 is the waste heat. Thus the efficiency is the fraction of heat energy converted into useful work.

We use the energy conservation

$$\Delta Q_1 + \Delta W_2 + \Delta Q_2 + \Delta W_1 = 0$$

which says that the total amount of energy entering the engine must be equal to the amount of energy leaving the engine in each cycle.

The efficiency is now calculated as

$$\begin{aligned} \eta &= \frac{-\Delta W_2 - \Delta W_1}{\Delta Q_1} = \frac{\Delta Q_1 + \Delta Q_2}{\Delta Q_1} = 1 + \frac{\Delta Q_2}{\Delta Q_1} \\ &= 1 + \underbrace{\frac{\int_{(2)} T dS}{\int_{(1)} T dS}}_{\leq 0} \end{aligned}$$

The index (1) denotes that the integration is done along the upper, high-temperature contour in the heat equation, while the index (2) denotes the integration along the lower, low-temperature contour of the heat equation. Because the entropy changes in opposite directions in the high-temperature and the low-temperature contour, the efficiency is always less than one.

If the heat transfer is isothermal, at an upper temperature T_2 and a lower temperature T_1, the efficiency[1] is simply

OPTIMUM EFFICIENCY

$$\eta = 1 - \frac{T_2}{T_1}$$

because both integrations extend over the same entropy interval.

Thus the optimum efficiency of a thermal engine is $\eta = 1 - \frac{T_2}{T_1}$. Unfortunately heat cannot be converted into work completely, because we can neither achieve infinite temperatures nor zero temperatures. However, if the lower temperature is the ambient temperature, we can increase the efficiency by raising the temperature. This is a considerable challenge for engineers and material scientists. For example the turbines of airplane engines are often run above the melting temperatures of its turbine blades, which can only be achieved by a clever coating and cooling of the turbine blades.

The optimum efficiency will only be reached by reversible engines. Since reversible engines lack a driving force, the entropy production, they are also very slow. Thus in order to make the engine not only efficient, but also fast we need to allow for irreversible processes. Irreversible processes, however, reduce the efficiency of the engine.

---

[1] german(efficiency)=Wirkungsgrad

4.4 Engines

Combustion engines use a gas container, the cylinder undergoes a cyclic process, where thermal energy resulting from burning fuel is converted into work.

4.4.1 Carnot cycle

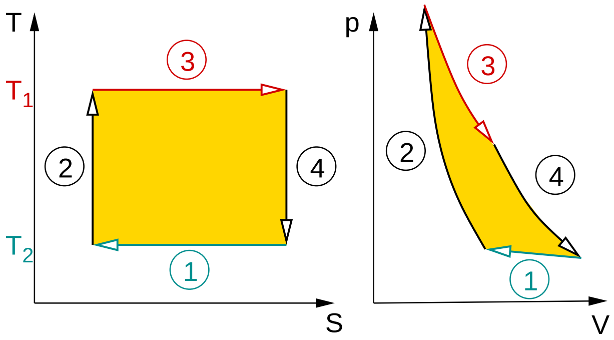


Fig. 4.5: Sketch of the heat equation and the work equation of the Carnot cycle.

The Carnot cycle[2] is the most efficient engine and is defined as a process that does not produce entropy, but merely translates the heat into work. What are the requirements for an engine that does not produce entropy?

- The transfer of heat between a heat source and the engine is most efficient if the temperature difference between the systems vanishes. If the heat source is a reservoir, the transfer heat must proceed at constant temperature.
- In order to cycle between the two temperatures of the two heat baths, entropy must not rise, so that the other processes must be isentropic.

Therefore the Carnot cycle consists of two isothermal and two isentropic processes.

A Carnot cycle can be implemented in any system, but here we illustrate the process on the example of a gas that produces work by repeatedly expansion and compression.

1: isothermal compression at low T: The piston moves inward, while the cylinder is connected to a cold heat bath.

2: isentropic compression The heat bath is decoupled while the compression continues. The temperature rises. The process is adiabatic.

3: isothermal expansion at high T: The piston moves outward while the temperature is controlled by a hot heat bath. While the gas expands, heat is transferred to the system to maintain the same temperature. If the heat source is at the same temperature as the gas, the entropy is transferred reversibly.

4: isentropic expansion The heat bath is decoupled while the expansion continues. The temperature falls.

---

[2] Sadi Nicolas Léonard Carnot. French Scientist, 1796-1832. His works were aimed at improving steam engines. With the process named under him he showed that the efficiency depends only on the temperature difference and not on the type of gas used. He also introduced the concept of reversibility.

The Carnot cycle reaches the theoretical possible efficiency, but it has the disadvantage that it is also infinitely slow, because there are no irreversible processes that drives that process. Thus it is an idealization of a true engine.

To avoid a confusion, the Carnot process can also be driven irreversibly, which leads to a lower than ideal efficiency. This is the case, if the temperature of the gas remains constant, but the heat source is at a higher temperature, so that the heat transfer is reversible.

4.4.2 Combustion engine: Otto engine

Here we want to investigate a real engine, the Otto motor. (see also `http:\\www.blueneptune.com/~xmwang/myGUI/OttoG.html`) (see also `http:\\www.blueneptune.com/~xmwang/myGUI/DieselG.html`)

Let us describe here the process in an Otto engine. We assume that a cylinder can cycle between two volumes V_1 and V_2, where $V_1 < V_2$.

1: Isobaric intake: The piston moves outward and a gas-fuel mixture is injected at constant pressure. No work is performed. (Step 6-1 in Fig. 4.6). The amount of gas is estimated from the ideal gas law $N = \frac{pV_2}{k_BT}$, where V_2 is the volume of the expanded cylinder, and p and T are ambient pressure and temperature. A turbocharger allows to raise the pressure during this step.

2: Adiabatic compression The piston moves inward and the gas is compressed. During compression it heats up. This means that the temperature rises, but no heat is lost to the environment in the process. The compression of the gas consumes work. (Step 1-2 in Fig. 4.6).

$$dU = -pdV \quad \overset{Eq.\ 2.17,2.23}{\Rightarrow} \quad \frac{3}{2}Nk_BdT = -\frac{Nk_BT}{V}dV$$
$$\Rightarrow \frac{3}{2T}dT = -\frac{1}{V}dV \quad \Rightarrow \quad d\left(\frac{3}{2}\ln[T]\right) = -d\left(\ln[V]\right)$$
$$T(V) = T_1\left(\frac{V}{V_>}\right)^{-\frac{2}{3}} \Rightarrow T_2 = T_1\left(\frac{V_>}{V_<}\right)^{\frac{2}{3}}$$
$$dW = pdV \overset{Eq.\ 2.23}{=} \frac{Nk_BT(V)}{V}dV = Nk_BT_1V_>^{\frac{2}{3}}V^{-\frac{5}{3}}dV$$
$$\Delta W = Nk_BT_1V_>^{\frac{2}{3}}\left[-\frac{3}{2}\left(V_<^{-\frac{2}{3}} - V_>^{-\frac{2}{3}}\right)\right]$$
$$= \frac{3}{2}Nk_BT_1\left(1-\left(\frac{V_>}{V_<}\right)^{\frac{2}{3}}\right)$$

3: Iso-choric heating: The compressed gas is ignited and therefore reaches a high temperature. The combustion energy ΔU, which can be estimated from the amount of gasoline injected, is directly converted into heat. The prototypical reaction would be

$$CH_3(CH_2)_{n-2}CH_3 + \tfrac{3n+1}{2}O_2 \rightarrow nCO_2 + (n+1)H_2O$$

The new temperature can be estimated from the caloric equation of state

$$T_3 = T_2 + \frac{2\Delta U}{3Nk_B} = T_1\left(\frac{V_>}{V_<}\right)^{\frac{2}{3}} + \frac{2\Delta U}{3Nk_B}$$

(Step 2-3 in Fig. 4.6) (A Diesel engine replaces this step by an isobaric heating step)

4: Adiabatic expansion: The pressure that builds up due to the high temperature is converted into work as the piston moves outward. This is the step that drives the engine. We can estimate the work performed in this step analogous to the adiabatic compression.

$$\Delta W_{3-4} = \frac{3}{2}Nk_BT_3\left(1-\left(\frac{V_<}{V_>}\right)^{\frac{2}{3}}\right)$$

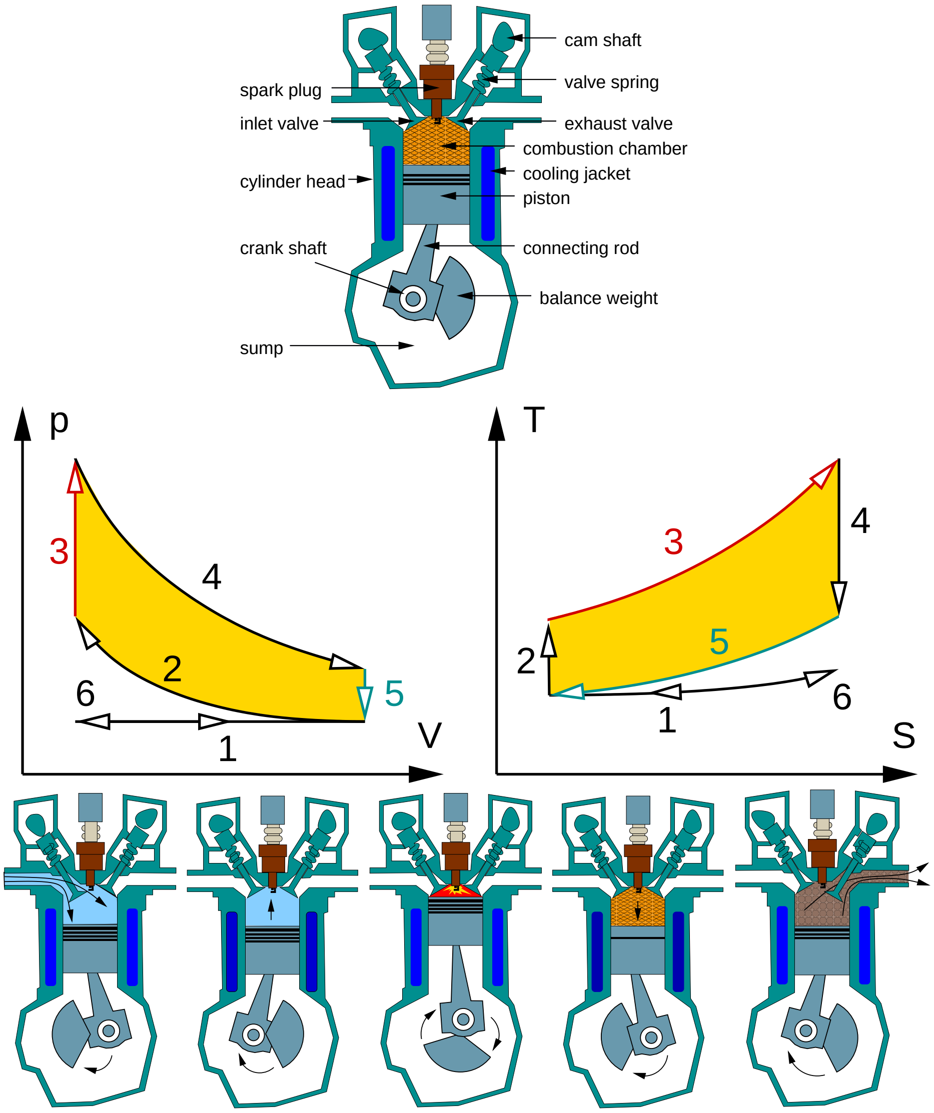


Fig. 4.6: Schematic drawing of a cylinder of a car engine. Below: p-V and $T - S$ diagram of the otto process. 1) Isobaric intake: Gasoline is injected into the cylinder, while the piston moves outward. (2) Adiabatic compression. The gasoline-air vapor is compressed at constant entropy. (3) Isochoric heating: The compressed gas is ignited. 4) Adiabatic compression: The piston moves outward and performs work. 5) isochoric cooling. 6) isobaric exhaustion: the combustion gases are exhausted. Below schematic drawings of the the states 1-6 of the cylinder from left to right. (Translation to German: Cam shaft=Nockenwelle, crank shaft=Kurbelwelle, piston=Kolben, valve spring= Ventilfeder, combustion chamber=Brennkammer, inlet valve=Einlassventil, exhaust valve= Auslassventil, spark plug=Zündkerze)

(Step 3-4 in Fig. 4.6)

5: Iso-choric cooling The gas cools down, giving heat to the cooling system. (Step 4-5 in Fig. 4.6)

6: Isobaric exhaustion: The piston moves inward pushing the exhaust gases out.(Step 5-6 in Fig. 4.6)

We can now evaluate the efficiency of the engine by comparing the energy used by combustion with the work done Thus we can calculate the total work per amount of energy ΔU supplied by gasoline.

$$\begin{aligned}
\Delta W_{1-6} &= \Delta W_{1-2} + \Delta W_{3-4} \\
&= \frac{3}{2} N k_B T_1 \left(1 - \left(\frac{V_>}{V_<}\right)^{\frac{2}{3}}\right) + \frac{3}{2} N k_B T_3 \left(1 - \left(\frac{V_<}{V_>}\right)^{\frac{2}{3}}\right) \\
&= \frac{3}{2} N k_B \left[T_1 \left(1 - \left(\frac{V_>}{V_<}\right)^{\frac{2}{3}}\right) + \left(T_1 \left(\frac{V_<}{V_<}\right)^{\frac{2}{3}} + \frac{2\Delta U}{3 N k_B}\right) \left(1 - \left(\frac{V_<}{V_>}\right)^{\frac{2}{3}}\right)\right] \\
&= \frac{3}{2} N k_B \frac{2\Delta U}{3 N k_B} \left(1 - \left(\frac{V_<}{V_>}\right)^{\frac{2}{3}}\right) \\
&= \left(1 - \left(\frac{V_<}{V_>}\right)^{\frac{2}{3}}\right) \Delta U
\end{aligned}$$

An optimum process would convert all energy supplied ΔU into work ΔW, so that $\Delta W_{1-6}/\Delta U = 1$. The most important factor for the efficiency is the compression ratio $V_>/V_<$, which should be as large as possible. There are a number of limitations to improve the compression ratio:

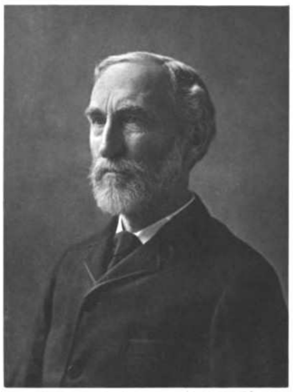
- We need to inject the proper amount of oxygen to burn the gasoline. This places a lower bound to $\frac{N}{\Delta U}$, where N is the amount of gas in the filled cylinder. Thus in order to increase the power of the engine, we either need to increase the frequency of the cycle, increase the pressure of injection using a turbocharger, or increase the volume of the cylinder. An optimum process will use a defined ratio $C = \frac{N}{\Delta U}$.
- the highest temperature T_3 is limited by the thermal stability of the materials in the cylinder. Let us call the highest allowed temperature T^{max}. Then the requirement is $T_3 < T_{max}$. T_3 is determined on the one hand by the compression ratio, and on the other hand the amount of fuel burned per amount of gas, namely C.

$$\begin{aligned}
\frac{V_>}{V_<} &= \left(\frac{T_{max}}{T_1} - \frac{2C}{3 k_B T_1}\right)^{\frac{3}{2}} \\
\frac{\Delta W_{1-6}}{\Delta U} &= \left(1 - \frac{1}{\frac{T_{max}}{T_1} - \frac{2C}{3 k_B T_1}}\right)
\end{aligned}$$

Thus we see that the critical factor for the efficiency of the engine is the maximum operating temperature T_{max}. Efficient cooling of the engine, that is lowering T_1 is important too.

Chapter 5

The Language of Thermodynamics



J. Willard Gibbs

Fig. 5.1: Josiah Gibbs, 1839-1903

Thermodynamics deals with equilibrium states. Equilibrium states are those, for which both conjugate pairs of intensive and extensive variables are defined. This implies that the probabilities are given by the maximum-entropy principle for a given set of constraints.

$$P_{\vec{n}}(T,f_i)=\frac{1}{Z(T,f_i)}\mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}+\sum_k f_kX_{k,\vec{n}})}$$

The expectation values of the observables are obtained via

$$\begin{aligned}U(T,f_k)&=\sum_{\vec{n}}P_{\vec{n}}(T,f_i)E_{\vec{n}}\\X_k(T,f_k)&=\sum_{\vec{n}}P_{\vec{n}}(T,f_i)X_{k,\vec{n}}\end{aligned}$$

and the entropy is

$$S(T,f_k)=-k_B\sum_{\vec{n}}P_{\vec{n}}(T,f_i)\ln\Big(P_{\vec{n}}(T,f_i)\Big)$$

Thus, in thermodynamics, we deal with a subset of possible statistical distributions. The idea behind this restriction is that any system will become more random with time and thus approach **thermal equilibrium**. The equilibrium state is the most random state compatible with the restrictions posed by conservation laws.

This implies that every non-equilibrium state will undergo transitions that change the probabilities in the direction of the equilibrium state. Thus a non-equilibrium state is not stationary, if left alone[1]. Equilibrium states, on the other hand, are stationary. Thermodynamics is therefore a theory of static states.

5.1 Laws of Thermodynamics

Here we derived thermodynamics from a fundamental concept, the maximum-entropy principle. Historically, thermodynamics developed as an empirical theory. The underlying concepts have not been known at that time. Also as empirical theory thermodynamics provides a consistent framework that can be derived from a number of postulates:

---

[1] Stationary non-equilibrium states can be constructed by creating a flux through the system. Life is an example. Our equilibrium state, which we approach after death, is dust. In order to maintain our highly ordered state, and to avoid an increase in entropy, we have to constantly consume food and air and excrement less energetic and less ordered material.

- First Law: Energy conservation
- Second Law: maximum-entropy principle
- Third Law (**Nernst's Theorem**): The entropy vanishes at zero temperature, or, more exactly $\lim_{T\to 0}\frac{S(N)}{k_BN}=0$.

The energy conservation follows from the Noether's theorem assuming time-translation symmetry, which is equivalent to saying that the laws of nature do not change with time. Thermodynamics defines a new form of energy, namely heat. This energy is due to microscopic degrees of freedom such as the energy due thermal motion of atoms in a solid or gas.

The maximum-entropy principle is based on the assumption that a system will completely randomize after a while and will undergo all transition compatible with conservation laws. The second law of thermodynamics was first formulated by Rudolf Clausius[2] in 1865.

Nernst's theorem[3] is based that at zero temperature only the ground state is occupied. If the ground-state is non-degenerate, the entropy vanishes. If it is degenerate, with degeneracy g, the entropy is $S(T=0)=k_B\ln[g]$. Mathematical physics says that all interacting systems have non-degenerate ground states.

5.2 Fundamental relation and equations of state

All thermodynamic properties are contained in a thermodynamic potential and its fundamental relation Eq. 1.73.

The fundamental relation contains the information on the equations of state

$$
\begin{aligned}
&dU=TdS-\sum_{k=1}^{t}f_kdX_k\\
\Rightarrow\quad &T=\left.\frac{dU}{dS}\right|_{\{f_k\}}\qquad f_k=-\left.\frac{dU}{dX_k}\right|_{S,\{f_{j\neq k}\}}
\end{aligned}
$$

For the internal energy as function of S,V,N we obtain the equations of state from its fundamental relation

$$
\begin{aligned}
&dU=TdS-pdV+\mu dN\\
\Rightarrow\quad &T=\left.\frac{dU}{dS}\right|_{V,N}\qquad p=-\left.\frac{dU}{dV}\right|_{S,N}\qquad \mu=\left.\frac{dU}{dN}\right|_{S,V}
\end{aligned}
$$

Note that knowing the function $U(T,V,N)$ is not sufficient since there is no fundamental relation for it.

The equations of state allow to evaluate the corresponding thermodynamic potentials up to a constant by integration along a line integral in the space of extensive variables.

$$
U(S,X_k)=U_0+\int_{S_0,\{X_{k,0}\}}^{S,\{X_k\}}\left(T(S,\{f_k\})dS-\sum_k f_kdX_k\right) \tag{5.1}
$$

The integral is analogous to

$$
f(\vec r)=f(\vec r_0)+\int_{\vec r_0}^{\vec r}d\vec r\,'\,\vec\nabla\big|_{\vec r\,'}f=f(\vec r_0)+\int_{\vec r_0}^{\vec r}\sum_i\left.\frac{df}{dr_i}\right|_{\vec r\,'}dr_i
$$

---

[2] Rudolf Clausius **Biographical note**

[3] **Editor: biographical note on Nernst**

Eq. 5.1 is used to determine thermodynamic potentials experimentally. During the process one determines the intensive and extensive parameters f_k and X_k respectively, and the temperature. The only difficulty is that the entropy S cannot be measured. For homogeneous systems, the entropy can be reconstructed from the Gibbs-Duhem equation derived in section 5.4. For other systems, we need to use other measures such as measuring the heat flow $dQ = TdS$ and performing the process quasistatically to avoid spontaneous entropy production via irreversible processes.

5.3 Euler Equation

The Euler Equation[4] rests on the fact that the entropy is an extensive variable.

Important for the Euler equations are the **assumptions**: The system under study must be **homogeneous**.

HOMOGENEOUS SYSTEM

A system is homogeneous, if it is composed out of many identical and statistically uncorrelated subsystems. A homogeneous system is uniform and has a single phase.

Homogeneous systems are, for example for a gas, a liquid or a homogeneous solid. The requirement of statistically uncorrelated systems breaks down on a length scale comparable to the de-Broglie wavelength.

The property that the entropy is extensive, implies that a system which is twice as large as the original system has twice the entropy of the original system. This is a direct consequence of the additivity of the entropy Eq. 1.11. This additivity is, however, only valid for statistically uncorrelated systems. This implies that the presence of the second subsystem does not affect the the first subsystem.

In order to perform the proof of Euler's equation, we need to identify all extensive quantities in the expression of the entropy. These are the observables X_k for which we fix the mean values, and other parameters, Y_k such as the volume. Note that, as an example, the number of particles can be treated either as observable, for which we fix the mean value, or as a parameter, if its value is fixed exactly.

The additivity of the entropy is expressed by[5]

$$S(\lambda U, \{\lambda X_i\}, \{\lambda Y_j\}) = \lambda S(U, \{X_k\}, \{Y_j\})$$

where λ is an arbitrary scale factor. Similarly we can write

$$U(\lambda S, \{\lambda X_k\}, \{\lambda Y_k\}) = \lambda U(S, \{X_k\}, \{Y_k\})$$

If we now form the derivative with respect to λ at the value $\lambda = 1$ we obtain the **Euler equation**.

---

[4] Leonhard Euler: 1707-1783 Swiss mathematitian and physicist. He invented much of the terminology of modern analysis and termed the notation of a mathematical function. He contributed to mechanics, fluid dynamics, optics and astronomy.

[5] The reader may verify this property for the entropy of the Boltzmann gas given by the Sackur-Trtrode equation Eq. 2.19 on p.56.

EULER EQUATION

$$U=\underbrace{\frac{\partial U}{\partial S}}_{T}S+\sum_k\underbrace{\frac{\partial U}{\partial X_k}}_{-f_k}X_k+\sum_j\frac{\partial U}{\partial Y_j}Y_j=TS-\sum_k f_kX_k+\sum_j\frac{\partial U}{\partial Y_j}Y_j \qquad (5.2)$$

The variables Y_k are external extensive parameters. They act in the same fashion as the extensive variables whose mean values are fixed.
Note the assumption of the Euler equation, namely that the system is homogenous.

For systems described by the extensive variables S, V, N, we obtain.

$$\begin{aligned}U&=TS+\mu N+\frac{dU}{dV}V=TS+\mu N-pV\\A&=U-TS=\mu N-pV\\H&=U+pV=TS+\mu N\\G&=U-TS+pV=\mu N\\\Phi&=U-TS-\mu N=-pV\end{aligned}$$

The reader is encouraged to verify these relations for the Boltzmann gas.

5.4 Gibbs-Duhem relation

We use the Euler equation Eq. 5.2 to evaluate the differential change of the internal energy. The latter can also be expressed by the fundamental equation Eq. 1.73. If we subtract the two from each other we are lead to the Gibbs-Duhem relation. For reasons of simplicity, we drop the external variables Y_k.

$$\begin{aligned}dU&\overset{Eq.\ 5.2}{=}TdS+SdT-\sum_k X_kdf_k-\sum_k f_kdX_k\\dU&\overset{Eq.\ 1.73}{=}TdS-\sum_k f_kdX_k\\\Rightarrow 0&=\quad SdT-\sum_{k=1}^{t}X_kdf_k\end{aligned}$$

which leads to the so-called **Gibbs-Duhem relation**[6]

GIBBS-DUHEM RELATION

$$0=SdT-\sum_{k=1}^{t}X_kdf_k \qquad (5.3)$$

Note that the Gibbs Duhem relation underlies the same assumptions as the Euler equation, namely that the system is homogeneous in the sense that it is composed of many identical subsystems, whose interaction is negligible.

The Gibbs-Duhem relation is important for determining the equations of state from experiments. We can measure X_k, f_k and the temperature. The entropy, however, is not accessible. Thus we can

---

[6] Pierre Maurice Marie Duhem. French scientist 1861-1916. His first PhD thesis was rejected due to the influence of Marcellin Berthelot, as it contradicted, even though correctly, to his theories. Berthelot became French Education Minister, which did not make Duhem's life simpler. He became Professor for Theoretical Physics in Bordeaux.

only obtain one equation of state less than required to determine the thermodynamic potential. The Gibbs Duhem relation provides the entropy via

$$S = \sum_k X_k \frac{df_k}{dT}$$

and thus provides the extra required equation of state.

5.5 Thermodynamic potentials

The thermodynamic potentials can be converted into each other by Legendre transforms.

$$\begin{aligned} F &= U(S, X_i) + f_i X_i \\ dF &= dU + X_i df_i + f_i dX_i \\ &= \underbrace{T dS - \sum_{k=1}^{t} f_k dX_k}_{dU} + X_i df_i + f_i dX_i \\ &= T dS - \sum_{k \neq i} f_k dX_k + X_i df_i \end{aligned}$$

From this expression we see that the extensive variable of U, namely X_i, has been replaced by the corresponding intensive variable f_i in F. The new function does not depend any more on the extensive variable X_i. Similar to replacing X_i by f_i one can also replace S by T.

Note that the complete Legendre transform of all extensive variables is constant due to the Euler equation.

$$F(T, \{f_k\}) = U(S, X_i) \underbrace{-TS + \sum_k f_k X_k}_{-U(S, X_k)} \overset{Eq.\ 5.2}{=} 0$$

The thermodynamic potentials for a system described by volume, entropy and number of particles are described in section. **??**.

5.6 Maxwell relations

If the variables only change in a small window of values, it is often helpful to approximate the thermodynamic variables by a second order polynomial.

The second order coefficients are the response functions. Some of them we have already encountered in Section 2.5. Tabulated values of many response function can be found in the literature.

The idea of Maxwell relations[7] is that the mixed partial derivatives of a thermodynamic potential can be interchanged. The first derivatives are related to the corresponding conjugate variables. This shall be shown for one example.

$$\frac{\partial^2 U}{\partial S \partial V} = \frac{\partial^2 U}{\partial V \partial S} \qquad \overset{p=-\frac{\partial U}{\partial V};T=\frac{\partial U}{\partial S}}{\Rightarrow} \qquad -\left.\frac{\partial p}{\partial S}\right|_{V,N} = \left.\frac{\partial T}{\partial V}\right|_{S,N}$$

If one starts with a potential $U(S, V, N)$ which depends on three variables, there are 9 second derivatives, of which 6 are pairwise identical. Thus we obtain three Maxwell relations. For t variables of a thermodynamic potential we obtain $t(t-1)/2$ Maxwell relations.

Thus one obtains three Maxwell relations, namely

$$
\begin{aligned}
\frac{\partial^2 U}{\partial S\partial V}=\frac{\partial^2 U}{\partial V\partial S} \qquad &\Rightarrow \qquad -\left.\frac{\partial p}{\partial S}\right|_{V,N}=\left.\frac{\partial T}{\partial V}\right|_{S,N}\\
\frac{\partial^2 U}{\partial S\partial N}=\frac{\partial^2 U}{\partial N\partial S} \qquad &\Rightarrow \qquad \left.\frac{\partial \mu}{\partial S}\right|_{V,N}=\left.\frac{\partial T}{\partial N}\right|_{S,V}\\
\frac{\partial^2 U}{\partial V\partial N}=\frac{\partial^2 U}{\partial N\partial V} \qquad &\Rightarrow \qquad \left.\frac{\partial \mu}{\partial V}\right|_{S,N}=-\left.\frac{\partial p}{\partial N}\right|_{S,V}
\end{aligned}
$$

Similarly we can use any other thermodynamic potential to create three other Maxwell's relations. Note that there are only $t(t+1)/2$ independent second derivatives. Thus all response function can be traced back to a small number of basic response functions using Maxwell's relation.

Fundamental response functions

All response functions for a U, V system at constant N can be traced back to three basic response functions, namely the thermal expansion coefficient α, the specific heat at constant volume c_V,and the isothermal compressibility κ_T

$$
\begin{aligned}
\alpha_T &\overset{\text{Eq. 2.26}}{=} \frac{1}{V}\left.\frac{dV}{dT}\right|_{p,N}\\
c_V &\overset{\text{Eq. 2.29}}{=} \left.\frac{dU}{dT}\right|_{V,N}\\
\kappa_T &\overset{\text{Eq. 2.27}}{=} -\frac{1}{V}\left.\frac{dV}{dp}\right|_{T,N}
\end{aligned}
$$

For this purpose we expand the Helmholtz free energy $A(T, V, N)$ to second order in T and V about some reference state defined by T_0 and V_0. For this quadratic thermodynamic potential we can work out all response functions, by Legendre transform to a suitable thermodynamic potential, forming the equations of state, and then forming the desired derivative. The second derivatives in this Taylor expansion can then be identified with the above response functions as shown below.

$$
\begin{aligned}
A(T,V,N) &= A_0 - S_0(T-T_0) - p_0(V-V_0)\\
&\quad -\frac{c_V}{2T_0}(T-T_0)^2 - \frac{\alpha}{\kappa_T}(T-T_0)(V-V_0) - \frac{1}{2V_0\kappa_T}(V-V_0)^2\\
&\quad +O(T-T_0, V-V_0)^3)\\
&= \left[A_0 + S_0T_0 + p_0V_0 - \frac{c_V}{2T_0}T_0^2 - \frac{\alpha}{\kappa_T}T_0V_0 - \frac{1}{2V_0\kappa_T}V_0^2\right]\\
&\quad + \left[-S_0 + \frac{c_V}{T_0}T_0 + \frac{\alpha}{\kappa_T}V_0\right]T + \left[-p_0 + \frac{\alpha}{\kappa_T}T_0 + \frac{1}{V_0\kappa_T}V_0\right]V\\
&\quad -\frac{c_V}{2T_0}T^2 - \frac{\alpha}{\kappa_T}TV - \frac{1}{2V_0\kappa_T}V^2\\
&\quad +O(T-T_0, V-V_0)^3)
\end{aligned}
$$

In the following we will show that the second derivatives of the Helmholtz potential can be represented by α, κ_T and c_V as

$$\frac{\partial^2 A}{\partial T^2} = -\frac{c_V}{T} \quad ; \quad \frac{\partial^2 A}{\partial T \partial V} = -\frac{\alpha}{\kappa_T} \quad ; \quad \frac{\partial^2 A}{\partial V^2} = +\frac{1}{V\kappa_T}$$

We begin with relating specific heat at constant volume, c_V, to the second derivatives of A:

$$U(T,V,N) = A(T,V,N) - T\underbrace{\frac{\partial A(T,V,N)}{\partial T}}_{-S(T,V,N)}$$

$$c_V \overset{\text{Eq. 2.29}}{=} \left.\frac{dU}{dT}\right|_{V,N} = \frac{\partial A}{\partial T} - \frac{\partial A}{\partial T} - T\frac{\partial^2 A}{\partial T^2} \quad \Rightarrow \quad \frac{\partial^2 A}{\partial T^2} = -\frac{c_V}{T} \tag{5.4}$$

Then we relate the isothermal compressibility κ_T to the second derivatives of A

$$-V\kappa_T \overset{\text{Eq. 2.27}}{=} \left.\frac{dV}{dp}\right|_{T,N} = \left(\left.\frac{dp}{dV}\right|_{T,N}\right)^{-1} = \left(\frac{d}{dV}\underbrace{\left(-\frac{dA}{dV}\right)}_{p}\right)^{-1} = \left(-\frac{\partial^2 A}{\partial V^2}\right)^{-1}$$

$$\Rightarrow \quad \frac{\partial^2 A}{\partial V^2} = \frac{1}{V\kappa_T} \tag{5.5}$$

cont'd

cont'd

Since we need to understand the system at constant pressure and number of particles, we express the coefficient of thermal expansion α_T by the Gibbs free energy $G(T,p,N)=A(T,V,N)+pV$

$$G(T,p,N)=\min_V\Big[A(T,V,N)+pV\Big]=A(T,V^0,N)+pV^0 \tag{5.6}$$

$$\frac{\partial G}{\partial p}=\underbrace{\left[\frac{\partial A}{\partial V}+p\right]}_{=0}\frac{\partial V^0}{\partial p}+V^0=V^0 \tag{5.7}$$

where $V^0(T,p,N)$ is the volume, which maximizes $A(T,V,N)+pV$ for a given T,p,N.

$$\begin{aligned}
V\alpha_T &\overset{\text{Eq. 2.26}}{=} \left.\frac{dV}{dT}\right|_{p,N} \overset{\text{Eq. 5.7}}{=} \frac{\partial^2 G}{\partial T\partial p}=\frac{\partial^2 G}{\partial p\partial T}\\
&= \frac{d}{dp}\frac{d}{dT}\Big(A(T,V^0(T,p,N),N)+pV^0(T,p,N)\Big)\\
&\overset{\text{Eq. 5.6}}{=} \frac{d}{dp}\left[\frac{\partial A(T,V^0,N)}{\partial T}+\left(\frac{\partial A(T,V^0,N)}{\partial V}+p\right)\frac{dV^0(T,p,N)}{dT}\right]\\
&= \frac{\partial^2 A}{\partial V\partial T}\underbrace{\frac{\partial V^0}{\partial p}}_{-V\kappa_T}+\underbrace{\left(\underbrace{\frac{\partial^2 A}{\partial V^2}}_{+\frac{1}{V\kappa_T}}\underbrace{\frac{\partial V^0}{\partial p}}_{-V\kappa_T}+1\right)}_{=0}\underbrace{\frac{\partial V^0}{\partial T}}_{-V\alpha}+\underbrace{\left(\frac{\partial A}{\partial V}+p\right)}_{=0}\frac{d^2V^0}{\partial T\partial p}\\
&= -V\kappa_T\frac{\partial^2 A}{\partial V\partial T}\\
\Rightarrow\frac{\partial^2 A}{\partial V\partial T} &= -\frac{\alpha}{\kappa_T}
\end{aligned}$$

With this we obtained all the second derivatives of the Helmholtz potential. *q.e.d.*

5.7 Summary

Important equations

$$\begin{aligned}
U &= TS-\sum_k f_k X_k\\
0 &= SdT-\sum_k X_k df_k
\end{aligned}$$

5.8 Exercise

Legendre transforms

Use the quadratic form for the Helmholtz free energy to calculate the corresponding internal energy, and express the second derivatives of the internal energy by α, κ_T and c_V.

This exercise shall train to perform the Legendre transforms.

Maxwell relations

Use the quadratic form for the Helmholtz free energy to calculate the

- adiabatic bulk modulus defined as

$$\beta_S = -V \left.\frac{dp}{dV}\right|_{S,N}$$

- and the specific heat at constant pressure

$$c_p = -V \left.\frac{dS}{dT}\right|_{p,N}$$

and express it by α, κ_T and c_V.

This exercise shall train selecting the correct thermodynamic potential, to perform Legendre transform to exploit maxwell's relation to obtain arbitrary response functions.

Determine the entropy of the Boltzmann gas from experiment

The molecules in most gases interact only weakly with each other. Therefore, their thermodynamic properties can be described well by a model of a gas of completely non-interacting molecules.

The ideal monatomic gas is characterized by two equations of state, which connect the equilibrium values of intrinsic and extrinsic variables.

$$\begin{aligned} PV &= NRT \\ U &= \frac{3}{2}NRT \end{aligned}$$

where $R = k_B/mole$ is the ideal gas constant. We assume that the two equations of state have been determined experimentally. Apparently we have three sets of variables, namely (U, T), (V, P) and (N, μ). Therefore the entropy depends on three extrinsic variables (U, V, N) However, only two equations of state are known. How do we obtain the entropy?

Let us bring the equations of state in a standard form

$$\begin{aligned} \frac{P}{T} &= R\frac{N}{V} \\ \frac{1}{T} &= \frac{3}{2}R\frac{N}{U} \end{aligned}$$

We are missing the third equation of state for $\frac{\mu}{T}$. Here we can use the Gibbs Duhem equation, which determines one intensive variable from all the others.

$$\begin{aligned} 0 &= Ud\frac{1}{T} + Vd\frac{P}{T} - Nd\frac{\mu}{T} \\ d\frac{\mu}{T} &= \frac{U}{N}d\frac{1}{T} + \frac{V}{N}d\frac{P}{T} = \frac{3}{2}R\frac{U}{N}d\frac{N}{U} + R\frac{V}{N}d\frac{N}{V} \\ &= \frac{3}{2}Rd\ln[\frac{N}{U}] + Rd\ln[\frac{N}{V}] \\ \frac{\mu}{T} &= -\frac{3}{2}R\ln[\frac{U}{N}] - R\ln[\frac{V}{N}] + C \end{aligned}$$

Thus we obtained the third equation of state up to an unknown constant C.

By inserting the three equations of state into the Euler equation, we obtain

$$\begin{aligned}
S(U,V,N) &= \frac{U}{T} + \frac{P}{T}V - \frac{\mu}{T}N \\
&= \Big[\frac{3}{2}R\frac{N}{U}\Big]U + \Big[R\frac{N}{V}\Big]V + \Big[-\frac{3}{2}R\ln[\frac{U}{N}] - R\ln[\frac{V}{N}] + C\Big]N \\
&= RN\Big[\frac{5}{2} - \frac{3}{2}\ln[\frac{U}{N}] - \ln[\frac{V}{N}] + C\Big] \\
&= RN\Big[\frac{5}{2} + C - \ln[\Big(\frac{U}{N}\Big)^{-\frac{3}{2}}\Big(\frac{V}{N}\Big)^{-1}]\Big] \\
&= RN\Big[\frac{5}{2} + C + \ln[U^{\frac{3}{2}}VN^{-\frac{5}{2}}]\Big]
\end{aligned}$$

We express C by the entropy of some reference state

$$S_0 = RN_0\Big[\frac{5}{2} + C + \ln[U_0^{\frac{3}{2}}V_0N_0^{-\frac{5}{2}}]\Big]$$

and obtain

$$S(U,V,N) = \frac{N}{N_0}S_0 + NR\ln[\Big(\frac{U}{U_0}\Big)^{\frac{3}{2}}\Big(\frac{V}{V_0}\Big)\Big(\frac{N}{N_0}\Big)^{-\frac{5}{2}}]$$

Note that the entropy violates the third law of thermodynamics. This is because the equations of state, we used as starting point, are not correct at low temperature.

Chapter 6

The Language of Statistical Physics

What we described in the previous section relied on the existence of discrete states with equal a-priori probabilities. Because quantum mechanical systems are quantized, these preconditions are valid and we can apply the power of information theory developed sofar. The statistical description of classical mechanics can be derived rigorously out of a quantum mechanical description, which is shown in a later section.

Fig. 6.1: John von Neumann, 1903-1957.

6.1 The state operator

The fundamental quantity in statistical physics is the **state operator**[8], introduced by John von Neumann.[1] As we describe a classical system by the coordinates and momenta of all the particles, or a quantum mechanical state by a wave function, a statistical quantum mechanical state is described by the state operator $\hat{\rho}$. Once the state operator is known, we can in principle derive all observable quantities from it.

Consider a quantum mechanical system described by a single, well defined (many-particle) state $|\Psi\rangle$. Such a system is called a **pure state** or **microstate**. We can describe a quantum mechanical state completely by the density operator

$$\hat{\rho} = |\Psi\rangle\langle\Psi|$$

The expectation value of some observable A is then obtained as the trace[2] of the product of the density operator and the observable.[3]

$$\langle A\rangle = \langle\Psi|\hat{A}|\Psi\rangle = \text{Tr}[|\Psi\rangle\langle\Psi|\hat{A}] = \text{Tr}[\hat{\rho}\hat{A}]$$

---

[1] John von Neumann, 1903-1957 hungarian-born american mathematician. He made major contributions to mathematics, physics, computer science and economics. He gave quantum mechanics a sound basis by introducing the operator theory.

[2] The German word for "Trace" is "Spur". If A is a matrix, $\text{Tr}[A] = \sum_i A_{i,i}$. The trace of an operator is defined by the sum of its diagonal matrix elements in any orthonormal basis, i.e. $Tr[\hat{A}] = \sum_k \langle\chi_k|\hat{A}|\chi_k\rangle$. (It can be shown that a change of the representation , i.e. a unitary transformation of the states leaves the value of the trace unchanged.)

[3] The trace of an operator $|\psi\rangle\langle\phi|$ can be written as $Tr[|\psi\rangle\langle\phi|] = \langle\phi|\psi\rangle$. To show this we introduce a complete orthogonal basis $|u_n\rangle$.

$$Tr[|\psi\rangle\langle\phi|] = \sum_n \langle u_n|\psi\rangle\langle\phi|u_n\rangle = \sum_n \langle\phi| \underbrace{\left(\sum_n |u_n\rangle\langle u_n|\right)}_{=1} |\psi\rangle = \langle\phi|\psi\rangle$$

In the example we can set $|\phi\rangle = \hat{A}|\psi\rangle$.

If we do not have complete knowledge of the system, we may describe the system by the probabilities P_i for the system to be in a particular quantum mechanical states $|\Psi_i\rangle$. Such a probabilistic description of the system is called a **statistical mixture** or a **macrostate**.

In order to describe a statistical mixture of quantum mechanical states we can sum the density operators of the microstates, weighted with their probabilities p_i, to obtain the density operator.

DENSITY OPERATOR

, **State operator**, also called the **statistical density operator**, **density matrix**, or **statistical operator**

$$\hat{\rho} = \sum_i P_i|\Psi_i\rangle\langle\Psi_i| = \sum_i |\Psi_i\rangle P_i\langle\Psi_i| \tag{6.1}$$

The probabilities P_i are positive $P_i \geq 0$ and must sum up to one $\sum_i P_i = 1$. The states are normalized, but are otherwise arbitrary. Each state is a quantum state of the entire system. Usually the states are many-particle states and no one-particle states.
Im Deutschen sind die Begriffe Dichtematrix oder Dichteoperator gebräuchlich.

The statistical expectation value of an observable $\hat{A}$ is obtained from the statistical density operator just as it is obtained from the density operator of a microstate:

$$\langle A\rangle = \sum_i P_i\langle\Psi_i|\hat{A}|\Psi_i\rangle = \mathrm{Tr}[|\Psi_i\rangle P_i\langle\Psi_i|\hat{A}] = \mathrm{Tr}[\hat{\rho}\hat{A}] \tag{6.2}$$

The operator expression is useful, because it is independent of the representation. If convenient, we can introduce at any point a suitable orthonormal basis.

- For example, if we choose eigenstates[4] $|a_i\rangle$ of $\hat{A}$ with eigenvalues a_i we only need to evaluate the diagonal elements of the state operator in this representation.

  $$\langle\hat{A}\rangle = Tr[\hat{\rho}\hat{A}] = \sum_{i,j}\langle a_i|\hat{\rho}|a_j\rangle \underbrace{\langle a_j|\hat{A}|a_i\rangle}_{a_i\delta_{i,j}} = \sum_i\langle a_i|\hat{\rho}|a_i\rangle a_i$$

  Thus we obtain the expectation value as a weighted sum of eigenvalues of $\hat{A}$. In this case the matrix element $\langle a_i|\hat{\rho}|a_i\rangle$ is the probability of the eigenvalue a_i.

- On the other hand we may also use eigenstates $|\Phi_i\rangle$ of the density operator with eigenvalues[5][6] p_i and obtain

  $$\langle\hat{A}\rangle = Tr[\hat{\rho}\hat{A}] = \sum_{i,j}\langle\Phi_i|\hat{\rho}|\Phi_j\rangle\langle\Phi_j|\hat{A}|\Phi_j\rangle = \sum_i p_i\langle\Phi_i|\hat{A}|\Phi_i\rangle$$

  We can interpret the eigenvalues p_i as the probability for the system to be in one of the eigenstates $|\Phi_i\rangle$ of the density operator.

---

[4] We use that $\sum_i |a_i\rangle\langle a_i| = \hat{1}$ is the unit operator.

[5] The eigenstates are defined by

$$\hat{\rho}|\Phi_i\rangle = |\Phi_i\rangle p_i$$

Note that the eigenstates are many particle states are many particle states and may not even have sharp values for the particle number.

[6] If the states in Eq. 6.1 are already orthogonal, these states are automatically eigenstates of the density operator. Their eigenvalue is P_i. The states orthogonal to those explicitely included in the sum Eq. 6.1 have a zero eigenvalue.

PROPERTIES OF THE STATISTICAL DENSITY OPERATOR

The statistical density operator must have three properties

1. The density operator is hermitian
$$\hat\rho=\hat\rho^\dagger$$
2. Its trace is one
$$\mathrm{Tr}[\hat\rho]=1$$
which is the normalization condition.
3. The probability for the system to be in one state is non-negative, that is the density operator is non-negative
$$\forall_{|\chi\rangle}\quad \langle\chi|\hat\rho|\chi\rangle\geq 0$$
4. If the statistical operator describes a single quantum state, it is **idempotent**, that is
$$\hat\rho^2=\hat\rho$$
and if the density operator is idempotent, it can be described by a microstate. If the density operator is not idempotent it describes a statistical mixture of several quantum states.

In the following we will prove these statements based on Eq. 6.1:

1. The statement $\hat\rho=\hat\rho^\dagger$ follows from the fact that it is the sum of hermitian operators[7] $|\Psi_i\rangle\langle\Psi_i|$ with real coefficients.
$$\begin{aligned}\langle\phi|\hat\rho|\chi\rangle &= \langle\phi|\sum_i|\Psi_i\rangle P_i\langle\Psi_i|\chi\rangle=\sum_i\langle\phi|\Psi_i\rangle P_i\langle\Psi_i|\chi\rangle=\sum_i\langle\chi|\Psi_i\rangle^*P_i\langle\Psi_i|\phi\rangle^*\\ &\overset{P_i=P_i^*}{=} \Big[\langle\chi|\sum_i|\Psi_i\rangle P_i\langle\Psi_i|\phi\rangle\Big]^*=\Big[\langle\chi|\hat\rho|\phi\rangle\Big]^*\\ \Rightarrow\hat\rho &= \hat\rho^\dagger\end{aligned}$$
The property that the operator is hermitian has the important consequence that the density operator can be diagonalized such that the eigenvalues p_i are real and the eigenstates $|\Phi_i\rangle$ are orthonormal (respectively, they can be orthonormalized).
$$\hat\rho|\Phi_i\rangle=|\Phi_i\rangle p_i\qquad\Leftrightarrow\qquad\hat\rho=\sum_i|\Phi_i\rangle p_i\langle\Phi_i|$$
The eigenvalues p_i can be identified with the probability that the system is in the eigenstate $|\Phi_i\rangle$. Note that the eigenstates $|\Phi_i\rangle$ may be different from the states $|\Psi_i$ from which the density operator has been constructed.
2. The statement $Tr[\hat\rho]=1$ follows from the requirement that the weights P_i of the contributing wave functions sum up to one.
$$Tr[\hat\rho]=Tr[\sum_i|\Psi_i\rangle P_i\langle\Psi_i|]=\sum_i P_i\underbrace{Tr[|\Psi_i\rangle\langle\Psi_i|]}_{=1}=\sum_i P_i=1$$

---

[7] The hermitean conjugate $\hat A^\dagger$ of an operator $\hat A$ is defined by the identity $\langle\phi|\hat A^\dagger|\chi\rangle=[\langle\chi|\hat A|\phi\rangle]^*$ for an arbitray pair of states. An operator is hermitean, when it is identical to its hermitean conjugate.

3. The statement $\langle\Phi|\hat{\rho}|\Phi\rangle \geq 0$ follows from the positiveness of the weights P_i of the states that contributed to the density operator.

$$\langle\Phi|\hat{\rho}|\Phi\rangle = \sum_i \langle\Phi|\Psi_i\rangle P_i \langle\Psi_i|\Phi\rangle = \sum_i |\langle\Phi|\Psi_i\rangle|^2 P_i \overset{P_i\geq 0}{\geq} 0$$

4. We can differentiate a pure state from a statistical mixture by testing the **idem-potency condition**

$$\hat{\rho}^2 = \hat{\rho}$$

If the idem-potency is fulfilled, the system is in a pure state and can always be described by a single quantum state. Otherwise it is in a statistical mixture.

Let us diagonalize the density operator and thus obtain its eigenvalues p_i and eigenstates $|\Phi_i\rangle$. Then we can write the density operator as

$$\hat{\rho} = \sum_i |\Phi_i\rangle p_i \langle\Phi_i|$$

and from that calculate its square

$$\hat{\rho}^2 = \sum_{i,j} |\Phi_i\rangle p_i \langle\Phi_i|\Phi_j\rangle p_j \langle\Phi_j| = \sum_i |\Phi_i\rangle p_i^2 \langle\Phi_i|$$

From the idem-potency condition $\hat{\rho}^2 = \rho$ and the linear independence of the eigenstates we thus obtain

$$p_i^2 = p_i$$

The equation $x^2 = x$ has only two possible solutions, namely $x = 0$ and $x = 1$. Hence p_i of an idempotent density operator can be either zero or one. Since the probabilities sum up to one and are non-negative, only the probability for a single state may be one, while all others vanish. Thus the system can be described by a single quantum state, namely the eigenstate with eigenvalue $p_i = 1$.

6.2 Reduced one-particle density matrix

A common source of confusion is the similarity of the density operator and the reduced one-particle density operator. In solid state physics one often uses the reduced one-particle density operator for many-electron systems. It is typically called density matrix without further specification.

Let us introduce here a notation that allows us to treat spatial and spin coordinates on the same footing and thus simplify the expressions: We introduce a new variable

$$\vec{x} \overset{\text{def}}{=} (\vec{r}, \sigma)$$

which consists of the three dimensional continuous variable $\vec{r}$ and the discrete spin variable σ. For electrons, σ can assume the values ↑ and ↓, which correspond to the two possible eigenvalues of the $\hat{S}_z$ spin operator in z-direction. An integral over all states will simply be written as

$$\int d^4x \overset{\text{def}}{=} \sum_\sigma \int d^3r$$

The reduced one-particle density matrix $\hat{\rho}^{(1)}$ for a system of indistinguishable particles can be constructed from the full density operator $\hat{\rho}$ as follows:

ONE-PARTICLE REDUCED DENSITY MATRIX

The one-particle reduced density operator $\hat{\rho}^{(1)}$ is defined as

$$\hat{\rho}^{(1)} \overset{\text{def}}{=} \int d^4x \int d^4x' \; |\vec{x}\rangle \; \rho^{(1)}(\vec{x},\vec{x'}) \; \langle\vec{x'}| \tag{6.3}$$

$$\hat{\rho}^{(1)}(\vec{x},\vec{x'}) \overset{\text{def}}{=} \sum_{N=0}^{\infty} N \int d^4x_2 \cdots \int d^4x_N \; \langle\vec{x},\vec{x}_2,\dots,\vec{x}_N|\hat{\rho}|\vec{x'},\vec{x}_2,\dots\vec{x}_N\rangle \tag{6.4}$$

where $\rho(\vec{x},\vec{x'})$ is the one-particle reduced density matrix. The terms "density matrix" and "density operator" are often used synonymous. $\hat{\rho}$ is the complete density operator.
The expression contains bra's and ket's with different number of arguments. This is because we need to use states that span the entire Fock-space, which includes states with various particle numbers. States with different particle numbers are orthonormal.[a]
In Eq. 6.4 we exploited that the contribution to the one-particle density matrix from each particle is identical to that of each other. Therefore we have singles out the first particle, and then we multiplied the result by the particle number N.

---

[a] The orthonormality of states with different particle numbers follows from the fact that the states are eigenstates of a hermitean particle number operator. In ΦSX: Quantum Physics, we have shown that eigenstates for an hermitean operator with different eigenvalues are orthogonal.

The one-particle reduced density operator is an operator in the one-particle Hilbert space. The one-particle reduced density operator is hermitean. Therefore, its eigenstates $|\phi_i\rangle$ can be chosen orthonormal and its eigenvalues f_i are real.

$$\hat{\rho}^{(1)}|\phi_i\rangle = |\phi_i\rangle f_i$$

The eigenvalues are the **occupations** of the orbitals and the eigenstates of the one-particle reduced density operator are the so-called **natural orbitals**. With the eigenvalues and eigenstates, we can respresent $\hat{\rho}^{(1)}$ as

$$\hat{\rho}^{(1)} = \sum_i |\phi_i\rangle f_i \langle\phi_i|$$

This expression looks fairly similar to the expression for the full density operator. However, the states $|\phi_i\rangle$ are one-particle states that can be represented by wave functions with a single three-dimensional spatial coordinate and a double values spin index. In the expression for the full density operator the states $|\Psi_i\rangle$ are many-particle states from the Fock space. The occupartions add upt to the expectation of the particle number $\langle N\rangle = \sum_i f_i$, while the probabilities in the expression for the full density operator add up to one, i.e. $\sum_i P_i = 1$.

The expectation values of *one-particle* operators can be obtained from the one-particle reduced density matrix as

$$\langle A\rangle = \mathrm{Tr}[\hat{\rho}^{(1)}\hat{A}]$$

The result is the expectation value of $\hat{A}$ for all particles. This expression, however, does not work with many-particle operators such as the Coulomb interaction between electrons.

6.3 Entropy functional

Now we can define the entropy functional analogous to the principles derived in Chapter 1.

ENTROPY FUNCTIONAL

The entropy is defined[8] as

$$\mathcal{S}[\hat{\rho}] = -k_B \mathrm{Tr}[\hat{\rho} \ln[\hat{\rho}]] \tag{6.5}$$

This quantum mechanical expression has been named "von Neumann entropy".

The equivalence of this definition of the entropy with the one we learned sofar, $S = -k \sum_x P_x \ln(P_x)$ is shown as follows:

Let me first remind of a fundamental operator identity: The function $f(\hat{A}) = \sum_i |a_i\rangle f(a_i)\langle a_i|$ of an operator $\hat{A} = \sum_i |a_i\rangle a_i \langle a_i|$ is defined in a representation of the eigenstates $|a_i\rangle$, i.e. $\hat{A}|a_i\rangle = |a_i\rangle a_i$, as the operator formed when each eigenvalue is replaced by the function of its eigenvalue.

With this in mind, we transform into a representation of eigenstates $|\Phi_i\rangle$ of the density operator with eigenvalues p_i.

$$\begin{aligned}
-k_B \mathrm{Tr}[\hat{\rho}\ln(\hat{\rho})] &= -k_B \mathrm{Tr}\left[\sum_i |\Phi_i\rangle p_i \langle\Phi_i| \ln[\sum_j |\Phi_j\rangle p_i \langle\Phi_j|]\right] \\
&= -k_B \mathrm{Tr}[\sum_i |\Phi_i\rangle p_i \langle\Phi_i| \sum_j |\Phi_j\rangle \ln[p_i] \langle\Phi_j|] \\
&= -k_B \sum_k \sum_{ij} \underbrace{\langle\Phi_k|\Phi_i\rangle}_{\delta_{k,i}} p_i \underbrace{\langle\Phi_i|\Phi_j\rangle}_{\delta_{i,j}} \ln[p_j] \underbrace{\langle\Phi_j|\Phi_k\rangle}_{\delta_{j,k}} \\
&= -k_B \sum_i p_i \ln[p_i]
\end{aligned}$$

6.3.1 Maximum-entropy principle

We can now formulate the maximum-entropy principle for quantum systems: In thermal equilibrium, the density matrix minimizes the entropy under the constraints imposed. According to the method of Lagrange multipliers, we need to find the extremum of the entropy augmented by the constraint terms.

$$\begin{aligned}
\mathcal{S}[\hat{\rho}, U, \{\bar{X}_k\}] &= -k_B \mathrm{Tr}[\hat{\rho} \ln \hat{\rho}] - \lambda\Big(\mathrm{Tr}[\hat{\rho}] - 1\Big) &(6.6)\\
&\quad -\frac{1}{T}\Big(\mathrm{Tr}[\hat{\rho}\hat{H}] - \bar{U}\Big) - \sum_k \frac{f_k}{T}\Big(\mathrm{Tr}[\hat{\rho}\hat{X}_k] - \bar{X}_k\Big) &(6.7)
\end{aligned}$$

with respect to $\hat{\rho}, \lambda, T, f_1, f_2, \ldots$.

Setting the variation with respect to the density matrix to zero yields the equilibrium condition. Note that we have generalized the concept of a functional to that of a function of an operator! The derivative with respect to an operator is an object, which uses an additional level of abstraction. Hence we need to be extra careful about the operations that we are allowed to perform.

This is why we extract the derivative from the variation of the entropy with respect to a change of $\hat{\rho}$. Before we continue, let us investigate a fundamental operation that will be needed later. Let us consider the variation of a trace of a function of an operator.

$$d\mathrm{Tr}[f(\hat{A})] = Tr[\left.\frac{df}{dx}\right|_{\hat{A}} d\hat{A}] \tag{6.8}$$

This identity only holds for the trace and not for the operators themselfes.

In order to prove this identity Eq. 6.8, let us begin with a simple power of an operator, i.e. $f(x) = x^n$.

$$\begin{aligned}
d\mathrm{Tr}[\hat{A}^n] &= \mathrm{Tr}[d\hat{A}\hat{A}^{n-1} + \hat{A}d\hat{A}\hat{A}^{n-1} + \ldots + \hat{A}^{n-2}d\hat{A}\hat{A} + \hat{A}^{n-1}d\hat{A}] \\
&= \mathrm{Tr}[d\hat{A}\hat{A}^{n-1}] + \mathrm{Tr}[\hat{A}d\hat{A}\hat{A}^{n-1}] + \ldots + \mathrm{Tr}[\hat{A}^{n-2}d\hat{A}\hat{A}] + Tr[\hat{A}^{n-1}d\hat{A}] \\
&\overset{\text{cycl. perm}}{=} \underbrace{\mathrm{Tr}[\hat{A}^{n-1}d\hat{A}] + \mathrm{Tr}[\hat{A}^{n-1}d\hat{A}] + \ldots + \mathrm{Tr}[\hat{A}^{n-1}d\hat{A}] + \mathrm{Tr}[\hat{A}^{n-1}d\hat{A}]}_{n-\text{times}} \\
&= \mathrm{Tr}[n\hat{A}^{n-1}d\hat{A}] = \mathrm{Tr}[\left.\frac{dx^n}{dx}\right|_{\hat{A}} d\hat{A}]
\end{aligned}$$

It is a simple matter to extend this result to a power-series expansion, and thus to an arbitrary function, which leads to Eq. 6.8.

Now we return to our original problem, namely to determine the variation of the entropy Eq. 6.7 with respect to the density operator. We use Eq. 6.8 for the function $f(x) = x\ln[x]$ and obtain

$$0 = d\mathcal{S} = \mathrm{Tr}\left[\left(-k_B\Big(\ln[\hat{\rho}] + \hat{1}\Big) - \lambda\hat{1} - \frac{1}{T}\hat{H} - \sum_k \frac{f_k}{T}\hat{X}_k\right) d\hat{\rho}\right]$$

The requirement that the variation of the entropy vanishes for any variation of the density operator implies that the prefactor in the trace vanishes itself. Hence,

$$\begin{aligned}
&\Rightarrow \quad 0 = \ln[\hat{\rho}] + \hat{1} + \frac{\lambda}{k_B}\hat{1} + \frac{1}{k_B T}\Big(\hat{H} + \sum_k f_k\hat{X}_k\Big) \\
&\Rightarrow \quad \hat{\rho} = \mathrm{e}^{-1-\frac{\lambda}{k_B}}\mathrm{e}^{-\frac{1}{k_B T}\left(\hat{H} + \sum_k f_k\hat{X}_k\right)}
\end{aligned}$$

The normalization factor $\mathrm{e}^{-1-\lambda/k_B}$ can be identified, via the constraint condition $\mathrm{Tr}[\hat{\rho}] = 1$, with the inverse $Z^{-1}(T, \{f_k\})$ of the partition function.

Thus we can summarize our results as follows:

QUANTUM EXPRESSIONS FOR PARTITION FUNCTION, STATE OPERATOR AND FREE ENERGY

The partition function has the form

$$Z(T, \{f_k\}) = \mathrm{Tr}\left[\mathrm{e}^{-\frac{1}{kT}(\hat{H} + \sum_k f_k\hat{X}_k)}\right] \tag{6.9}$$

The state operator in thermal equilibrium has the form

$$\hat{\rho}^{\mathrm{eq}}(T, \{f_k\}) = \frac{1}{Z(T, \{f_k\})}\mathrm{e}^{-\frac{1}{kT}(\hat{H} + \sum_k f_k\hat{X}_k)} = \mathrm{e}^{-\frac{1}{kT}(\hat{H} + \sum_k f_k\hat{X}_k - F(T, \{f_k\}))} \tag{6.10}$$

The free energy has the form

$$F(T, \{f_k\}) = -k_B T\ln[Z(T, \{f_k\})] = -k_B T\ln\left[\mathrm{Tr}[\mathrm{e}^{-\frac{1}{kT}(\hat{H} + \sum_k f_k\hat{X}_k)}]\right] \tag{6.11}$$

Once the free energy is known, we can proceed as before and calculate the equations of state, which yields the mean values

$$\begin{aligned}
U(T,\{f_k\}) &= \langle E\rangle_{T,f_k} = \mathrm{Tr}[\hat{H}\,\hat{\rho}^{\mathrm{eq}}(T,\{f_k\})] \\
&\overset{Eq.~1.67}{=} F(T,\{f_k\}) + TS(T,\{f_k\}) - \sum_j f_j X_j(T,\{f_k\}) \\
\bar{X}_j(T,\{f_k\}) &= \langle X\rangle_{T,f_k} = \mathrm{Tr}[\hat{X}_j\,\hat{\rho}^{\mathrm{eq}}(T,\{f_k\})] \overset{Eq.~1.66}{=} \frac{\partial F(T,f_k)}{\partial f_i} \\
S(T,\{f_k\}) &= -k_B \mathrm{Tr}\left[\hat{\rho}^{\mathrm{eq}}(T,\{f_k\})\ln\left(\hat{\rho}^{\mathrm{eq}}(T,\{f_k\})\right)\right] \overset{Eq.~1.65}{=} -\frac{\partial F(T,\{f_k\})}{\partial T}
\end{aligned}$$

as function of the intensive variables.

6.3.2 Mean values

Above we obtained the relation of the free energy to the mean values in analogy to Eqs. 1.67, 1.65, and 1.66. Here I will show this connection in a more explicit form.

Expectation values of an observable $\hat{X}$ are obtained from the density operator as

$$\langle X\rangle \overset{\text{Eq. 6.2}}{=} \mathrm{Tr}[\hat{\rho}\hat{X}] \tag{6.12}$$

The expectation value of X in thermal equilibrium can alternatively be derived from the free energy obtained from an **extended Hamiltonian**. The resulting expression

$$\langle\hat{X}\rangle_T = \left.\frac{\partial F(T,f)}{\partial f}\right|_{f=0} \tag{6.13}$$

is completely analogous to the equation of state Eq. 1.66 on p. 39. Here, Eq. 6.13 will be derived from the density operator formalism in its full generality.

The identity above is often used as follows: if one is interested in a particular expectation value $\langle X\rangle$, we simply add an additional term $f\hat{X}$ to the Hamiltonian. The resulting extended Hamiltonian has the form

$$\hat{H}(f) = \hat{H}_0 + f\hat{X}$$

The extended Hamiltonian is not necessarily physical. Rather the additional term can be regarded as a mathematical trick.

Then one works out the free energy $F(T,f)$ as function of f for the extended Hamiltonian. Finally we form the derivative with respect to this f and evaluate it for $f=0$.

$$\langle X\rangle = \left.\frac{\partial F}{\partial f}\right|_{f=0}$$

At first sight, using Eq. 6.13 to obtain an expectation value seems more complicated than using Eq. 6.2 directly. However, it is often easier to determine the free energy with an additional variable than to work out the trace in Eq. 6.2.

The free energy $F(T)$ depends on the temperature and the Hamiltonian. The Hamiltonian describes the system completely, but it is independent from the statistics. Let us modify the Hamilton operator by coupling the system via some operator $\hat{X}_k$ to an external variable f_k so that [8]

$$\hat{H}(\{f_k\}) = \hat{H}_0 + \sum_j f_j\hat{X}_j \tag{6.16}$$

---

[8] There is a subtle point that needs attention:

- We can consider the terms $\sum_k f_k X_k$ either as part of the Hamiltonian $\hat{H}(\{f_k\})$. In this formulation, the internal energy is the expectation value of $\hat{H}(\{f_k\})$ and it is obtained from the free energy as

$$U_f(T,\{f_k\}) = F(T,\{f_k\}) + TS(T,\{f_k\}) \tag{6.14}$$

- Or we can consider the terms $-\sum_k f_k X_k$ as one part of the constraint terms $-\frac{1}{T}[(E_{0,\vec{n}}-U_0)-\sum_k f_k(X_{k,\vec{n}}-\bar{X}_k)]$

Note here that the internal energy is the expectation value of $\hat{H}_0$, while the expectation value of $\hat{H}(\{f_k\})$ is

$$\langle\hat{H}(\{f_k\})\rangle_{T,\{f_k\}}=U(T,\{f_k\})+\sum_k f_kX_k(T,\{f_k\})$$

We can represent the Hamiltonian $\hat{H}(\{f_k\})$ by its eigenstates $|\psi_{\vec{n}}(\{f_k\})\rangle$ with eigenvalues $E_{\vec{n}}(\{f_k\})$ as

$$\hat{H}(\{f_k\})=\sum_{\vec{n}}\Big|\psi_{\vec{n}}(\{f_k\})\Big\rangle E_{\vec{n}}(\{f_k\})\Big\langle\psi_{\vec{n}}(\{f_k\})\Big|$$

This allows us to write the free energy in the conventional form

$$\begin{aligned}F(T,\{f_k\})&=-k_BT\ln\Big[\mathrm{Tr}[\mathrm{e}^{-\frac{1}{k_BT}(\hat{H}_0+\sum_k f_k\hat{X}_k)}]\Big]\\&=-k_BT\ln\Big[\mathrm{Tr}[\mathrm{e}^{-\frac{1}{k_BT}\hat{H}(\{f_k\})}]\Big]=-k_BT\ln\Big[\sum_{\vec{n}}\mathrm{e}^{-\frac{E_{\vec{n}}(\{f_k\})}{k_BT}}\Big]\end{aligned}$$

Let us now evaluate the derivative of the free energy with respect to one of the external parameters.

$$\begin{aligned}\frac{\partial F}{\partial f_k}&=-k_BT\frac{1}{\sum_{\vec{n}}\mathrm{e}^{-\frac{E_{\vec{n}}(\{f_k\})}{k_BT}}}\sum_{\vec{n}}\left(-\frac{1}{k_BT}\frac{\partial E_{\vec{n}}(\{f_k\})}{\partial f_k}\mathrm{e}^{-\frac{E_{\vec{n}}(\{f_k\})}{k_BT}}\right)\\&=\sum_{\vec{n}}\frac{\partial E_{\vec{n}}(\{f_k\})}{\partial f_k}\frac{1}{\sum_{\vec{n}}\mathrm{e}^{-\frac{E_{\vec{n}}(\{f_k\})}{k_BT}}}\mathrm{e}^{-\frac{E_{\vec{n}}(\{f_k\})}{k_BT}}\end{aligned}\tag{6.17}$$

The dependence of the eigenvalues $E_{\vec{n}}(\{f_k\})$ on f_k is obtained by first order perturbation theory.[9]

$$\begin{aligned}\frac{dE_{\vec{n}}}{df_k}&=\frac{d}{df_k}\Big\langle\Psi_{\vec{n}}(\{f_k\})\Big|\hat{H}(\{f_k\})\Big|\Psi_{\vec{n}}(\{f_k\})\Big\rangle\\&=\Big\langle\frac{d\Psi_{\vec{n}}(\{f_k\})}{df_k}\Big|\hat{H}(\{f_k\})\Big|\Psi_{\vec{n}}(\{f_k\})\Big\rangle+\Big\langle\Psi_{\vec{n}}(\{f_k\})\Big|\frac{d\hat{H}(\{f_k\})}{df_k}\Big|\Psi_{\vec{n}}(\{f_k\})\Big\rangle+\Big\langle\Psi_{\vec{n}}(\{f_k\})\Big|\hat{H}(\{f_k\})\Big|\frac{d\Psi_{\vec{n}}(\{f_k\})}{df_k}\Big\rangle\\&=\underbrace{\left(\Big\langle\frac{d\Psi_{\vec{n}}(\{f_k\})}{df_k}\Big|\Psi_{\vec{n}}(\{f_k\})\Big\rangle+\Big\langle\Psi_{\vec{n}}(\{f_k\})\Big|\frac{d\Psi_{\vec{n}}(\{f_k\})}{df_k}\Big\rangle\right)}_{=\frac{d}{df_k}\langle\Psi_{\vec{n}}(\{f_k\})|\Psi_{\vec{n}}(\{f_k\})\rangle=0}E_{\vec{n}}(\{f_k\})+\Big\langle\Psi_{\vec{n}}(\{f_k\})\Big|\hat{X}\Big|\Psi_{\vec{n}}(\{f_k\})\Big\rangle\\&=\Big\langle\Psi_{\vec{n}}(\{f_k\})\Big|\hat{X}\Big|\Psi_{\vec{n}}(\{f_k\})\Big\rangle\end{aligned}\tag{6.18}$$

If we insert this result into the above equation Eq. 6.17 for the derivative of the free energy, we obtain

$$\begin{aligned}\frac{\partial F}{\partial f_k}&\overset{\text{Eq. 6.17}}{=}\sum_{\vec{n}}\Big\langle\Psi_{\vec{n}}(\{f_k\})\Big|\hat{X}_k\Big|\Psi_{\vec{n}}(\{f_k\})\Big\rangle\frac{1}{Z(T,\{f_k\})}\mathrm{e}^{-\frac{E_{\vec{n}}(\{f_k\})}{k_BT}}\\&=\sum_{\vec{n}}\Big\langle\Psi_{\vec{n}}(\{f_k\})\Big|\hat{X}_k\underbrace{\frac{1}{Z(T,\{f_k\})}\mathrm{e}^{-\frac{\hat{H}(\{f_k\})}{k_BT}}}_{\hat{\rho}(\{f_k\})}\Big|\Psi_{\vec{n}}(\{f_k\})\Big\rangle\\&=\mathrm{Tr}\left[\hat{X}_k\hat{\rho}(\{f_k\})\right]=\langle\hat{X}_k\rangle\qquad q.e.d.\end{aligned}$$

---

in the entropy for the maximum-entropy principle. The internal energy is the expectation value of $\hat{H}_0$ which is obtained from the free energy

$$U_0(T,\{f_k\})=F(T,\{f_k\})+TS(T,\{f_k\})-\sum_k f_kX_k(T,\{f_k\})\tag{6.15}$$

Thus there is a difference in the definition of the internal energy, namely $U_f(T,\{f_k\})=U_0(T,\{f_k\})-\sum_j f_jX_j(T,\{f_k\})$. Both formulations have the same density operator, and both are internally consistent. However, it is important not to mix the expressions from one formulation with those of the other.

[9] Attention is required for degenerate states. In this case the states in the multiplet need to be chosen such, that they refer to different eigenstates of $\hat{A}$.

This is the correct quantum mechanical version of the equations of state.

Let us illustrate this finding on a few examples:

- consider a charged system for which the free energy as function of temperature and external potential is known. The coupling term is $\Delta H = -\int d^3r\ \Phi(\vec{r})\rho(\vec{r})$, where $\Phi(\vec{r})$ is the electrostatic potential and $\rho(\vec{r})$ is the charge density. Hence the mean value of the density at point $\vec{r}$ is

$$\langle\rho(\vec{r})\rangle = \frac{\delta F(T,[\Phi(\vec{r})])}{\delta\Phi(\vec{r})}$$

- similarly we obtain the magnetization as the derivative of the free energy with respect to a magnetic field. The coupling term is $\Delta H = -\int d^3r\ \vec{m}(\vec{r})\vec{B}(\vec{r})$. Hence

$$\langle m_i(\vec{r})\rangle = \frac{\delta F(T,[\vec{B}(\vec{r})])}{\delta B_i(\vec{r})}$$

The derivation of $S = -\frac{\partial F}{\partial T}$ is analogous to that of Eq. 1.65, when we use a representation of Eigenstates of the full Hamiltonian.

6.4 Fluctuations and correlations

We would like to know the uncertainty of the observable as well. The mean square deviation is[10]

$$\langle (A - \langle A\rangle)^2\rangle = \langle A^2\rangle - \langle A\rangle^2$$

We have to consider two distinct sources for the uncertainty of an observable.

- in a statistical mixture, each state has a different expectation value of a variable $\hat{A}$. The spread is then given by the probabilities to find a state with a given value.

- even in a pure state, there may be a quantum mechanical uncertainty, if the state is not an eigenstate of the operator $\hat{A}$.

Consider two different observables $\hat{A}$ and $\hat{B}$. We would like to know if the two quantities are correlated. Such a correlation would allow us to measure A and draw conclusions on the property B. The **correlation** is defined as

$$C_{AB} = \frac{1}{2}\Big(\langle (A - \langle A\rangle)(B - \langle B\rangle)\rangle + \langle (B - \langle B\rangle)(A - \langle A\rangle)\rangle\Big) = \langle \frac{AB + BA}{2}\rangle - \langle A\rangle\langle B\rangle$$

The above expression simplifies to the classical expression $\langle AB\rangle - \langle A\rangle\langle B\rangle$, if A and B commute.

Here the more complex quantum mechanical form has to be used if A and B are not compatible, that is if they do not commute with each other.

Let us first work out the fluctuations for operators that commute with the Hamiltonian. (Remember this important restriction!!) If the operators commute with the Hamiltonian, we can introduce a

---

[10]

$$\Big\langle (\hat{A} - \langle\hat{A}\rangle)^2\Big\rangle = \langle (\hat{A}^2 - 2\hat{A}\langle\hat{A}\rangle + \langle\hat{A}\rangle^2)\rangle = \langle\hat{A}^2\rangle - 2\langle\hat{A}\rangle\langle\hat{A}\rangle + \langle\hat{A}\rangle^2 = \langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2$$

basis of common eigenstates of the operators and the Hamiltonian.

$$
\begin{aligned}
\frac{\partial\langle X_j\rangle}{\partial f_i} &= \frac{\partial}{\partial f_i}\sum_{\vec{n}} X_{j,\vec{n}} \mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}(\{f_k\})-F(T,f))} \\
&= \sum_{\vec{n}} X_{j,\vec{n}} \left[-\frac{1}{k_BT}\left(\underbrace{X_{i,\vec{n}}}_{\frac{\partial E_{\vec{n}}}{\partial f_i}} - \frac{\partial F}{\partial f_i}\right)\right] \mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}(\{f_k\})-F(T,f))} \\
&= -\frac{1}{k_BT}\Big[\underbrace{\sum_{\vec{n}} X_{j,\vec{n}} X_{i,\vec{n}} \mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}(\{f_k\})-F(T,f))}}_{\langle X_iX_j\rangle} + \underbrace{\frac{\partial F}{\partial f_i}}_{\langle X_i\rangle} \underbrace{\sum_{\vec{n}} X_{j,\vec{n}} \mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}(\{f_k\})-F(T,f))}}_{\langle X_j\rangle}\Big] \\
&= -\frac{1}{k_BT}\Big[\langle X_iX_j\rangle - \langle X_i\rangle\langle X_j\rangle\Big] \\
&= -\frac{1}{k_BT}\Big\langle (X_i - \langle X_i\rangle)(X_j - \langle X_j\rangle)\Big\rangle
\end{aligned}
$$

This leads us to the linear response theorem.

LINEAR RESPONSE THEOREM

The **linear response theorem** relates the response of an extensive variable to the fluctuations.

$$
\frac{\partial\langle X_i\rangle}{\partial f_j} = -\frac{1}{k_BT}\Big\langle (X_i - \langle X_i\rangle)(X_j - \langle X_j\rangle)\Big\rangle
$$

Thus, if the equilibrium fluctuations of a quantity are large, its response to an external force will be strong as well. If fluctuations of two quantities are strongly correlated, the response of one quantity to a force acting on the other will be strong.

Note that, in this derivation, X_i can also be the energy. so that

$$
\frac{d\langle E\rangle}{df_j} = \frac{dU}{df_j} = -\frac{1}{k_BT}\Big\langle (E - \langle E\rangle)(X_j - \langle X_j\rangle)\Big\rangle
$$

The physical background of the linear response theorem is the following: On the one hand, fluctuations are large, if they do not require large energies. On the other hand, if a variation does not require a lot of energy, the expectation value of that variable can also be easily shifted by an external "force" such as f_k.

Note, that the second derivatives of the free energy are related to the response functions via

$$
\frac{\partial^2 F}{\partial f_i \partial f_j} = \frac{d\langle X_i\rangle}{df_j} = -\frac{1}{k_BT}\Big\langle (X_i - \langle X_j\rangle)(X_j - \langle X_j\rangle)\Big\rangle
$$

and

$$
\frac{dU}{df_i} = \langle X_i\rangle - T\frac{\partial^2 F}{\partial f_i \partial T}
$$

[11]. Thus we can calculate the fluctuations directly from the thermodynamic potentials.

---

[11] The equations of state specify $\langle X_i(T,f_k)\rangle = \frac{\partial F(T,f_k)}{\partial f_i}$, so that $\frac{\langle X_i\rangle}{\partial f_j} = \frac{\partial^2 F}{\partial f_i \partial f_j}$. Similarly $\langle E\rangle = U(T,f_k) = F(T,f_k) - T\frac{\partial F(T,f_k)}{\partial T}$, so that $\frac{dU}{df_i} = \frac{\partial F}{\partial f_i} - T\frac{\partial^2 F}{\partial f_i \partial T} = \langle X_i\rangle - T\frac{\partial^2 F}{\partial f_i \partial T}$.

Let us now form the second derivative involving the temperature

$$
\begin{aligned}
U_f &= \langle \hat{H}(\{f_k\})\rangle = \sum_{\vec{n}} E_{\vec{n}}(\{f_k\}) \mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}(\{f_k\})-F(T,f))} \\
\frac{dU_f}{dT} &= \frac{d\langle \hat{H}(\{f_k\})\rangle}{dT} = \sum_{\vec{n}} E_{\vec{n}}(\{f_k\}) \left(\frac{1}{k_BT^2}(E_{\vec{n}}(\{f_k\})-F) + \frac{1}{k_BT}\frac{\partial F}{\partial T}\right) \mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}(\{f_k\})-F(T,f))} \\
&= \frac{1}{k_BT^2}\Big[\underbrace{\sum_{\vec{n}} E^2_{\vec{n}}(\{f_k\}) \mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}(\{f_k\})-F(T,f))}}_{\langle E^2(\{f_k\})\rangle} \\
&\qquad - \underbrace{\left(F - T\frac{\partial F}{\partial T}\right)}_{=F+TS=U_f=\langle E(\{f_k\})\rangle} \underbrace{\sum_{\vec{n}} E_{\vec{n}} \mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}(\{f_k\})-F(T,f))}\Big]}_{\langle E(\{f_k\})\rangle} \\
&= \frac{1}{k_BT^2}\left[\langle E^2(\{f_k\})\rangle - \langle E\rangle^2\right] = \frac{1}{k_BT^2}\Big\langle (E-\langle E\rangle)^2 \Big\rangle
\end{aligned}
\tag{6.19}
$$

Since $c_V \overset{\text{def}}{=} dU/dT|_V$ is the specific heat, we obtain

ENERGY FLUCTUATIONS AND SPECIFIC HEAT

$$
\Big\langle (E(\{f_k\}) - \langle E(\{f_k\})\rangle)^2 \Big\rangle = k_BT^2 c_V \tag{6.20}
$$

Here we used the specific heat at constant volume, which is one of the possible special cases. Note here that if we are interested in the fluctuation E_0 from this expression, we would need to include the fluctuations of all the other variables as well.

We see that the fluctuations grow with temperature, and that they are larger with a larger specific heat. By making a dimensional argument[12] we can convince ourselves that the relative mean square deviation decreases with system size.

$$
\frac{\sqrt{\langle (E-\langle E\rangle)^2\rangle}}{\langle E\rangle} \overset{\text{Eq. 6.19}}{=} \frac{k_BT}{U}\sqrt{\frac{1}{k_B}\frac{dU}{dT}} \sim \frac{1}{\sqrt{N}}
$$

Let us consider a Boltzmann gas. The energy is $U = \langle E\rangle = \frac{3}{2}Nk_BT$ so that we obtain

$$
\frac{\sqrt{\langle (E-\langle E\rangle)^2\rangle}}{\langle E\rangle} = \frac{k_BT}{U^{BG}}\sqrt{\frac{1}{k_B}\frac{dU^{BG}}{dT}} = \sqrt{\frac{2}{3N}}
$$

For $N = 10^{23}$ particles we find that the energy fluctuates by less than a factor 10^{-10}. This is an indication that the difference between the canonical and the micro-canonical ensemble disappear for large systems. For a single molecule, however, the relative fluctuation is almost one!

Let us now consider the variation in the number of particles. We set $X = N$ and $f = \mu$ and obtain

$$
\langle (N-\langle N\rangle)^2\rangle = k_BT\frac{dN}{d\mu}
$$

A similar dimensional argument as the one given for the energy leads to

$$
\frac{\sqrt{\langle (N-\langle N\rangle)^2\rangle}}{\langle N\rangle} \sim \frac{1}{\sqrt{N}}
$$

---

[12] $E \sim N$, $U \sim N$, $T \sim 1$

6.4.1 Examples for the linear response theorem

6.5 Thermodynamic limit

Macroscopic systems are considered identical copies of many smaller systems. These copies are in contact with each other and can exchange any conserved quantity. Thus, the corresponding Lagrange multipliers are identical for all the subsystems. Hence, the composed system is characterized in thermal equilibrium by a set of **extensive variables** and **intensive variables**.

- The extensive variables are the values of the conserved quantities $E_{tot} = \sum_j E_j$ and $X_{k,tot} = \sum_j X_{k,j}$ summed over all subsystems, where j is the sum over the subsystems and the index i specifies the type of the conserved quantity. The conserved quantity may be energy, volume etc. The value of an extensive variable scales with the size of the system.

- The intensive variables are the Lagrange multipliers T, f_k Since the subsystems are in contact, the values of the Lagrange multipliers of all the subsystems are identical in thermal equilibrium, i.e. $T = T_j$ and $f_{k,tot} = f_{k,j}$. This is also the value we attribute to the combined system.

Sofar we have identified U, X_k with a statistical mean value of E and X_k respectively. Thus in any particular measurement we may measure a value of E, X_k that differs from the mean value. In the thermodynamic limit, however, the mean square deviation from the average value vanishes as we have seen in the previous section on fluctuations. Thus we obtain with certainty the same value in each measurement. This is the **thermodynamic limit** In the thermodynamic limit, we need not distinguish between average and instantaneous values.

The underlying reason is the central limit theorem, which is proven below. It says that the mean value $\frac{1}{N}\sum_{i=1}^{N} A_i$ of the result obtained from N measurements, where $N \to \infty$, in a random experiment has a Gaussian distribution with variance σ^2/N, where $\sigma^2 = \langle (A - \langle A \rangle)^2 \rangle$ is the variance of A a given experiment. We identify each individual measurement yielding A_i with a subsystem of a large system in thermal contact. The central limit theorem says that the relative standard deviation $\frac{\sigma}{\langle A \rangle}$ vanishes proportional to $\frac{1}{\sqrt{N}}$. Multiplication of mean value and standard deviation with N leaves this ratio untouched, so that we find that while the spread of values grows with the size of the system, the relative error vanishes.

6.5.1 Central limit theorem

The **central limit theorem** is one of the hallmarks of statistical theory, and it explains, why we obtain Gaussian distributed results in many experiments. It also explains, why we can reduce the uncertainty of an experiment by performing many experiments and averaging the result.

The central limit theorem says that if we perform N times an identical experiment with a probability $p(x)$, then the probability for the average value of x is normal distributed with the mean value equal to that of the individual experiment and and a standard deviation that vanishes as $1/\sqrt{N}$. The requirement is that each experiment is independent of each other and that N is sufficiently large.

The **variance** is defined as $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$ The variable $\sigma = \sqrt{\sigma^2}$ is called the **standard deviation**.

Let us prove the central limit theorem. Let us say we perform N independent experiments, and obtain for the i-th experiment the result x_i. The probability distributions $p_i(x)$ are identical, that is $p_i(x) = p_j(x)$ so that we use only a single distribution $p(x) = p_i(x)$. Furthermore, they are statistical uncorrelated. Hence the probability $P(x_1, x_2, \dots)$ to obtain x_1 in the first experiment, x_2 in the second, etc., is given by the products of the distributions of the individual experiment, that is

$$P(x_1, x_2, \dots) = p(x_1)p(x_2)\dots \tag{6.21}$$

We define the mean value X of the values x_i of the subsystems

$$X=\frac{1}{N}\sum_{i=1}^{N}x_i$$

Now we are ready to start: Let $P(X)$ be the probability that X has a certain value. We first construct the Fourier transform

$$P(k)=\int dX\mathrm{e}^{-ikX}P(X)$$

We can write the Fourier transform $P(k)$ as the expectation value of e^{-ikX}.

$$\begin{aligned}
P(k) &= \langle \mathrm{e}^{-ikX}\rangle=\langle \mathrm{e}^{-i\frac{k}{N}\sum_{i=1}^{N}x_i}\rangle=\langle\prod_{i=1}^{N}\mathrm{e}^{-i\frac{kx_i}{N}}\rangle\\
&\overset{\text{Eq. 6.21}}{=} \int dx_1\ldots\int dx_N\ p(x_1)\ldots p(x_N)\mathrm{e}^{-ikx_1/N}\ldots\mathrm{e}^{-ikx_N/N}\\
&= \Big[\int dxp(x)\mathrm{e}^{-ikx/N}\Big]^N=\exp\Big[N\ln[\int dxp(x)\mathrm{e}^{-ikx/N}]\Big]\\
&\overset{\text{Eq. 6.23}}{=} \exp\Big[N\ln[\sum_{j=0}^{\infty}\frac{1}{j!}(\frac{-ik}{N})^j\int dxp(x)x^j]\Big]=\exp\Big[N\underbrace{\ln[\sum_{j=0}^{\infty}\frac{1}{j!}(\frac{-ik}{N})^j\langle x^j\rangle]}_{=:A}\Big] \qquad (6.22)
\end{aligned}$$

We will come back to the expression indicated by A below. We have used the power series expansion

$$\mathrm{e}^x=\sum_{j=0}^{\infty}\frac{1}{j!}x^j \qquad (6.23)$$

of the exponential function. The expectation values of powers of a variable are called the **moments** of the probability distribution. For example, the mean value $\langle x\rangle$ is called the first moment, $\langle x^2\rangle$ is called the second moment, etc.

Now we use the power series expansion of the logarithm

$$\ln[1+x]=x-\frac{1}{2}x^2+O(x^3)$$

to obtain the highest orders in $\frac{1}{N}$, which dominate for $N\to\infty$.[13]

$$\begin{aligned}
A &\overset{\text{see Eq. 6.22}}{=} \ln[\sum_{j=0}^{\infty}\frac{1}{j!}(\frac{-ik}{N})^j\langle x^j\rangle]\\
&= \ln[1-\frac{ik\langle x\rangle}{N}-\frac{k^2\langle x^2\rangle}{2N^2}+O(\frac{1}{N^3})]\\
&= \Big(-\frac{ik\langle x\rangle}{N}-\frac{k^2\langle x^2\rangle}{2N^2}\Big)-\frac{1}{2}\Big(-\frac{ik\langle x\rangle}{N}\Big)^2+O(\frac{1}{N^3})\\
&= -\frac{ik}{N}\langle x\rangle-\frac{1}{2}\frac{k^2}{N^2}\Big(\langle x^2\rangle-\langle x\rangle^2\Big)+O(\frac{1}{N^3})\\
&= \frac{1}{N}\Big(-ik\langle x\rangle-\frac{1}{2}k^2\frac{\sigma^2}{N}\Big)+O(\frac{1}{N^3})
\end{aligned}$$

where $\sigma^2=\langle x^2\rangle-(\langle x\rangle)^2$ is the variance of x_i in each subsystem. Inserting the result for A into Eq. 6.22 yields

$$P(k)\overset{\text{Eq. 6.22}}{=}\mathrm{e}^{NA}=\mathrm{e}^{-ik\langle x\rangle-\frac{1}{2}k^2\frac{\sigma^2}{N}}+O(\frac{1}{N^2})$$

---

[13] This statement is not mathematically precise. Some further assumptions on the probability distribution are required.

$P(X)$ is obtained by the Fourier back-transform

$$P(X) = \int \frac{dk}{2\pi} \mathrm{e}^{ikX} P(k) = \int \frac{dk}{2\pi} \mathrm{e}^{i(X-\langle x\rangle)k - \frac{1}{2}\frac{\sigma^2}{N}k^2 + O(\frac{1}{N^2})}$$

Now we use the relation

$$\int_{-\infty}^{\infty} dy\; \mathrm{e}^{iay-by^2} = \sqrt{\frac{\pi}{b}} \mathrm{e}^{-\frac{a^2}{4b}}$$

which is the Fourier transform of a Gauss function. We replace $y = k$, $a = X - \langle x\rangle$ and $b = \sigma^2/(2N)$ in the above equation, which yields

$$\begin{aligned} P(X) &= \frac{1}{2\pi}\sqrt{\frac{\pi}{\sigma^2/(2N)}} \mathrm{e}^{-\frac{(X-\langle x\rangle)^2}{4\sigma^2/(2N)}} + O(\frac{1}{N^2}) \\ &= \sqrt{\frac{1}{2\pi\sigma^2/N}} \mathrm{e}^{-\frac{(X-\langle x\rangle)^2}{2\sigma^2/N}} + O(\frac{1}{N^2}) \end{aligned}$$

Thus we obtain for large N a normal distributed probability, irrespective of the probability $p(x)$ of the subsystems. q.e.d.

CENTRAL LIMIT THEOREM

$$P(X) = \sqrt{\frac{1}{2\pi\sigma^2/N}} \exp\left(-\frac{(X-\langle x\rangle)^2}{2\sigma^2/N}\right) + O(\frac{1}{N^2}) \tag{6.24}$$

Normal distributed means that the probability distribution has the shape of a Gaussian. The mean value of the distribution of X is the mean value of the individual distribution, and the standard deviation of the distribution is given by $\sigma_X = \sigma_x/\sqrt{N}$. In other words the percentage deviation vanishes like $1/\sqrt{N}$. σ_X is the standard deviation of the variable X and σ_x is the standard deviation of the variables x_i.

6.6 The classical Limit

Many systems can be described adequately without referring to quantum mechanics. Often a problem, that is intractable on a quantum-mechanical level, is still accessible with classical physics. Therefore we investigate here the classical limit $\hbar \to 0$.

In order to make the connection to thermodynamics, we need to determine an expression, Eq. 6.43, of the partition sum in terms of classical variables, the coordinates and momenta. Therefore we define position and momentum eigenstates $|x\rangle$ and $|p\rangle$. In order to simplify the discussion we start out with only one dimension. The result can easily be generalized to arbitrary many dimensions. The states are defined via their matrix elements. The amplitude of the momentum eigenstate $|p\rangle$ at position x is a plane wave is defined as the matrix element $\langle x|p\rangle$

6.6.1 Basics

When we introduced quantum mechanics, we started from classical mechanics, and deduced a quantum theory by introducing the wave nature of "particles" and by treating them as wave packets. Later we arrived at the more abstract bracket notation, in which wave functions appeared only in one special representation. Here we will go all the way back. For this reason I start here from the bracket notation and I will introduce position and momenta. besides greatly simplifying the proof that will follow later, it shall also remind the reader of the fundamentals of quantum mechanics.

Before we start with the derivation of the classical limit let us define the basisset for position and momentum. We do this consistently within the bracket notation, because that will also train the latter.

In the following we need the position and momentum eigenstates defined by

$$\hat{x}|x\rangle = |x\rangle x \tag{6.25}$$

$$\langle x|x'\rangle = \delta(x-x') \tag{6.26}$$

$$\hat{1} = \int dx\; |x\rangle\langle x| \tag{6.27}$$

and

$$\hat{p}|p\rangle = |p\rangle p \tag{6.28}$$

$$\langle p|p'\rangle = 2\pi\hbar\delta(p-p') \tag{6.29}$$

$$\hat{1} = \int \frac{dp}{2\pi\hbar}\; |p\rangle\langle p| \tag{6.30}$$

The relation between position and momentum representation is given by

$$\langle x|p\rangle = \mathrm{e}^{\frac{i}{\hbar}px} \tag{6.31}$$

The position and momentum operators can the be written as

$$\hat{x} = \int dx\; |x\rangle x\langle x| \tag{6.32}$$

$$\hat{p} = \int dx\; |x\rangle \frac{\hbar}{i}\partial_x\langle x| \tag{6.33}$$

$$\hat{H} \stackrel{\text{def}}{=} H(\hat{p},\hat{x}) = \int dx\; |x\rangle H(\frac{\hbar}{i}\partial_x, x)\langle x| \tag{6.34}$$

If any of these formulas are not known, please review Appendix B on p. 221 or the corresponding chapters in ΦSX: Quantum Theory.

Derivative of $\mathrm{e}^{-\beta\hat{H}}$

Later we need the derivative of $\mathrm{e}^{-\beta\hat{H}}$ with respect to $\beta = \frac{1}{k_BT}$.

$$\partial_\beta \mathrm{e}^{-\beta\hat{H}} = -\hat{H}\mathrm{e}^{-\beta\hat{H}} \tag{6.35}$$

The detailed derivation is as follows: We use the power series expansion of the exponential function

$$\mathrm{e}^x = \sum_{j=0}^{\infty}\frac{1}{j!}x^j$$

and obtain

$$\begin{aligned}
\partial_\beta \mathrm{e}^{-\beta\hat{H}} &= \partial_\beta \sum_{j=0}^{\infty}\frac{1}{j!}\left(-\beta\hat{H}\right)^j = \sum_{j=1}^{\infty}\frac{1}{j!}\left(j\beta^{j-1}\right)\left(-\hat{H}\right)^j \\
&= -\hat{H}\sum_{j=1}^{\infty}\frac{1}{(j-1)!}\left(-\beta\hat{H}\right)^{j-1} = -\hat{H}\sum_{j=0}^{\infty}\frac{1}{j!}\left(-\beta\hat{H}\right)^j \\
&= -\hat{H}\mathrm{e}^{-\beta\hat{H}}
\end{aligned}$$

Some more preparation

The second relation needed is[14]

$$\langle p|x\rangle H(\frac{\hbar}{i}\partial_x,x)=H(p-i\hbar\partial_x,x)\langle p|x\rangle \tag{6.36}$$

We first derive a simpler case, of the equation above before we generalize it to obtain Eq. 6.36.

$$\begin{aligned}(p-i\hbar\partial_x)\langle p|x\rangle &\overset{\text{Eq. 6.31}}{=} (p-i\hbar\partial_x)\mathrm{e}^{-ipx/\hbar}\\ &= p\mathrm{e}^{-ipx/\hbar}-\underbrace{i\hbar\left(\partial_x\mathrm{e}^{-ipx/\hbar}\right)}_{=p\mathrm{e}^{-ipx/\hbar}}-\mathrm{e}^{-ipx/\hbar}i\hbar\partial_x\\ &= -\mathrm{e}^{-ipx/\hbar}i\hbar\partial_x=\langle p|x\rangle\frac{\hbar}{i}\partial_x\end{aligned}$$

Thus we obtain

$$\langle p|x\rangle\frac{\hbar}{i}\partial_x=(p-i\hbar\partial_x)\langle p|x\rangle \tag{6.37}$$

This result can be generalized to arbitrary functions of $\frac{\hbar}{i}\partial_x$ and x. A function of a differential operator is defined, as any function of an operator, by its power series expansion. If a function $H(p,x)$ has a power series expansion, then the function of an operator is defined by replacing the argument with the corresponding differential operator.

$$\begin{aligned}H(p,x) &\overset{\text{def}}{=} \sum_j a_j(x)p^j \qquad\Leftrightarrow\qquad H(\frac{\hbar}{i}\partial_x,x)=\sum_j a_j(x)\left(\frac{\hbar}{i}\partial_x\right)^j\\ \langle p|x\rangle H(\frac{\hbar}{i}\partial_x,x) &= \sum_j a_j(x)\langle p|x\rangle(\frac{\hbar}{i}\partial_x)^j \overset{\text{Eq. 6.37}}{=} \sum_j a_j(x)(p-i\hbar\partial_x)\langle p|x\rangle(\frac{\hbar}{i}\partial_x)^{j-1}=\cdots\\ &\overset{\text{Eq. 6.37}}{=} \sum_j a_j(x)(p-i\hbar\partial_x)^j\langle p|x\rangle=H(p-i\hbar\partial_x,x)\langle p|x\rangle\end{aligned}$$

Thus we have obtained Eq. 6.36

6.6.2 Derivation of the partition sum in the classical limit

Now we are in the position to work out the partition function

$$Z(T)\overset{\text{Eq. 6.9}}{=}\mathrm{Tr}[\mathrm{e}^{-\beta\hat{H}}]\overset{Eq.\ 6.30}{=}\int\frac{dp}{2\pi\hbar}\langle p|\mathrm{e}^{-\beta\hat{H}}|p\rangle\overset{Eq.\ 6.27}{=}\int\frac{dp}{2\pi\hbar}\int dx\,\underbrace{\langle p|x\rangle\langle x|\mathrm{e}^{-\beta\hat{H}}|p\rangle}_{I(p,x,\beta)} \tag{6.38}$$

The dependence of the partition function on the other intrinsic variables f_k in Eq. 6.9 has been absorbed, for the sake of simplicity, into the definition of the Hamiltonian $\hat{H}(\{f_k\})=\hat{H}_0+\sum_k f_k\hat{X}_k$.

Let us take a closer look at the integrand

$$I(p,x,\beta)\overset{\text{def}}{=}\langle p|x\rangle\langle x|\mathrm{e}^{-\beta\hat{H}}|p\rangle \tag{6.39}$$

---

[14] To make the distinction between Hamilton function and Hamilton operator specific, note the following: $H(p,x)$ is the Hamilton function. If the arguments of the Hamilton function are operators, $\hat{H}\overset{\text{def}}{=}H(\hat{p},\hat{x})$ is the Hamilton operator.

and form the derivative with β:

$$
\begin{aligned}
\frac{\partial I}{\partial \beta} &\overset{Eq.\ 6.35}{=} -\langle p|x\rangle\langle x|\hat{H}\mathrm{e}^{-\beta\hat{H}}|p\rangle \\
&\overset{Eq.\ 6.34}{=} -\int dx'\langle p|x\rangle \underbrace{\langle x|x'\rangle}_{\delta(x-x')} H(\frac{\hbar}{i}\partial_{x'},x')\langle x'|\mathrm{e}^{-\beta\hat{H}}|p\rangle \\
&= -\langle p|x\rangle H(\frac{\hbar}{i}\partial_{x},x)\langle x|\mathrm{e}^{-\beta\hat{H}}|p\rangle \\
&\overset{Eq.\ 6.36}{=} -H(p-i\hbar\partial_x,x)\langle p|x\rangle\langle x|\mathrm{e}^{-\beta\hat{H}}|p\rangle \\
&= -H(p-i\hbar\partial_x,x)I(x,p,\beta) \\
&\overset{Taylor}{=} -H(p,x)I(p,x,\beta)-\frac{\partial H(p,x)}{\partial p}\underbrace{\frac{\hbar}{i}}_{-i\hbar}\partial_x I(p,x,\beta)+O(\hbar^2)
\end{aligned}
$$

Thus we obtained a partial differential equation for $I(x,p,\beta)$ in x and β.

$$
\frac{\partial I(p,x,\beta)}{\partial\beta}+\frac{\hbar}{i}\frac{\partial H(p,x)}{\partial p}\frac{\partial I(p,x,\beta)}{\partial x}=-H(p,x)I(p,x,\beta)+O(\hbar^2)
$$

Now we perform the **classical limit** by retaining only the leading order in $\hbar$. For each value for x and p, we arrive at a differential equation in β,

$$
\frac{\partial I}{\partial\beta}\overset{\hbar\to 0}{=}-H(p,x)I(p,x,\beta)
$$

Note that this equation does no more contain the differential operator in x. Thus, for each value of the classical variables x and p, we can solve the differential equation and arrive at

$$
I(p,x,\beta)=C(p,x)\mathrm{e}^{-\beta H(p,x)} \tag{6.40}
$$

The integration constant $C(p,x)$ is obtained from the high-temperature limit, which we transform using the the eigenstates $|n\rangle$ of the Hamiltonian with energies E_n.

$$
\begin{aligned}
C(p,x) &= \lim_{\beta\to 0} I(p,x,\beta) \overset{Eq.\ 6.39}{=} \lim_{\beta\to 0}\sum_n \langle p|x\rangle\langle x|\mathrm{e}^{-\beta\hat{H}}|n\rangle\langle n|p\rangle \\
&= \lim_{\beta\to 0}\sum_n \langle p|x\rangle\langle x|n\rangle\mathrm{e}^{-\beta E_n}\langle n|p\rangle \\
&\overset{\lim_{\beta\to 0}\mathrm{e}^{-\beta E_n}=0}{=} \sum_n \langle p|x\rangle\langle x|n\rangle\langle n|p\rangle \\
&= \langle p|x\rangle\langle x|p\rangle \overset{Eq.\ 6.31}{=} \mathrm{e}^{-ipx/\hbar}\mathrm{e}^{+ipx/\hbar}=1
\end{aligned} \tag{6.41}
$$

Thus we obtain

$$
I(p,x,\beta)\overset{Eqs.\ 6.40,6.41}{=}\mathrm{e}^{-\beta H(p,x)} \tag{6.42}
$$

Note that I is a function of real numbers, and not an operator function.

Thus we obtain our final result, the **partition function in the classical limit**,

$$
Z(\beta)\overset{Eqs.\ 6.38,6.42}{=}\int dx\int\frac{dp}{2\pi\hbar}\,\mathrm{e}^{-\beta H(p,x)}
$$

The prefactor ensures that the partition function dimension-less. Like $\hbar$ the product $dxdp$ has the units of an angular momentum.

PARTITION FUNCTION IN THE CLASSICAL LIMIT

The partition function of N classical particles, whose dynamics is governed by the Hamilton function $H(\vec{p}_1, \dots, \vec{p}_N, \vec{r}_1, \dots, \vec{r}_N)$, is

$$Z(\beta) = \int d^3r_1 \cdots d^3r_N \int \frac{d^3p_1}{2\pi\hbar} \cdots \frac{d^3p_N}{2\pi\hbar} \, \mathrm{e}^{-\beta H(\vec{p}_1,\dots,\vec{p}_N,\vec{r}_1,\dots,\vec{r}_N)} \tag{6.43}$$

For identical particles we need to divide by the number of permutations, so that

$$Z(\beta) = \frac{1}{N!} \int d^3r_1 \cdots d^3r_N \int \frac{d^3p_1}{2\pi\hbar} \cdots \frac{d^3p_N}{2\pi\hbar} \, \mathrm{e}^{-\beta H(\vec{p}_1,\dots,\vec{p}_N,\vec{r}_1,\dots,\vec{r}_N)} \tag{6.44}$$

Note that, even in the classical limit the partition function depends on $\hbar$. This factor does not affect the physics, because in the free energy $F = -k_BT \ln[Z]$ it contributes a constant and only the derivatives of the free energy have physical meaning.

6.6.3 Expectation values

Now we would like to obtain mean values an observable $\hat{A} \stackrel{\text{def}}{=} A(\hat{p}, \hat{x})$.

We use the **extended Hamiltonian**

$$\hat{H}(\lambda) \stackrel{\text{def}}{=} \hat{H}_0 + \lambda\hat{A} = H_0(\hat{p}, \hat{x}) + \lambda A(\hat{p}, \hat{x})$$

The derivative of the free energy with respect to λ yields the expectation value of the corresponding observable $\hat{A}$ via Eq. 6.13. This statement is analogous to an equation of state Eq. 1.66 on p. 39.

Now we apply this principle to the expressions in the classical limit Eq. 6.43. We do not differentiate between identical and non-identical particles, because the factor $1/N!$ contributes only a constant to the free energy, which does not appear in the derivative.

$$\begin{aligned}
\langle A \rangle_T &\stackrel{\text{Eq. 6.13}}{=} \frac{\partial F(T,\lambda)}{\partial \lambda} \\
&= \frac{\partial}{\partial \lambda}\Big|_{\lambda=0} \left(-k_BT \ln[Z(\beta,\lambda)]\right) \\
&\stackrel{\text{Eq. 6.43}}{=} \frac{\partial}{\partial \lambda}\Big|_{\lambda=0} \left(-k_BT \ln\Big[\int dx \int \frac{dp}{2\pi\hbar} \mathrm{e}^{-\beta(H_0(p,x)+\lambda A(p,x))}\Big]\right) \\
&= \int dx \int \frac{dp}{2\pi\hbar} A(p,x) \frac{1}{Z(T,0)} \mathrm{e}^{-\beta H_0(p,x)} \\
&= \int dx \int dp \; A(p,x) P(p,x,T)
\end{aligned}$$

The probability density $P(x_0, p_0)$ can be obtained as the expectation value of an observable $A(p,x) = \delta(p - p_0)\delta(x - x_0)$. Thus we can define a probability density in phase space

$$P(p,x,T) = \frac{1}{2\pi\hbar Z(T,0)} \mathrm{e}^{-\beta H_0(p,x)} = \frac{\mathrm{e}^{-\beta H_0(p,x)}}{\int dx \int dp \; \mathrm{e}^{-\beta H_0(p,x)}} \tag{6.45}$$

PROBABILITY DENSITY IN PHASE SPACE

We can generalize this result Eq. 6.45 to N-particle systems

$$\langle A\rangle=\int d^{3N}p\int d^{3N}r\;P(\vec p_1,\dots\vec p_N,\vec r_1,\dots,\vec r_N)\;A(\vec p_1,\dots\vec p_N,\vec r_1,\dots,\vec r_N) \tag{6.46}$$

with

$$P(\vec p_1,\dots\vec p_N,\vec r_1,\dots,\vec r_N)=\frac{\mathrm{e}^{-\beta H(\vec p_1,\dots\vec p_N,\vec r_1,\dots,\vec r_N)}}{\int d^{3N}p\int d^{3N}r\;\mathrm{e}^{-\beta H(\vec p_1,\dots\vec p_N,\vec r_1,\dots,\vec r_N)}} \tag{6.47}$$

The **ergodic theorem** says that the ensemble average of an observable is identical to the time average, that is

$$\int dx\int dp\;A(p,x)P(p,x,T)=\lim_{T\to\infty}\frac{1}{T}\int_0^T dt\;A(p(t),x(t))$$

where $x(t),p(t)$ is any trajectory of the system solving Hamilton's equation. This allows to determine average values from a molecular dynamics simulations.

6.6.4 Separating position and momentum variables

If the Hamilton function has the special form

$$H(p,x)=\frac{p^2}{2m}+V(x)$$

The probability density can be split into a density in position space and a velocity distribution

$$P(p,x,T)=\frac{\mathrm{e}^{-\beta\left(\frac{p^2}{2m}+V(x)\right)}}{\int dx\int dp\;\mathrm{e}^{-\beta\left(\frac{p^2}{2m}+V(x)\right)}}=\underbrace{\frac{\mathrm{e}^{-\beta\frac{p^2}{2m}}}{\int dp\;\mathrm{e}^{-\beta\frac{p^2}{2m}}}}_{P^{kin}(p,T)}\underbrace{\frac{\mathrm{e}^{-\beta V(x)}}{\int dx\;\mathrm{e}^{-\beta V(x)}}}_{P^{pot}(x,T)}$$

Thus, the momentum distribution is identical in each point in space and it is independent of the potential.

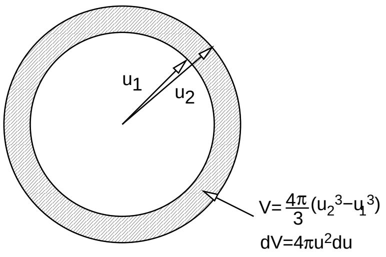
The momentum distribution can be translated into a velocity distribution using $p=mv$.

$$\begin{aligned}
P(p)dp&=P(v)dv\qquad\Rightarrow P(v)=P(p)\frac{dp}{dv}=mP(p)\\
P(v)&=\frac{m\mathrm{e}^{-\beta\frac{p^2}{2m}}}{\int dp\;\mathrm{e}^{-\beta\frac{p^2}{2m}}}=\frac{m\mathrm{e}^{-\frac{mv^2}{2k_BT}}}{m\int dv\;\mathrm{e}^{-\frac{mv^2}{2k_BT}}}=\frac{\mathrm{e}^{-\frac{mv^2}{2k_BT}}}{\int dv\;\mathrm{e}^{-\frac{mv^2}{2k_BT}}}=\frac{\mathrm{e}^{-\frac{mv^2}{2k_BT}}}{\sqrt{\frac{2k_BT}{m}}\int dy\;\mathrm{e}^{-y^2}}\\
&=\sqrt{\frac{m}{2\pi k_BT}}\mathrm{e}^{-\frac{mv^2}{2k_BT}}
\end{aligned}$$

In higher dimension the kinetic energy can be factorized into one-dimensional terms. If we consider the velocity distribution for one particle in three dimensions we obtain the **Maxwell-Boltzmann distribution of molecular velocities**

$$P(\vec v)=\left(\frac{m}{2\pi k_BT}\right)^{\frac{3}{2}}\mathrm{e}^{-\frac{m\vec v^2}{2k_BT}} \tag{6.48}$$

Note that the Maxwell-Boltzmann distribution of velocities is not only valid in an ideal gas, but for all non-interacting particles. It requires, however, that kinetic and potential energy can be separated.



By integrating over all directions, we obtain the **Maxwell distribution of speeds**.[15]

$$P(|\vec{v}|) = \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}} 4\pi\vec{v}^2 \mathrm{e}^{-\frac{m\vec{v}^2}{2k_BT}} \tag{6.49}$$

The probability in position space only depends on the potential without considering the velocity distribution

$$P(x) = \frac{\mathrm{e}^{-\beta V(x)}}{\int dx \mathrm{e}^{-\beta V(x)}}$$

Of particular interest is to express the partition function directly by the potential and to integrate out the momentum variables

$$\begin{aligned} Z(\beta) &= \left[\int dx\ \mathrm{e}^{-\beta V(x)}\right]\left[\int \frac{dp}{2\pi\hbar}\mathrm{e}^{-\beta\frac{p^2}{2m}}\right] \\ &= \int dx\ \mathrm{e}^{-\beta V(x)} \frac{\sqrt{2\pi m k_B T}}{2\pi\hbar} \\ &= \underbrace{\sqrt{\frac{mk_BT}{2\pi\hbar^2}}}_{\lambda_T^{-1}} \int dx\ \mathrm{e}^{-\beta V(x)} \end{aligned}$$

With the help of the **de-Broglie Wavelength** from Eq. 2.7, defined previously in the derivation of the Boltzmann gas, we can write the partition function as

$$Z(\beta) = \int \frac{dx}{\lambda_T}\ \mathrm{e}^{-\beta V(x)} \tag{6.50}$$

We can generalize the result to many-particle systems and obtain

---

[15] Consider a probability distribution $P^{\vec{v}}(\vec{v})$ of the three dimensional velocity space. We wish to construct the probability density $P^{(u)}(u)$ for the absolute value $u \overset{\text{def}}{=} |\vec{v}|$ of the velocity. Let us calculate the probability of finding the velocity between u_1 and u_2, which we evaluate in Polar coordinates. We denote the angular integration by $d\Omega \overset{\text{def}}{=} d\phi d\theta \sin(\theta)$.

$$\int_{u_1}^{u_2} du\ P^{(u)}(u) \overset{!}{=} \int_{u_1<|\vec{v}|<u_2} d^3v\ P^{\vec{v}}(\vec{v}) = \int_{u_1}^{u_2} du \int d\Omega\ u^2 P^{\vec{v}}$$

Now we exploit that the probability density $P^{\vec{v}}$ only depends on the absolute value, so that we can perform the angular integration.

$$\int_{u_1}^{u_2} du\ P^{(u)}(u) = \int_{u_1}^{u_2} du\ 4\pi u^2 P^{\vec{v}}(\vec{v})$$

Thus we can translate the probability densities by

$$P(|\vec{v}|) \overset{\text{def}}{=} P^{(u)}(u) = 4\pi u^2 P^{\vec{v}}(\vec{v}) =: 4\pi\vec{v}^2 P(\vec{v})$$

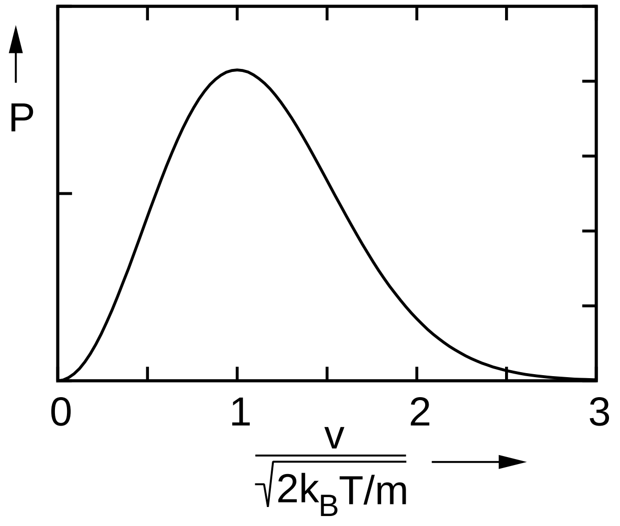


Fig. 6.2: Maxwell distribution of speeds.

PARTITION SUM IN THE CLASSICAL LIMIT EXPRESSED IN TERMS OF THE POTENTIAL

$$Z(\beta)=\left(\prod_{i=1}^{N}\frac{1}{\lambda^3_{T,i}}\right)\int d^{3N}r\ \mathrm{e}^{-\beta V(r_1,\dots,r_{3N})}$$

For identical particles we need to divide by the number of permutations

$$Z(\beta)=\frac{1}{N!}\left(\prod_{i=1}^{N}\frac{1}{\lambda^3_{T,i}}\right)\int d^{3N}r\ \mathrm{e}^{-\beta V(r_1,\dots,r_{3N})}$$

PROBABILITY DENSITY IN CONFIGURATION SPACE

The probability density of a system with a conventional kinetic energy can be written as particles as

$$P(\vec{r}_1,\dots,\vec{r}_N,T)=\frac{\mathrm{e}^{-\beta V(\vec{r}_1,\dots,\vec{r}_N)}}{\int d^3r_1\cdots\int d^3r_N\ \mathrm{e}^{-\beta V(\vec{r}_1,\dots,\vec{r}_N)}}$$

Note that this result is valid both for identical and non-identical particles, because the factor $N!$ is canceled by the normalization.

With the help of the probability density Eq. 6.51, one can determine expectation values of observables that do not depend on the velocities. Expectation values for some operator can now be obtained as

$$\begin{aligned}\langle A\rangle&=\mathrm{Tr}[\hat{A}\hat{\rho}]\\&=\int d^{3N}r\int d^{3N}p\ P(\vec{r}_1,\dots,\vec{r}_N,\vec{p}_1,\dots,\vec{p}_N,T)A(\vec{r}_1,\dots,\vec{r}_N,\vec{p}_1,\dots,\vec{p}_N)\end{aligned}$$

However, for many particles one has to solve integrals in very high dimensions. Direct evaluation of these integrals is often not possible due to the dimension bottle neck.[16]

---

[16] The effort to integrate a function in $3N$ dimensions grows exponentially with N. Thus the limitations for any

The expression for the probability density lends itself for a number of approximative techniques. One efficient way to evaluate highly dimensional integrals is the **Monte-Carlo method**, which amounts to averaging the values of the observable for randomly selecting configurations in such a way that their frequency is proportional to the Boltzmann factor. While the basic idea seems simple, the mathematical background and the methodology is fairly sophisticated.

Another common approximation is to exploit that the Boltzmann factor effectively vanishes when the potential is high. The so-called **saddle-point approximation** approximates the potential by harmonic oscillators centered at the local minima. This turns the integrals into Gauss integrals that can be solved analytically. This method is at its best at low temperatures.

6.6.5 Speed of molecules

It may be of interest to investigate the average velocity of molecules in air.

$$\begin{aligned}
\langle|v|\rangle &= \int_0^\infty dv\; P(v)v \\
&\overset{\text{Eq. 6.49}}{=} 4\pi\left(\frac{m}{2\pi k_BT}\right)^{\frac{3}{2}}\int_0^\infty dv\; v^3\mathrm{e}^{-\frac{mv^2}{2k_BT}} \\
&\overset{x=mv^2/(2k_BT)}{=} 4\pi\left(\frac{m}{2\pi k_BT}\right)^{\frac{3}{2}}\left(\frac{2k_BT}{m}\right)^2\underbrace{\int_0^\infty dx\; x^3\mathrm{e}^{-x^2}}_{=\frac{1}{2}} \\
&= 4\sqrt{\frac{k_BT}{2\pi m}}
\end{aligned}$$

The value of the Boltzmann constat is $k_B \approx 1.38\times10^{-23}\frac{\mathrm{J}}{\mathrm{K}}$. Air consists predominantly of nitrogen molecules N_2. The atomic weight of a nitrogen atom is 14amu. 1amu is the **atomic mass unit**[17] or **universal mass unit** which has the value $1\text{amu}\approx1.66\times10^{-27}\text{kg}$.

Thus we obtain as the average speed of a nitrogen molecule

$$\begin{aligned}
\langle|v|\rangle &= 4\sqrt{\frac{1.38\times10^{-23}\frac{\mathrm{J}}{\mathrm{K}}300\mathrm{K}}{2\times3.14\times2\times14\text{amu}\times1.66\times10^{-27}\frac{\text{kg}}{\text{amu}}}} \\
&= 4\sqrt{\frac{1.38\times3}{2\times3.14\times2\times14\times1.66}}\sqrt{\frac{10^{-23}10^{2}}{10^{-27}}}\sqrt{\frac{\frac{\mathrm{J}}{\mathrm{K}}\mathrm{K}}{\text{amu}\frac{\text{kg}}{\text{amu}}}} \\
&= 0.48\times10^3\frac{m}{s} = 1700\frac{\text{km}}{\text{h}}
\end{aligned}$$

Thus the average velocity of a gas particle is about 1.4 times the speed of sound or mach 1.5!

This result is at first surprising. However, it is intuitively clear that the gas molecules must be faster than the speed of sound, because a sound wave propagates via gas molecules.

6.7 Ensembles

We can now select statistical mixtures that maximize the entropy for a given set of constraints. Such mixtures are called **ensembles**. The French word "Ensemble" is translated in German as "Gesamtheit" and in English as "the whole". One can think of an ensemble as a particular macrostate, that is all possible states of the system with their probabilities. It is often used in connection with certain couplings of a system with its surrounding.

---

computer are reached already for fairly small systems.

[17] The term atomic mass unit is very misleading because it is easily confused with the mass unit of the Hartree atomic unit system. The latter is the mass of an electron, which is about 1900 times smaller than 1amu.

- Consider a system that does not have any interaction with its surrounding. If it is in thermal equilibrium, we call its macrostate a **micro-canonical ensemble**. Such a system is, in general, characterized by its energy and the number of particles.
- Consider a system that can exchange heat with its surrounding. If it is in thermal equilibrium, we call its macrostate a **canonical ensemble**. Such a system can be characterized by its temperature and its number of particles.
- Consider a system that can exchange heat and particles with its surrounding. If it is in thermal equilibrium, we call its macrostate a **macro canonical ensemble**. Such a system can be characterized by its temperature and the **chemical potentials** of its particles.

Since these three ensembles are frequently we discuss them in detail.

6.7.1 The micro-canonical ensemble

The micro-canonical ensemble describes a completely isolated system. There is no thermal coupling towards the outside nor can particles escape through the walls containing the system.

In the micro-canonical ensemble the energy of the system has a well defined value. It is used for a closed system not in thermal contact with the outside world. The density operator of the micro-canonical ensemble is a mixture of all eigenstates $|\Psi_i\rangle$ of the Hamiltonian with a given energy E

$$\hat{\rho} = \frac{1}{N_E} \sum_{i;E_i=E} |\Psi_i\rangle\langle\Psi_i|$$

where N_E is the number of states with energy E, i.e. the degeneracy of that energy eigenvalue. Because the coupling with the surrounding is completely absent, the energy and particle numbers are conserved. Note that the energy E is identical to the internal energy U. Since the energy has a sharp value, its mean value U is equal to the actual value E. Nevertheless it is customary to use the symbol E in the context of a micro-canonical ensemble.

The micro-canonical ensemble plays an important role for simulations. For example, a molecular dynamics simulation produces a micro-canonical ensemble. In a molecular dynamics simulation we solve Newton's equations of motion, for which the energy is conserved in the absence of friction terms. While the choice of the ensemble becomes irrelevant for very large systems, it makes a difference for small systems that are accessible to simulations. There are techniques to extend such molecular dynamics simulations to other ensembles. For example, the Nose Thermostat[9, 10] allows to construct a canonical ensemble, which will be discussed next.

6.7.2 The canonical ensemble

A canonical ensemble corresponds to a system in thermal contact with its surrounding. The system can exchange energy, but the mean value of the energy is defined, and therefore also its temperature. The temperature of the system is identical to that of the surrounding, which acts as a heat bath. The mean value of the energy is denoted by U and is called the **internal energy**.

The state operator in thermal equilibrium is obtained from the maximum-entropy principle. The maximum-entropy principle requires that the density matrix maximizes the entropy Eq. 6.5 under the constraint that the mean value of the energy is equal to U, and the constraint that the trace of ρ is equal to one.

$$\mathcal{S}_{tot}[\hat{\rho}, U, T, \lambda] = \mathcal{S}[\hat{\rho}] - \frac{1}{T}\Big(\mathrm{Tr}[\hat{\rho}\hat{H}] - U\Big) - \lambda\Big(\mathrm{Tr}[\hat{\rho}] - 1\Big)$$

This equation is only valid of $\partial S_{tot}/\partial\frac{1}{T} = 0$ and $\partial S_{tot}/\partial\lambda = 0$.

Let us define here the **Helmholtz Free energy** [18]. The free energy functional $\mathcal{F}$ is related to the entropy via

$$\mathcal{F}(T,[\hat\rho]) = \underbrace{\mathrm{Tr}[\hat\rho\hat H]}_{U} - T\mathcal{S}[\hat\rho] = \mathrm{Tr}\Big[\hat\rho\hat H - k_BT\hat\rho\ln(\hat\rho)\Big]$$

In order to obtain the density matrix in thermal equilibrium we need to minimize the free energy with respect to the desnity matrix for the given energy and normalization constraints.

$$\mathcal{F}(T,[\hat\rho]) = \mathrm{Tr}[\hat\rho\hat H] + k_BT\mathrm{Tr}[\hat\rho\ln[\hat\rho]] + \lambda T\Big(\mathrm{Tr}[\hat\rho]-1\Big)$$

The minimum free energy principle determines the density matrix as

$$\frac{\partial\mathcal{F}}{\partial\hat\rho} = 0 \qquad \Rightarrow \qquad \hat\rho_{eq} = \mathrm{e}^{-(1+\lambda/k_B)}\mathrm{e}^{-\frac{1}{k_BT}\hat H} = \frac{\mathrm{e}^{-\frac{1}{k_BT}\hat H}}{\mathrm{Tr}[\mathrm{e}^{-\frac{1}{k_BT}\hat H}]}$$

where we inserted λ in order to enforce the normalization condition.

If we insert the state operator that fulfills the minimum free energy principle we obtain the free energy $F(T)$ as

$$\begin{aligned}F(T) &= \max_{\hat\rho}\mathcal{F}(T,\hat\rho) = \mathcal{F}(T,\hat\rho_{eq})\\ &= \mathrm{Tr}[\Big(\hat H + k_BT\ln[\hat\rho_{eq}]\Big)\hat\rho_{eq}]\\ &= \mathrm{Tr}[\Big(\hat H + k_BT\ln[\mathrm{e}^{-\frac{\hat H}{k_BT}}] - k_BT\ln[Z]\Big)\hat\rho_{eq}]\\ &= -k_BT\ln\big[\mathrm{Tr}[\mathrm{e}^{-\frac{1}{k_BT}\hat H}]\big]\end{aligned}$$

Note the difference between $F(T)$ and the free energy functional $\mathcal{F}(T,[\hat\rho])$. $F(T)$ is the maximum value of the functional $\mathcal{F}(T,[\hat\rho])$ for all density matrices that fulfill the constraints.

FREE ENERGY

The free energy can be written as

$$F(T) = -k_BT\ln[Z(T)]$$

where the partition function is defined as

$$Z(T) = \mathrm{Tr}[\mathrm{e}^{-\frac{\hat H}{k_BT}}]$$

In a representation of eigenstates $|\Psi_i\rangle$ of the Hamiltonian $\hat H$, the partition function can be expressed by the eigenvalues of the Hamiltonian.

$$Z(T) = \mathrm{Tr}[\mathrm{e}^{-\frac{\hat H}{k_BT}}] = \sum_i \mathrm{e}^{-\frac{E_i}{k_BT}}$$

In this representation, the density operator in thermal equilibrium has the form

$$\hat\rho_{eq}(T) = \frac{1}{Z(T)}\sum_i |\Psi_i\rangle\mathrm{e}^{-\frac{E_i}{k_BT}}\langle\Psi_i|$$

---

[18] Helmholtz**Editor: Here some biographical stuff**

6.7.3 The grand-canonical ensemble

The grand canonical ensemble corresponds to a system in thermal contact with its surrounding in such a way that it can exchange energy and particles with it. The surrounding represents a heat bath and a particle reservoir. There may be one particle reservoir for every type of particle that can be exchanged with the environment. The temperature of the system and the chemical potentials for the particle reservoirs have well defined values.

$$\hat{\rho}_{eq}(T,\mu)=\sum_i |\Psi_i\rangle \frac{\mathrm{e}^{-\frac{E_i-\mu N_i}{k_BT}}}{\sum_j \mathrm{e}^{-\frac{E_j-\mu N_j}{k_BT}}}\langle\Psi_i|$$

Here the states are many-particle states including states for all number of particles. This implies that the eigenfunctions span the **Fock space** and not only the Hilbert space for a given number of particles. The Fock space is defined as the space spanned by the wave functions for all possible number of particles. Thus it contains one-particle states, two-particle states etc. The Fock space is the superset of the Hilbert spaces for different number of particles.

For the grand canonical ensemble we introduce the grand-canonical potential $\Phi(T,\mu)=-k_BT\ln[Z(T,\mu)]$ with the partition function

$$Z(T,\mu)=\mathrm{Tr}[\mathrm{e}^{-\frac{1}{k_BT}(\hat{H}-\mu\hat{N}}]=\sum_i \mathrm{e}^{-\frac{1}{k_BT}(E_i-\mu N_i)}$$

6.8 Dynamics

Editor: This section should be placed into a chapter dynamics!

6.9 H-theorem

Our motivation for the second law of thermodynamics, namely the maximum-entropy principle, was based on a purely cognitive concept, namely the assumption that the uncertainty about the state of the system increases. However, there must be a deeper reason for this concept of increasing entropy, because the dynamics of a physical system seems to drive a system objectively towards a state of maximum entropy. Hence the origin of the entropy increase must be contained within the dynamical equations of motion.

When look at the equations of motion such as the Schrödinger equation or Newton's equations of motion we immediately notice a paradox: The equations of motion are time reversible which contradicts a quantity that grows only in one direction of the time arrow. The explanation of irreversible behavior from the microscopic equations is non-trivial.

At the core of the origin of irreversible behavior is the **H-theorem** (pronounced "Eta"-theorem [19])of Boltzmann. What is shown here is the quantum-mechanical version of the H-theorem of Boltzmann.

We consider the dynamics beginning with a unique initial state described by a density operator ρ_0 under an *ensemble of hamiltonians*. The ensemble of hamiltonians describes our uncertainty of the dynamics, that is the coupling to an environment.

We show first that the entropy remains constant, if the dynamics is governed by a single Hamiltonian, that may or may not change with time. The dynamical equation for the density matrix is the

---

[19] The H-theorem is pronounced "Eta-theorem". The greek letter Eta is written "H". There seems to be some confusion about the origin of the letter H and the name Eta-theorem[11].

von Neumann equation $i\hbar\partial_t\hat\rho = [\hat H, \hat\rho]$, which follows directly from the Schrödinger equation.

$$\begin{aligned}
\frac{dS}{dt} &\overset{\text{Eq. 6.5}}{=} \frac{d}{dt}\left(-k_B\mathrm{Tr}[\hat\rho\ln(\hat\rho)]\right) = -k_B\mathrm{Tr}[\left(\frac{d}{dt}\hat\rho\right)(\ln(\hat\rho)+1)] \\
&\overset{i\hbar\dot{\hat\rho}=[\hat H,\hat\rho]}{=} -k_B\mathrm{Tr}[\left(-\frac{i}{\hbar}[\hat H,\hat\rho]\right)\ln(\hat\rho)] - k_B\frac{d}{dt}\underbrace{\mathrm{Tr}[\hat\rho]}_{=1} \\
&= \frac{ik_B}{\hbar}\mathrm{Tr}\left[\hat H\hat\rho\ln(\hat\rho) - \hat\rho\hat H\ln(\hat\rho)\right] \\
&\overset{\text{cycl.perm}}{=} \frac{ik_B}{\hbar}\mathrm{Tr}\left[\hat H\hat\rho\ln(\hat\rho) - \hat H\ln(\hat\rho)\hat\rho\right] \\
&= \frac{ik_B}{\hbar}\mathrm{Tr}\left[\hat H\underbrace{[\hat\rho,\ln(\hat\rho)]}_{=0}\right] = 0
\end{aligned} \tag{6.51}$$

We have shown with Eq. 6.51 that the entropy of a closed system remains constant. This reflects that the dynamics of quantum mechanics, i.e. the Schrödinger equation or the von Neumann equation, is deterministic. Important is that the Hamiltonian may be change with time. Thus even a fluctuating external potential will not increase the entropy.

In order to produce irreversible behavior, i.e. an entropy production, one can average the dynamics over an ensemble of Hamiltonians. As will be shown, the ignorance about the time evolution leads to an increasing ignorance about the states of the system, and hence an entropy increase.

The mathematical basis for the proof of entropy increase is the property[12] of the entropy that it is concave[20]:

$$S[\sum_k c_k\hat\rho_k] \geq \sum_k c_k S[\hat\rho_k] \qquad \text{with} \qquad \sum_k c_k = 1 \tag{6.52}$$

---

[20] A function is concave if it bends downward (negative curvature). A more precise definition, that also holds for non-continuous functions is the following: A function $f(x)$ is concave, if

$$f(x) \geq f(a) + \frac{x-a}{b-a}(f(b)-f(a)) \qquad \text{with} \qquad x\in[a,b]$$

Here we show that the entropy is concave, that is we prove Eq. 6.52. It is sufficient to show Eq. 6.52 for two density matrices ρ_1 and ρ_2

$$S(x\hat{\rho}_1 + (1-x)\hat{\rho}_2) \geq xS(\hat{\rho}_1) + (1-x)S(\hat{\rho}_2)$$

where x is any number between 0 and 1.
For two hermitian operators $\hat{A}$ and $\hat{B}$ with non-negative eigenvalues a and b and normalized eigenstates $|a\rangle$ and $|b\rangle$, that is $\hat{A}|a\rangle = |a\rangle a$ and $\langle a|a'\rangle = \delta_{a,a'}$ and similar expressions for $\hat{B}$, we can show that

$$\begin{aligned}
&\mathrm{Tr}[\hat{A}\ln\hat{B}] - \mathrm{Tr}[\hat{A}\ln\hat{A}]\\
&= \mathrm{Tr}\left[\left(\sum_a |a\rangle a\langle a|\right)\ln\left(\sum_b |b\rangle b\langle b|\right)\right] - \mathrm{Tr}\left[\left(\sum_a |a\rangle a\langle a|\right)\ln\left(\sum_{a'} |a'\rangle a'\langle a'|\right)\right]\\
&= \sum_a a\langle a|\left(\sum_b |b\rangle\ln(b)\langle b|\right)|a\rangle - \sum_a a\langle a|\left(\sum_{a'} |a'\rangle\ln(a')\langle a'|\right)|a\rangle\\
&= \sum_{a,b} a\langle a|b\rangle\ln(b)\langle b|a\rangle - \sum_{a,a'} a\langle a|a'\rangle\ln(a')\underbrace{\langle a'|a\rangle}_{=\delta_{a,a'}}\\
&= \sum_{a,b} a\ln(b)|\langle b|a\rangle|^2 - \sum_a a\ln(a)\underbrace{\sum_b|\langle b|a\rangle|^2}_{=1}
\end{aligned}$$

We used the inequality $\ln(x) \leq x-1$, which is valid for $x > 0$. Furthermore, the equality $\ln(x) = x-1$ only holds for $x = 1$. Thus the equality in the above equation holds only if the eigenvalues a and b are identical, whenever, $\langle a|b\rangle \neq 0$. It is possible to convince oneself, that this can only be achieved, if the eigenvectors are identical, except for a unitary transformation within one multiplet of degenerate states. If eigenvalues and eigenvectors are identical the two operators are identical. Hence the equality only holds if $\hat{A} = \hat{B}$.

$$\begin{aligned}
&\mathrm{Tr}[\hat{A}\ln\hat{B}] - \mathrm{Tr}[\hat{A}\ln\hat{A}]\\
&= \sum_{a,b}\left(a\ln(b) - a\ln(a)\right)|\langle b|a\rangle|^2\\
&= \sum_{a,b} a\underbrace{\ln(b/a)}_{\leq b/a-1}|\langle b|a\rangle|^2\\
&\leq \sum_{a,b}(b-a)|\langle b|a\rangle|^2\\
&= \sum_a\langle a|\left(\sum_b |b\rangle b\langle b\right)|a\rangle - \sum_b\langle b|\left(\sum_a |a\rangle a\langle a\right)|b\rangle\\
&= \mathrm{Tr}[B] - \mathrm{Tr}[A]
\end{aligned}$$

. . . cont'd

cont'd...

If we apply

$$\mathrm{Tr}[\hat{A}\ln\hat{B}]-\mathrm{Tr}[\hat{A}\ln\hat{A}]\leq\mathrm{Tr}[B]-\mathrm{Tr}[A]$$

to two density matrices with $\mathrm{Tr}[\hat{\rho}_1]=\mathrm{Tr}[\hat{\rho}_2]=1$, we obtain

$$\mathrm{Tr}[\hat{\rho}_1\ln(\hat{\rho}_2)]\leq\mathrm{Tr}[\hat{\rho}_1\ln(\hat{\rho}_1)] \tag{6.53}$$

Note that the eigenvalues of a density matrix are non-negative, so that the requirements of the above derivation are fulfilled.

$$\begin{aligned}S(x\rho_1+(1-x)\rho_2) &= -k_B\mathrm{Tr}\left[\Big(x\rho_1+(1-x)\rho_2\Big)\ln\Big(x\rho_1+(1-x)\rho_2\Big)\right]\\ &= -k_Bx\,\underbrace{\mathrm{Tr}\left[\overbrace{\rho_1}^{\hat{A}}\ln(\overbrace{x\rho_1+(1-x)\rho_2}^{\hat{B}})\right]}_{\leq\mathrm{Tr}[\rho_1\ln(\rho_1)]}\\ &\quad -k_B(1-x)\underbrace{\mathrm{Tr}\left[\rho_2\ln(x\rho_1+(1-x)\rho_1)\right]}_{\leq\mathrm{Tr}[\rho_2\ln(\rho_2)]}\\ &\overset{\text{Eq. 6.53}}{\geq} xS(\rho_1)+(1-x)S(\rho_2)\end{aligned}$$

which proves the concave nature of the entropy ...*q.e.d.*

Once we have established that the entropy is concave, the entropy increase for an ensemble of Hamiltonians follows relatively quickly: Let us now propagate the same initial density matrix $\hat{\rho}_0$ with different Hamilton operators $\hat{H}_k$. We obtain a number of density matrices $\hat{\rho}_k(t)$. If the Hamilton operators have positive weights c_k we obtain with $\sum_k c_k=1$

$$\begin{aligned}S[\sum_k c_k\hat{\rho}_k(t)] &\overset{Eq.\ 6.52}{\geq} \sum_k c_kS[\hat{\rho}_k(t)] \overset{E.\ 6.51}{=} \sum_k c_kS[\hat{\rho}_k(0)]=\sum_k c_kS[\hat{\rho}_0]\\ &= \underbrace{\left(\sum_k c_k\right)}_{=1}S[\underbrace{\left(\sum_k c_k\right)}_{=1}\hat{\rho}_0]=S[\sum_k c_k\hat{\rho}_k(0)]\end{aligned}$$

Thus we obtain

$$S(t)\geq S(0)$$

Thus the entropy can only increase with time. Since the equality only holds for identical density matrices, the above relation always holds, if the Hamiltonians lead to different time evolutions of the density matrices.

Why should we need to average over different Hamiltonians? Let me give an argument that I have not fully worked out and which should be taken with a grain of salt. If we consider our system as a subsystem of the universe, the environment, i.e. the rest of the universe, acts like an ensemble of Hamiltonians for the wave function of our subsystem. This interpretation implies that the overall entropy of the universe remains constant, but our system can transport information into the environment.

6.10 Summary

In this section we have learned that a system can be described by a density operator $\hat{\rho}$. The density operator for a particular system can be determined from the maximum-entropy principle, if the Hamiltonian of the system is known. If the state is describes by the resulting density operator $\hat{\rho}_{eq}$, the system is in thermal equilibrium. Depending on the ensemble that is chosen, the maximum-entropy principle can be converted into a minimum principle of a thermodynamic potential such as the free energy principle. The advantage of the free energy principle over the maximum-entropy principle is that it is, as most other physical theories, centered around an energy.

Once the density operator is known one can evaluate the expectation value of any observable from it as $\langle A\rangle = Tr[\hat{\rho}\hat{A}]$. We can also avoid any reference to individual microstates, if the free energy for an extended Hamiltonian $\hat{H}(\gamma) = \hat{H}_0 + \alpha\hat{A}$ is known. Then we can evaluate the mean values directly from the minimum free energy $F(T,\alpha)$ using Eq. **??**.

Because all macroscopic properties can be obtained from the free energy $F = -k_BT\ln[Z]$, we will spend some time in the next section to evaluate partition functions and the resulting thermodynamic potentials. Once they are known, we can describe the state of the system by only a few parameters, namely the extensive and intensive variables chosen in the maximum-entropy principle.

There are different ensembles, which depend on whether we use a fixed value of an operator or the mean value as constraint in the maximum-entropy principle. In the thermodynamic limit, that is for very large systems, the difference between the ensembles becomes irrelevant, because the variance of the expectation value vanishes.

While the theory has been based on a quantum mechanical description, we have finally shown how the classical partition function can be derived as limiting case $\hbar \to 0$.

Editorial remark: compare the concept of a state, an equation of motion and an observable between classical mechanics, quantum mechanics and statistical mechanics

6.10.1 Important formulas

$$
\begin{aligned}
\hat{\rho} &= \sum_i |\Psi_i\rangle P_i \langle\Psi_i| \\
\hat{\rho} &= \hat{\rho}^\dagger \\
\mathrm{Tr}[\hat{\rho}] &= 1 \\
\forall_{|\Psi\rangle} \langle\Psi|\hat{\rho}|\Psi\rangle &\geq 0 \\
\hat{\rho}^2 &= \hat{\rho} \qquad \text{idempotency for pure states} \\
S[\hat{\rho}] &= -k_B \mathrm{Tr}[\hat{\rho}\ln(\hat{\rho})] \\
\hat{\rho}(T, f_i) &= \frac{1}{Z(T, f_i)} \mathrm{e}^{-\frac{1}{kT}(\hat{H} - \sum_k f_k \hat{X}_k)} \\
\langle A\rangle &= \mathrm{Tr}[\hat{\rho}\hat{A}] \\
\frac{d\langle X_i\rangle}{df_j} &= \frac{1}{k_B T} \Big\langle (X_i - \langle X_i\rangle)(X_j - \langle X_j\rangle) \Big\rangle \\
\Big\langle (E - \langle E\rangle)^2 \Big\rangle &= k_B T^2 c_V \\
Z(T) &= \int dx \int \frac{dp}{2\pi\hbar}\, \mathrm{e}^{-\beta H(p,x)} \\
Z(T) &= \int \frac{dx}{\lambda_T} \mathrm{e}^{-\beta V(x)} \\
\langle A\rangle_T &= \int dx \int dp\; A(p,x) P(p,x,T) \\
P(p,x,T) &= \frac{\mathrm{e}^{-\beta H(p,x)}}{\int dx \int \frac{dp}{2\pi\hbar}\, \mathrm{e}^{-\beta H(p,x)}}
\end{aligned}
$$

Chapter 7

Non-interacting Model Systems

In this chapter we will work out the free energy and other thermodynamic quantities for a few non-interacting systems. The interaction between subsystems adds substantial complexity and also adds a number of new physical effects such as phase transitions. Without interactions we can often find analytical solutions. Therefore the non-interacting systems are the starting point for the investigation of interacting systems. There are a number of systems where interactions are of secondary importance. We will see later that one can describe interacting systems approximately by treating them as non-interacting with thermodynamic forces, that encapsulate the interaction in an effective way.

I will select a small number of model systems, that allow us to understand a wealth of physical phenomena. The model systems are (1) the two state system, (2) the harmonic oscillator, and (3) the free particle. Most problems in physics can be regarded as generalizations or modifications of these model systems.

Non-distinguishable particles add some complexity, because their behavior is not uncorrelated, even if they do not have an interaction. An example is the Pauli principle[1][2], which does not allow two Fermions to be at the same position. The reason is that the particle nature of Fermions and Bosons is just one aspect of the corresponding particle field, as it is described in quantum field theory. The model systems mentioned above will allow us, however, to describe their behavior in a natural way.

7.1 Two state system

The two state system can be applied to very different problems such as isolated spins in a magnetic field or to estimate the concentration of dilute defects in a solid or the adsorption of atoms on a surface.

Therefore we will first study the two state model in the abstract form and later show its application to physical problems.

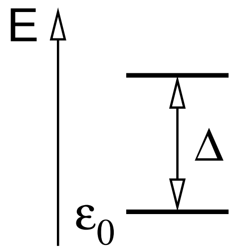
7.1.1 Thermodynamic potentials of the two state system

The system has, per site, two possible states with different energies. The ground state is denoted by ϵ_0 and the second state is denoted by $\epsilon_0 + \Delta$, so that Δ is the energy spacing. A state is characterized on each site by a parameter σ, which can have values 0 and 1. σ specifies if the system is in the ground state or the excited state. A state with N_s sites is therefore characterized by an N_s dimensional vector $\vec{\sigma}$.

---

[1] Wolfgang Pauli. Austrian Physicist 1900-??. Nobel price in Physics 1945. Professor for Theoretical Physics at the Federal Institute of Technology (ETH) Zurich. Professor at Princeton, USA 1935-1945. Returned to Zurich in 1945.

[2] The Pauli principle says that due to the antisymmetry of the wave function of Fermions, no two Fermions can occupy the same one-particle state.



The total energies are thus

$$E_{\vec{\sigma}} = \sum_{i=1}^{N_s} \epsilon_0 + \sigma_i \Delta \qquad \text{with} \qquad \sigma_i \in \{0,1\} \tag{7.1}$$

In addition we want to track the number of excited states

$$N_{\vec{\sigma}} = \sum_{i=1}^{N_s} \sigma_i \tag{7.2}$$

In the following we also allow to fix the number of excitations, which introduced the chemical potential as the Lagrange parameter for the number constraint. Fixing this number does not carry any physical meaning for excitations. If the number of excitations is not fixed we set μ=0, which effectively removes the constraint.

The rationale for introducing this flexibility is that in second quantization, certain excitations are identified with particles, which may be conserved. In order to prepare for this parallelism, we take this possibility into account.

Partition function and free energy

Now we can construct the partition function

$$\begin{aligned} Z(T,\mu) &= \sum_{\vec{\sigma}} \mathrm{e}^{-\beta(E_{\vec{\sigma}}-\mu N_{\vec{\sigma}})} = \sum_{\vec{\sigma}} \prod_{i=1}^{N_s} \mathrm{e}^{-\beta(\epsilon_0+\sigma_i\Delta-\mu\sigma_i)} \\ &= \prod_{i=1}^{N_s} \sum_{\sigma_i=0}^{1} \mathrm{e}^{-\beta(\epsilon_0+\sigma_i\Delta-\mu\sigma_i)} \\ &= \left[\mathrm{e}^{-\beta\epsilon_0}\left(1+\mathrm{e}^{-\beta(\Delta-\mu)}\right)\right]^{N_s} \end{aligned}$$

For the interchange of sum and product see below.[3]

From the partition function we immediately obtain the free energy of the two state system as

FREE ENERGY OF THE TWO-STATE SYSTEM

$$F(T,\mu) = -k_B T \ln[Z(T,\Delta,\mu)] = N_s \left\{\epsilon_0 - k_B T \ln\left[1+\mathrm{e}^{-\beta(\Delta-\mu)}\right]\right\} \tag{7.3}$$

---

[3] The terms have the following structure

$$\begin{aligned} \sum_{\vec{\sigma}} \prod_{i=1}^{n} f(\sigma_i) &= \sum_{\sigma_1} \cdots \sum_{\sigma_n} f(\sigma_1)\cdots f(\sigma_n) \\ &= \left(\sum_{\sigma_1} f(\sigma_1)\right) \cdots \left(\sum_{\sigma_n} f(\sigma_n)\right) = \prod_{i=1}^{n} \left(\sum_{\sigma_i} f(\sigma_i)\right) \end{aligned}$$

Number of excitations

Let us now determine the equations of state for the particle number N, which is the number of excitations

$$
\begin{aligned}
N(T,\mu) &= -\frac{\partial F}{\partial \mu} = N_s k_B T \frac{1}{1+\mathrm{e}^{-\beta(\Delta-\mu)}} \left(\frac{1}{k_B T}\right) \mathrm{e}^{-\beta(\Delta-\mu)} \\
&= N_s \frac{1}{1+\mathrm{e}^{+\beta(\Delta-\mu)}}
\end{aligned}
$$

The fractional occupation $f = N/N_s$ of the two-state system is nothing but the **Fermi distribution function**, which will play an important role in the statistics of Fermions

$$
f = \frac{N(T,\mu)}{N_s} = \frac{1}{1+\mathrm{e}^{+\beta(\Delta-\mu)}} \tag{7.4}
$$

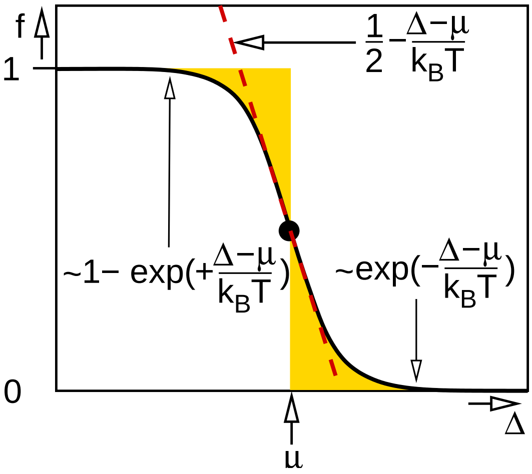


Fig. 7.1: Fermi distribution.

The Fermi-distribution function has a step-function like character. In the low-temperature limit, that is $k_BT << |\Delta-\mu|$, the Fermi distribution function deviates from the step function only by the Boltzmann factor. In the high temperature limit $k_BT >> |\Delta-\mu|$ its value approaches $\frac{1}{2}$ corresponding to a completely random occupation of ground and excited state. The width of the cross-over of Fermi distribution function is proportional to k_BT.[4]

$$
\begin{aligned}
f &= \frac{1}{1+\mathrm{e}^{+\beta(\Delta-\mu)}} \xrightarrow{\Delta-\mu>>k_BT} \mathrm{e}^{-\beta(\Delta-\mu)} \\
f &= \frac{1}{1+\mathrm{e}^{+\beta(\Delta-\mu)}} \xrightarrow{-(\Delta-\mu)<<-k_BT} 1-\mathrm{e}^{\beta(\Delta-\mu)} \\
f &= \frac{1}{1+\mathrm{e}^{+\beta(\Delta-\mu)}} \xrightarrow{|\Delta-\mu|<<k_BT} \frac{1}{2}-\frac{\Delta-\mu}{k_BT}
\end{aligned}
$$

Editor: This footnote is saved for later [5].

---

[4] As calculated from the tangent at $\Delta=\mu$.

[5]

$$
1-\tanh(x) = 1-\frac{\mathrm{e}^x-\mathrm{e}^{-x}}{\mathrm{e}^x+\mathrm{e}^{-x}} = \frac{\mathrm{e}^x+\mathrm{e}^{-x}-\mathrm{e}^x+\mathrm{e}^{-x}}{\mathrm{e}^x+\mathrm{e}^{-x}} = \frac{2\mathrm{e}^{-x}}{\mathrm{e}^x+\mathrm{e}^{-x}} = 2\frac{1}{1+\mathrm{e}^{2x}}
$$

Entropy

Let us now determine the equations of state

$$\begin{aligned}S(T,\mu) &= -\frac{\partial F}{\partial T}\\ &\overset{\text{Eq. 7.3}}{=} N_s k_B \ln\left[1+\mathrm{e}^{-\beta(\Delta-\mu)}\right] + N_s k_B T \frac{1}{1+\mathrm{e}^{-\beta(\Delta-\mu)}}\frac{\Delta-\mu}{k_B T^2}\mathrm{e}^{-\beta(\Delta-\mu)}\\ &= N_s k_B \left\{\ln\left[1+\mathrm{e}^{-\beta(\Delta-\mu)}\right] + \frac{\Delta-\mu}{k_B T}\frac{1}{1+\mathrm{e}^{+\beta(\Delta-\mu)}}\right\}\end{aligned}$$

It is sometimes useful to express the entropy directly by the fractional occupation.

$$\begin{aligned}f(x) &= \frac{1}{1+\mathrm{e}^x} \qquad \Rightarrow \qquad \frac{1}{f} = 1+\mathrm{e}^x \qquad \Rightarrow \qquad \mathrm{e}^x = \frac{1}{f}-1 = \frac{1-f}{f}\\ 1+\mathrm{e}^{-x} &= 1+\frac{1-f}{f} = \frac{1}{1-f}\\ x &= \ln(\mathrm{e}^x) = \ln\left(\frac{1-f}{f}\right) = \ln(1-f)-\ln(f)\\ \frac{1}{1+\mathrm{e}^x} &= f\end{aligned}$$

Thus the entropy obtains the form

$$\begin{aligned}S(f) &= N_s k_B \left\{\ln\left[1+\mathrm{e}^{-\beta(\Delta-\mu)}\right] + \frac{\Delta-\mu}{k_B T}\frac{1}{1+\mathrm{e}^{+\beta(\Delta-\mu)}}\right\}\\ &= N_s k_B \left\{-\ln(1-f) + [\ln(1-f)-\ln(f)]\,f\right\}\\ &= -N_s k_B \Big\{f\ln(f) + (1-f)\ln(1-f)\Big\}\end{aligned}$$

Thus we can express the entropy as function of the number of excited states $N = N_s f$.

ENTROPY OF THE TWO-STATE SYSTEM

$$S(f) = -N_s k_B\Big[f\ln(f) + (1-f)\ln(1-f)\Big] \tag{7.5}$$

Internal energy

The internal energy can be expressed simply by by counting the number of excited states

$$U = (N_s - N)\epsilon_0 + N(\epsilon_0+\Delta) = N_s\,(\epsilon_0 + f\Delta)$$

However, let us also consider the standard method

$$
\begin{aligned}
U(T,\mu) &= F + TS + \mu N \\
&= \underbrace{\Big(N_s\epsilon_0 - N_s k_B T \ln\Big[1 + \mathrm{e}^{-\beta(\Delta-\mu)}\Big]\Big)}_{F} \\
&\quad \underbrace{+ N_s k_B T \left\{\ln\Big[1 + \mathrm{e}^{-\beta(\Delta-\mu)}\Big] + \frac{\Delta-\mu}{k_B T}\frac{1}{1+\mathrm{e}^{+\beta(\Delta-\mu)}}\right\}}_{TS} \\
&\quad \underbrace{+ \mu N_s \frac{1}{1+\mathrm{e}^{+\beta(\Delta-\mu)}}}_{\mu N} \\
&= N_s\epsilon_0 + N_s\Delta \underbrace{\frac{1}{1+\mathrm{e}^{\beta(\Delta-\mu)}}}_{f}
\end{aligned}
\tag{7.6}
$$

which is identical to what we derived above as expectation value of the energy.

Number of excitations from the minimum free energy principle

It is instructive and often useful to consider the free energy as a functional of the fractional number of excitations, and to obtain the value of the latter by minimizing the free energy.

The free energy is given as

$$
\begin{aligned}
F(T,\mu,f) = U - TS - \mu N &= N_s\Big\{\epsilon_0 + \Delta f - k_B T\Big[f\ln(f) + (1-f)\ln(1-f)\Big] - \mu f\Big\} \\
&= N_s\Big\{\epsilon_0 + (\Delta-\mu)f - k_B T\Big[f\ln(f) + (1-f)\ln(1-f)\Big]\Big\}
\end{aligned}
$$

In the low-temperature limit, the entropy contribution vanishes and the free energy is a straight line with respect to f. The minimum of the free energy is then either $f = 0$ for $\Delta > \mu$ or $f = 0$ for $\delta < \mu$. When $\Delta - \mu = 0$, any value of f would be compatible with the minimum free energy principle. The result therefor depends on how we approach this state.

If we increase the temperature, the entropy will become important. The entropy favors an occupation of $f = \frac{1}{2}$.

7.1.2 Application: Spin in a magnetic field

Imagine a defect level in the band gap of a semiconductor or insulator, which is occupied by an unpaired electron. The electron has a magnetic moment which changes its energy by $\Delta E = -\vec{\mu}_e\vec{B}$, where μ_e is the magnetic moment of the electron.

The magnetic moment of an isolated electron

$$
\vec{\mu}_e = -\frac{g\mu_B}{\hbar}\vec{S}
$$

is the product of the **Lande g-factor** g, the Bohr magneton $\mu_B = \frac{e\hbar}{2m_e} \approx 9 \times 10^{-24} J/T$ and the electron spin S, which can have values $\pm\hbar/2$ along a given axis. The g-factor for a free electron is $g = 2.0023$.[6] Due to the negative charge, the magnetic moment of the electron is antiparallel to its angular momentum.

The magnetic moment of the electron interacts with an applied magnetic field, which favors a parallel alignment of the magnetic moment with the magnetic field. The energy of a spin in the

---

[6] Experimentalists usually use $g = 2$ for the free electron in their analysis!

magnetic field is

$$E = -\vec{\mu}_e\vec{B} = +g\mu_B\frac{1}{\hbar}\vec{S}\vec{B}$$

Thus the system has two eigenstates, namely $\epsilon_0 = -\frac{g}{2}\mu_B B$ and $\epsilon_1 = +\frac{g}{2}\mu_B B$. The splitting $\Delta = g\mu_B B$ between the the two levels is called the Zeeman splitting. It can be determined from the absorption bands of microwaves as function of an applied magnetic field B.

If the experiment is done not on a free electron but an electron in a defect level of a solid, the Zeeman splitting differs slightly from the result expected for a free electron. This deviation is mapped onto the g-factor, which now depends on the particular defect. The g-factor depends furthermore on the orientation of the defect relative to the magnetic field. Thus the g-factor is a tensor. The g factor is used to characterize defects. The symmetry of the g-factor reflects the symmetry of the defect and provides additional insight into the atomic structure of the defect.

What we discussed here is the principle of the Electron Spin Resonance (ESR) or Electron Paramagnetic Resonance (EPR) technique. The thermodynamic properties of these spins are also used the the Electrically Detected Magnetic Resonance (EDMR) technique. Here, a current of magnetized electrons is driven through a defect level. Due to the Pauli principle, an electron can tunnel into the defect only, if its spin is antiparallel to that already residing in the defect. Thus the defect mediated tunneling mechanism is basically blocked at low temperatures, because the tunneling electrons and those in the defect are aligned parallel due to the magnetic field. If, however, a microwave irradiation is in resonance with the Zeeman splitting, it flips the spin in the defect, resulting in a random orientation of the spins in the defect. Thus the current due to defect mediated tunneling is enhanced. This technique is particularly useful, if the probes are so small that absorption measurements can not be done. This is the case, for example, when one studies the defect in gate oxides of transistors. In order to make the technique quantitative, one needs to understand the alignment of the spins in the defects as function of temperature and magnetic fields.

Now we need to translate this problem into our two state model. We obtain

$$\begin{aligned}
\epsilon_0 &= E(S_z = -\frac{\hbar}{2}) = g\mu_B\frac{1}{\hbar}\underbrace{-\frac{\hbar}{2}}_{S_z}\underbrace{B}_{B_z} = -\mu_e B_z \\
\Delta &= E(S_z = +\frac{\hbar}{2}) - E(S_z = -\frac{\hbar}{2}) = 2g\mu_B\frac{1}{\hbar}\underbrace{B}_{B_z}\underbrace{\frac{\hbar}{2}}_{S_z} = 2\mu_e B_z \\
\mu &= 0
\end{aligned} \tag{7.7}$$

where

$$\mu_e = \frac{1}{2}g\mu_B$$

without vector arrow is the absolute value of the magnetic moment of the electron. It should not be confused with the magnetic moment of the free electron, which is a fundamental constant. In the ground state with energy ϵ_0, the magnetic moment is oriented parallel to the magnetic field, but the spin is antiparallel. In the excited state, the moment points parallel to the magnetic field. The magnetic moment in the direction of the magnetic field of the sample is

$$m_{\vec{\sigma}} = \mu_e\left(N_s - 2\sum_{i=1}^{N_s}\sigma_i\right) = \mu_e N_s - 2\mu_e N_{\vec{\sigma}} \tag{7.8}$$

The chemical potential μ has been set to zero, because there is no conservation law for the number of excited states or the magnetic moment. It will be useful for other problems.

We got rid of the vector notation: If we choose a basisset where the spin of the basis function is either parallel or antiparallel to the magnetic field, all vector quantities are aligned.

The total energy can now be expressed by

$$E_{\vec{\sigma}} \overset{Eq.\ 7.7}{=} \sum_i \mu_e B(2\sigma_i - 1) \overset{Eq.\ 7.8}{=} -m_{\vec{\sigma}} B \tag{7.9}$$

where σ_i is either 0 or 1.

This total energy describes the **Langevin paramagnetism**, which is also found in a gas of atoms with unpaired electrons, or in certain salts with lanthanide of actinide ions, which have partially occupied shells of f-electrons. These f-electrons are localized closely to the nucleus, so that they cannot form metallic bands. Because electrons in the localized f-shell have a strong electrostatic interaction among each other, only a fixed number of electrons can fit into the shell. The two state model can easily be generalized to higher angular momenta to describe magnetic atoms with a larger moment.

7.1.3 Thermodynamic functions

Let us translate the free energy of the two state model into our model

$$\begin{aligned}
F(T,B) &\overset{Eq.\ 7.3}{=} N_s\left\{\epsilon_0 - k_BT\ln\left[1+\mathrm{e}^{-\beta(\Delta-\mu)}\right]\right\} \\
&\overset{Eq.\ 7.7}{=} N_s\left\{-\mu_eB - k_BT\ln\left[1+\mathrm{e}^{-2\beta\mu_eB}\right]\right\} \\
&= N_s\left\{-\mu_eB - k_BT\ln\left[\mathrm{e}^{-\beta\mu_eB}\left(\mathrm{e}^{+\beta\mu_eB}+\mathrm{e}^{-\beta\mu_eB}\right)\right]\right\} \\
&\overset{\ln(ab)=\ln a+\ln b}{=} N_s\left\{\underbrace{-\mu_eB - k_BT\left(-\frac{\mu_eB}{k_BT}\right)}_{=0} - k_BT\ln\left[\underbrace{\mathrm{e}^{+\beta\mu_eB}+\mathrm{e}^{-\beta\mu_eB}}_{2\cosh(\beta\mu_eB)}\right]\right\} \\
&= -N_sk_BT\ln\left[2\cosh\left(\frac{\mu_eB}{k_BT}\right)\right]
\end{aligned} \tag{7.10}$$

Note that the free energy now also depends on the magnetic field.

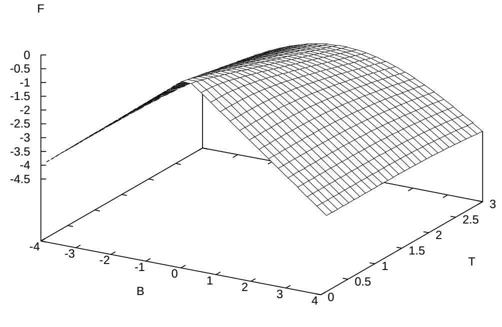


Fig. 7.2: Free energy of a two state system as function of magnetic field and temperature

7.1.4 Susceptibility of a paramagnet

Now we are able to derive some physically relevant quantities. Let us determine the magnetic susceptibility of a paramagnet. A system of dilute spins is a paramagnet.

Let us first review some of the relations from electrodynamics: The magnetic susceptibility is defined as

$$\chi_m = \left.\frac{dM}{dH}\right|_{H=0}$$

where the magnetization M is the magnetic moment per volume, that is

$$M = \frac{m}{V}$$

The magnetic field H is related to B via

$$\begin{aligned}
H &= \frac{1}{\mu_0}B - M \\
B &= \mu_0 (H + M) = \mu_0 \left(H + M_{rem} + \left.\frac{dM}{dH}\right|_{H=0} H + O(H^2) \right) \\
&= \mu_0 M_{rem} + \mu_0 \underbrace{(1 + \chi_m)}_{\mu_r} H + O(H^2) = \mu_0 M_{rem} + \mu_0 \mu_r H + O(H^2)
\end{aligned}$$

μ_0 is the permeability of vacuum and μ_r is the relative permeability. Our system does not have a remanent magnetization so that $M_{rem} = M(H = 0) = 0$. Thus we obtain

$$\begin{aligned}
\chi_m &= \frac{dM}{dB}\frac{dB}{dH} = \mu_0(1 + \chi_m)\frac{dM}{dB} \\
\Rightarrow \chi_m &= \frac{\mu_0 \frac{dM}{dB}}{1 - \mu_0 \frac{dM}{dB}}
\end{aligned}$$

We consider a system of dilute spins. Therefore we can assume that the magnetization and therefore also dM/dB is small. Thus we use in the following

$$\chi_m \approx \mu_0 \frac{dM}{dB} = \frac{\mu_0}{V}\frac{dm}{dB}$$

Thus, we need a relation between the magnetization and the magnetic field. From the free energy

$$F(T, B) \overset{Eq.~7.9}{=} -k_B T \ln \left[\sum_{\vec{\sigma}} \mathrm{e}^{-\frac{1}{k_B T}(-m_\sigma B)} \right]$$

we obtain the magnetic moment via

$$\frac{dF}{dB} = -\sum_{\vec{\sigma}} P_{\vec{\sigma}}(T, B) m_{\vec{\sigma}} = -m(T, B)$$

Thus the total magnetic moment can be obtained from the free energy as

$$m = -\frac{\partial F}{\partial B}$$

Note that this equation is an an equation of state. Thus magnetization and magnetic field are conjugate quantities.

From the formula given before we can also see that the magnetization goes into saturation for for strong fields and $dM/dB \to 0$ for $B >> k_B T/\mu_e$.

For the two state model we obtain the magnetic moment

$$m(T, B) = -\frac{\partial F}{\partial B} \overset{Eq.~7.10}{=} N_s \mu_e \tanh\left(\frac{\mu_e B}{k_B T}\right) \tag{7.11}$$

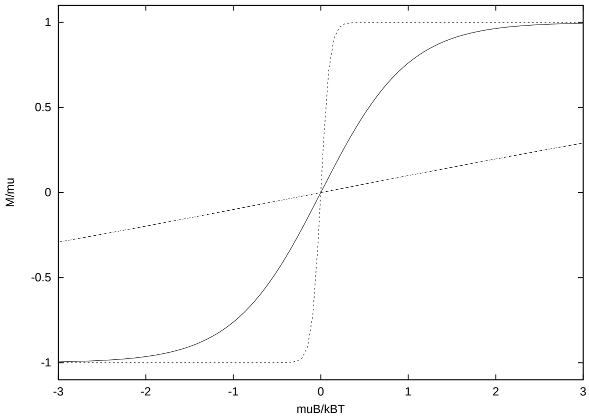


Fig. 7.3: Magnetic moment as function of magnetic field at finite temperature.

and from this we can directly evaluate the susceptibility. (For the derivative of the tanh see[7].)

$$\begin{aligned}\chi_m &= \frac{\mu_0}{V}\frac{dm}{dB} = \frac{\mu_0 N_s \mu_e}{V}\left[1-\tanh^2\left(\frac{\mu_e B}{k_B T}\right)\right]\left(\frac{\mu_e}{k_B T}\right)\\ &= \frac{\mu_0 N_s \mu_e^2}{k_B T V}\left[1-\tanh^2\left(\frac{\mu_e B}{k_B T}\right)\right]\\ &= \frac{\mu_0 N_s \mu_e^2}{k_B T V}\frac{1}{\cosh^2\left(\frac{\mu_e B}{k_B T}\right)}\end{aligned}$$

Thus we obtain the magnetic susceptibility at zero field as

$$\chi_m = \frac{\mu_0 N_s \mu_e^2}{k_B T V}$$

We can see the strong temperature dependence of a paramagnet. The susceptibility disappears with increasing temperature.

---

[7]

$$\begin{aligned}\partial_x \tanh(x) &= \partial_x \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^x + e^{-x}}{e^x + e^{-x}} - \frac{(e^x - e^{-x})^2}{(e^x + e^{-x})^2} = 1 - \tanh^2(x)\\ &= \frac{(e^x + e^{-x})^2}{(e^x + e^{-x})^2} - \frac{(e^x - e^{-x})^2}{(e^x + e^{-x})^2} = \frac{4}{(e^x + e^{-x})^2} = \frac{1}{\cosh^2(x)}\end{aligned}$$

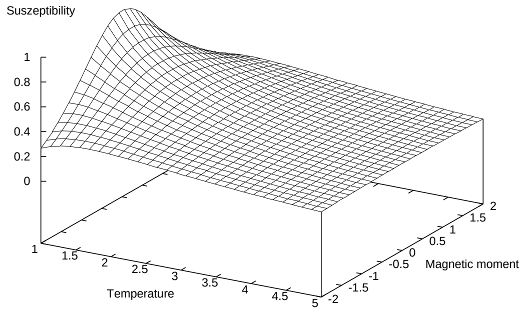


Fig. 7.4: Magnetic susceptibility as function of temperature and magnetic field.

7.1.5 Heat capacity

Now we can determine the internal energy.

$$
\begin{aligned}
U(T,B) &\overset{Eq.\ 7.6}{=} N_s\epsilon_0 + N_s\Delta\frac{1}{1+\mathrm{e}^{\beta(\Delta-\mu)}} \\
&\overset{Eq.\ 7.7}{=} N_s\left\{-\mu_e B + 2\mu_e B\frac{1}{1+\mathrm{e}^{2\beta\mu_e B}}\right\} \\
&= N_s\mu_e B\left\{\frac{1-\mathrm{e}^{2\beta\mu_e B}}{1+\mathrm{e}^{2\beta\mu_e B}}\right\} \\
&= -N_s\mu_e B\frac{\mathrm{e}^{\beta\mu_e B}-\mathrm{e}^{-\beta\mu_e B}}{\mathrm{e}^{\beta\mu_e B}+\mathrm{e}^{-\beta\mu_e B}} \\
&= -N_s\mu_e B\tanh\left(\frac{\mu_e B}{k_B T}\right)
\end{aligned}
$$

By comparison with the above equation Eq. 7.11 we find the expected result $U = -mB$.

The heat capacity at constant magnetic field is defined as

$$
c_B = \frac{\partial U}{\partial T}
$$

It is differential amount of energy taken up when the temperature is increased by a small amount, that is the linear response of the internal energy to a change of temperature.

$$
\begin{aligned}
U(T,B) &= -N_s\mu_e B\tanh\left(\frac{\mu_e B}{k_B T}\right) \\
c_B = \frac{dU}{dT} &= -N_s\mu_e B\left[1-\tanh^2\left(\frac{\mu_e B}{k_B T}\right)\right]\left(-\frac{\mu_e B}{k_B T^2}\right) \\
&= N_s k_B\left(\frac{\mu_e B}{k_B T}\right)^2\left[1-\tanh^2\left(\frac{\mu_e B}{k_B T}\right)\right] \\
&= N_s k_B\frac{\left(\frac{\mu_e B}{k_B T}\right)^2}{\cosh^2\left(\frac{\mu_e B}{k_B T}\right)}
\end{aligned}
$$

The heat capacity approaches zero as the temperature vanishes, in accordance with the third law of thermodynamics.

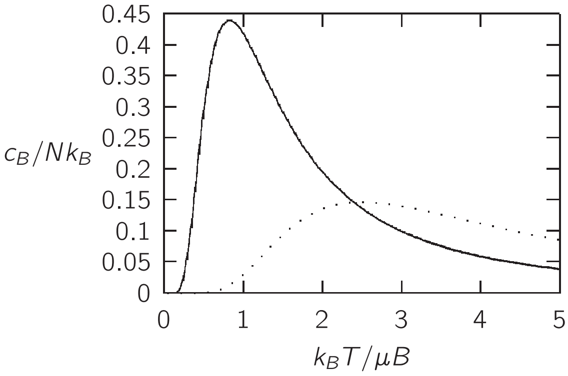


Fig. 7.5: Heat capacity as function of temperature for two values of the magnetic field. The dotted line is for a 3 times larger magnetic field.

7.1.6 Adiabatic demagnetization

The entropy is obtained as

$$S = -\frac{\partial F}{\partial T} = Nk_B \ln[2\cosh(\beta\mu B)] - \frac{N\mu B}{T}\tanh(\beta\mu B)$$

In the high temperature limit, the entropy is $S = -Nk_B \ln[2]$, which corresponds to 2^N accessible states.

The two state system can be used to cool systems to very low temperatures. The technique is called **adiabatic demagnetization**. The idea is to cool a paramagnet with to low temperatures using conventional techniques, such as immersing it into liquid helium. Then one applies a magnetic field at constant temperature, and finally one pumps out the helium to thermally insulate the system and one switches the magnetic field off. The last step is adiabatic that is without removing heat from the system. Adiabatic means that the entropy remains constant. Thus in the constant temperature step, we move from the upper curve in Fig. 7.6 vertically down, and in the adiabatic step horizontally to the left to the original line. From the coordinates we read that the last process has changed the temperature of the paramagnet to a much lower value. Ideally this technique would allow to move all the way to zero temperature. In real systems however, there are residual fields in the material itself, which limit the temperatures that can be reached. The best material would be one with a high density of large magnetic moments that however do not couple. A popular material is for example $Ce_2Mg_3(NO_3)_{12}(H_2O)_{24}$[13]. For even lower temperatures $< 10^{-3}$ K one can use the nuclear magnetic moments in a metal. The nuclear moments are even better decoupled from each other than the magnetic ions in the salt mentioned before. A metal is used for better heat transport through the electron system. This technique allows temperatures of 10^{-7} K to be reached [14].

7.1.7 Concentration of point defects

Editor: There are errors in this section. Rework!

Crystals have a certain number of native defects, such as vacant lattice sites, so-called vacancies, or atoms that squeeze in between occupied lattice sites, so-called interstitials. There may also be a certain number of foreign atoms solved in the lattice. It is important to understand the concentration of these defects. In doped semiconductors the concentration of dopant atoms determines the number of charge carriers. Other atoms may influence the diffusivity of dopant atoms. In a structural material defects may hinder the motion of dislocations and therefore affect the plastic behavior of materials. The plastic behavior affects the crack propagation and is thus responsible for the stability of structural materials such as steel.

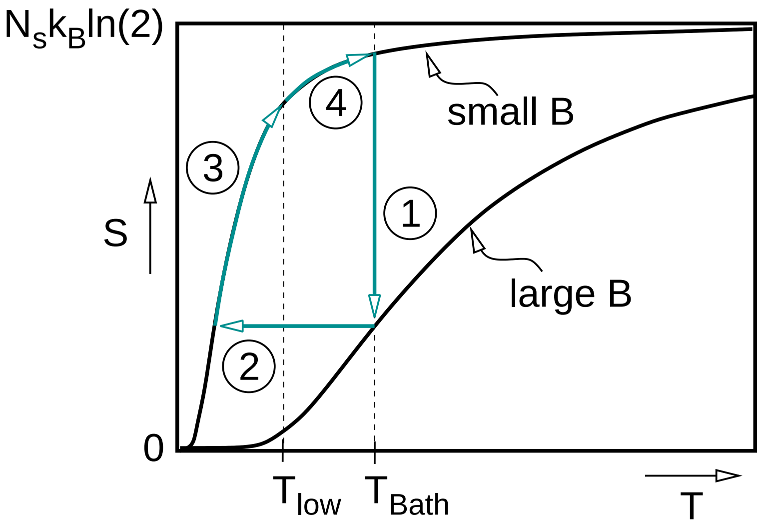


Fig. 7.6: Entropy as function of temperature for two values of the magnetic field. The field for the line denoted "large B" is twice that of the line denoted "small B". Shown is also the principle of adiabatic demagnetization. A system is cooled, for example by liquid Helium, to the temperature T_{bath}. (1) At this temperature the magnetic field is switched on, which reduces the entropy while the spins align. Then the helium is removed, in order to thermally insulate the system. (2) Now, the magnetic field is switched off. Because the system is adiabatic (thermally insulated), the entropy does not change during this process. The entropy drops to a very low value. Ideally the magnetic field should go to zero. In practice a small value due to intrinsic fields in the apparatus remains. (3) The system is brought into contact with the probe that is to be cooled. The probe may be at a temperature above T_{low}. Heat is transferred from the probe to the system. (4) If the probe and the coolant reach the same temperature, say at T_{low}, the coolant is decoupled from the probe, and brought into contact with helium.

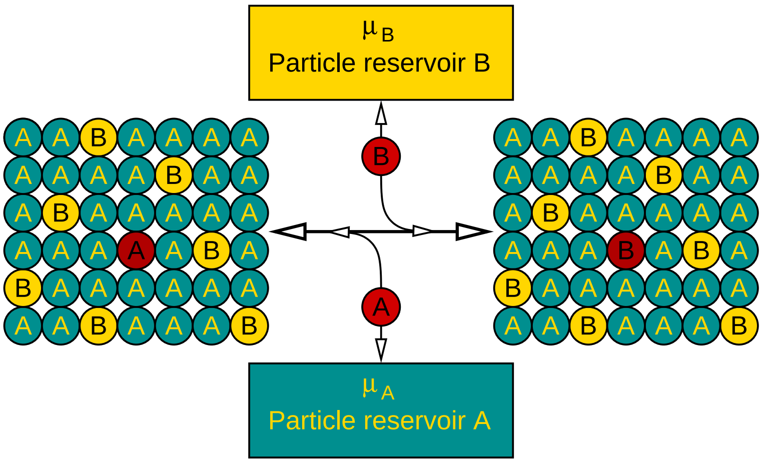


Fig. 7.7: Schematic diagram of an alloy, where one *A* atom is replaced by a B-type atom.

Let us consider a crystal made of atoms *A*. A small number of atoms *B* shall be dissolved in *A*. We can consider the insertion of defect atoms as a chemical reaction

$$A_{N_s} + B \leftrightarrow A_{N_s-1}B + A$$

where N_s is the number of lattice sites. The left hand side describes a material with N_s lattice sites and an atom *B* in the gas phase or another particle reservoir. The right-hand side describes the crystal with one substitutional[8] defect atom *B*.

The energy balance is

$$\Delta E = E[A_{N_s-1}B] + \mu_A - E[A_{N_s}] - \mu_B$$

Do not confuse μ_B with the Bohr magneton, which has the same symbol. Let us denote the energy per atom of the pure crystal with

$$E_x[A] \overset{\text{def}}{=} \frac{1}{N_s} E[A_{N_s}]$$

("x" stands for "X-tal" used for "Crystal".) The chemical potential of atom *A* is the lattice energy itself. Thus, in our model we obtain

$$\mu_A = \frac{dE[A_{N_A}]}{dN_A} \overset{N_s\to\infty}{=} E[A_{N_s+1}] - E[A_{N_s}] = (N_s+1)E_x[A] - N_s E_x[A] = E_x[A]$$

The formation energy of a defect *B* is defined as $E_F \overset{\text{def}}{=} E[A_{N_s-1}B] - E[A_{N_s}] + E_x[A]$, which can be obtained for example from the energy difference of two first-principles calculations.[9] Thus we obtain

$$\Delta E = E_F - \mu_B$$

The chemical potential of atom μ_B depends on the environment: The chemical potential differs, if the atoms come from the gas or from another solid or liquid material in contact with our probe.

Now we map this problem onto our two state model. We consider each lattice site as a two state system. It is in the ground state, if the lattice site is occupied by an atom *A* and in the excited state it is occupied with an atom *B*. Thus, the total energy is

$$E_{\vec\sigma} = \sum_{i=1}^{N_s} \sigma_i E_F$$

---

[8] A substitutional defect is one where a defect atom occupies a lattice site and thus displaced a host atom

[9] One calculates the energy of a lattice with N_s sites. Division by N_s yields $E_x[A] = E[A_{N_s}]/N_s$. Then one calculates the energy $E[A_{N_s-1}B]$ by replacing one atom *A* by an atom of type *B*.

If we compare this expression to the Eq. 7.1 on p. 162, we map the parameters fo this model onto those of the general two state model.

$$
\begin{aligned}
\epsilon_0 &= 0 \\
\Delta &= E_F + E_x[A] \\
\mu &= \mu_B
\end{aligned}
$$

The two state model will be reliable as long as the atoms B in the crystal A do not interact with each other. Thus the the concentration needs to be sufficiently low, so that the atoms B distributed over the lattice do not form pairs or larger aggregates. Secondly the energy of pairs or aggregates must be higher than that of an isolated B site, so that they do not form pairs spontaneously.

The number of particles can directly be obtained from $N(T, \mu)$ as

$$
\begin{aligned}
N(T, \mu_B) &= N_s \frac{1}{1 + \mathrm{e}^{\frac{\Delta - \mu}{k_B T}}} = N_s \frac{1}{1 + \mathrm{e}^{\frac{E_F + \mu_B}{k_B T}}} \\
c(T) &= \frac{N}{N_s} = \frac{1}{1 + \mathrm{e}^{\frac{E_F + \mu_B}{k_B T}}}
\end{aligned}
$$

which is the desired result.

Why does μ_B enter with a positive sign?? there is an error in the derivation!

The requirement that the concentration is low, corresponds to the condition $\mathrm{e}^{-\frac{E_F + \mu}{k_B T}} << 1$. If we expand in the small parameter, we obtain the Boltzmann factor

$$
c(T) \approx \mathrm{e}^{-\frac{E_F + \mu_B}{k_B T}}
$$

7.2 Rotations

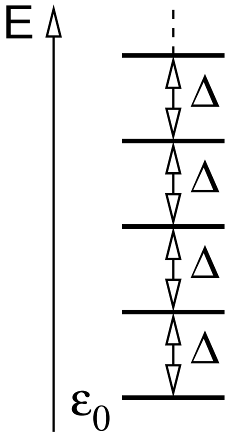
7.3 Harmonic oscillator

The harmonic oscillator is a model system for the vibrations of an atom in a molecule or a solid. The vibrational entropy is determines for example the concentration of defects, or the absolute value of a reaction rate.

Let us consider the harmonic oscillator, which has equidistant energy eigenvalues $E_i = \hbar\omega(i + \frac{1}{2})$ for states $i = 0, 1, 2, \ldots$. $\frac{1}{2}\hbar\omega$ is the zero-point vibrational energy and the spacing between the levels is $\hbar\omega$.

7.3.1 Thermodynamic potentials of the harmonic oscillator

The ground state energy $\frac{1}{2}\hbar\omega$ is denoted by ϵ_0 and the excitation energy $\hbar\omega$ is denoted by Δ.



Now we consider a system of N_s identical harmonic oscillators. The state of this system is uniquely defined by the N_s-dimensional vector $\vec{\sigma}$, which describes the excitation level of each harmonic oscillator. Thus, each element of $\vec{\sigma}$ can assume any non-negative integer value.

Thus the total energy of the system is

$$E_{\vec{\sigma}} = \sum_{i=1}^{N_s} (\epsilon_0 + \sigma_i \Delta) \qquad \text{with} \qquad \sigma_i \in \{0, 1, 2, \ldots\} \tag{7.12}$$

The number of excitation is defined as

$$N_{\vec{\sigma}} = \sum_{i=1}^{N_s} \sigma_i \tag{7.13}$$

Partition function and free energy

We construct the partition function

$$\begin{aligned} Z(T,\mu) &= \sum_{\vec{\sigma}} \mathrm{e}^{-\beta(E_{\vec{\sigma}} - \mu N_{\vec{\sigma}})} = \sum_{\vec{\sigma}} \prod_{i=1}^{N_s} \mathrm{e}^{-\beta(\epsilon_0 + \sigma_i \Delta - \mu\sigma_i)} \\ &= \prod_{i=1}^{N_s} \sum_{\sigma_i=0}^{\infty} \mathrm{e}^{-\beta(\epsilon_0 + \sigma_i(\Delta - \mu))} = \left[\mathrm{e}^{-\beta\epsilon_0} \sum_{\sigma=0}^{\infty} \left(\mathrm{e}^{-\beta(\Delta-\mu)} \right)^{\sigma} \right]^{N_s} \\ &= \left[\mathrm{e}^{-\beta\epsilon_0} \frac{1}{1 - \mathrm{e}^{-\beta(\Delta-\mu)}} \right]^{N_s} \end{aligned}$$

For the interchange of sum and product see the corresponding derivation for the two-state system. We used here the expression of a geometric series[10] Note that the result for $Z(T,\mu)$ is undefined, that is $Z(T,\mu) = \pm\infty$ for $\mu \geq \Delta$, because the geometric series is not defined. Thus the resulting expression is only valid if $\mu < \Delta$, despite the fact that we can evaluate the final expression also if the condition is violated.

---

[10]

$$(1-q)\sum_{i=0}^{n} q^i = \sum_{i=0}^{n} q^i - \sum_{i=1}^{n+1} q^i = 1 - q^{n+1} \qquad \Rightarrow \qquad \sum_{i=0}^{n} q^i = \frac{1-q^{n+1}}{1-q}$$

For an infinite geometric series $n \to \infty$, the term q^{n-1} vanishes, if $|q| < 1$. Otherwise the series diverges, and the result remains undefined.

The high temperature limit $\beta \rightarrow 0$ of the partition function for one site, $Z(T \rightarrow \infty, \mu) = \frac{k_B T}{\hbar\omega}$, we observe that it increases linearly with temperature. This is expected if the partition function is interpreted as the number of thermally accessible energy levels.

From the partition function we immediately obtain the free energy of the harmonic oscillator as

$$F(T, \mu) = -k_B T \ln[Z(T, \Delta, \mu)] = N_s \left\{ \epsilon_0 + k_B T \ln \left[1 - \mathrm{e}^{-\beta(\Delta-\mu)} \right] \right\} \tag{7.14}$$

Compare this expression to that for the two-state system Eq. 7.3 on p. 162 and explore the similarities and differences.

Number of excitations

Let us now determine the equations of state for the particle number N, which is the number of excitations

$$\begin{aligned} N(T, \mu) &= -\frac{\partial F}{\partial \mu} = -N_s k_B T \frac{1}{1 - \mathrm{e}^{-\beta(\Delta-\mu)}} \left(-\frac{1}{k_B T} \right) \mathrm{e}^{-\beta(\Delta-\mu)} \\ &= N_s \frac{1}{\mathrm{e}^{+\beta(\Delta-\mu)} - 1} \end{aligned}$$

The average excitation level $b = N/N_s$ of the harmonic oscillator is the **Bose distribution function**, which will play an important role in the statistics of Bosons

$$b(\Delta, T, \mu) = \frac{N(T, \mu)}{N_s} = \frac{1}{\mathrm{e}^{+\beta(\Delta-\mu)} - 1} \tag{7.15}$$

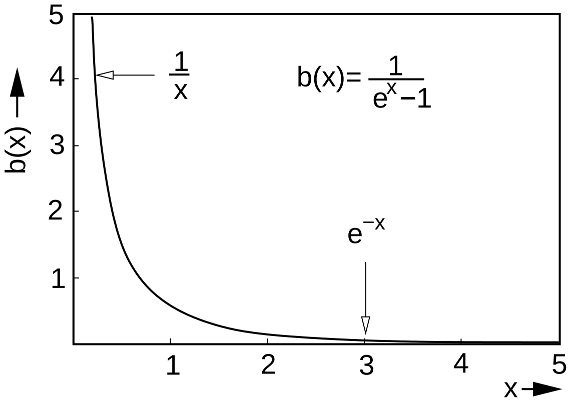


Fig. 7.8: Bose distribution function as function of $x = (\epsilon - \mu)/k_B T$. There is a singularity at $x = 0$. The distribution is negative for $x < 0$. For large arguments, i.e. $x >> 1$ the Bose distribution becomes identical to the Boltzmann factor.

The Bose distribution function as function of Δ has a singularity at the chemical potential.[11]

$$b(\Delta, T, \mu) = \frac{1}{\mathrm{e}^{\beta(\Delta-\mu)} - 1} \rightarrow \left\{ \begin{array}{lll} \frac{k_B T}{\Delta-\mu} & \text{for} & |\Delta - \mu| << k_B T \\ \mathrm{e}^{-\beta(\Delta-\mu)} & \text{for} & \Delta - \mu >> k_B T \\ -1 & \text{for} & \Delta - \mu << -k_B T \end{array} \right\}$$

---

[11] In the case $|\Delta - \mu| >> k_B T$, $\alpha \overset{\text{def}}{=} \mathrm{e}^{-\beta(\Delta-\mu)}$ is a small quantity.

$$\frac{1}{\alpha^{-1} - 1} = \frac{\alpha}{1 - \alpha} = \alpha(1 + \alpha + O(\alpha^2) = \alpha + O(\alpha^2$$

Interestingly we obtain the Boltzmann factor for $\Delta-\mu >> k_BT$, which is the same result as for the two state system. This observation can be rationalized as follows: In this limit the number of excitations is small. If we identify the number of excitations with a number of particles we would have a dilute system. The number of double and higher excitations can be neglected compared to the first excitation. As a consequence the second and higher terms, i.e. $\sigma \geq 2$, in the partition sum can be neglected relative to the first two terms. The first two terms correspond exactly to the two state system. This shows that nearly all systems[12] behave like the two-state system, if the number of excitations is small. This is the case for $\Delta > \mu$ in the low-temperature limit.

Values with $\Delta \leq \mu$ are not meaningful for bosons, as the number of excitations is negative. The singularity at $\Delta = \mu$ will have important consequences as it will lead to the Bose-Einstein condensation.

Entropy

Let us now determine the entropy from the equations of state

$$\begin{aligned} S(T,\mu) &= -\frac{\partial F}{\partial T} \\ &\overset{\text{Eq. 7.14}}{=} -N_sk_B \ln\left[1-\mathrm{e}^{-\beta(\Delta-\mu)}\right] - N_sk_BT\frac{1}{1-\mathrm{e}^{-\beta(\Delta-\mu)}}\left(\frac{\Delta-\mu}{k_BT^2}\right)\mathrm{e}^{-\beta(\Delta-\mu)} \\ &= -N_sk_B\left\{\ln\left[1-\mathrm{e}^{-\beta(\Delta-\mu)}\right] - \frac{\Delta-\mu}{k_BT}\frac{1}{\mathrm{e}^{+\beta(\Delta-\mu)}-1}\right\} \end{aligned}$$

It is sometimes useful to express the entropy directly by the number of excitations $N(T,\mu)/N_s$, which is the Bose distribution function.

$$\begin{aligned} b(x) = \frac{1}{\mathrm{e}^x-1} \qquad &\Rightarrow \qquad \frac{1}{b} = \mathrm{e}^x - 1 \qquad \Rightarrow \qquad \mathrm{e}^x = \frac{1}{b}+1 = \frac{1+b}{b} \\ 1-\mathrm{e}^{-x} &= 1-\frac{b}{1+b} = \frac{1}{1+b} \\ x &= \ln(\mathrm{e}^x) = \ln\left(\frac{1+b}{b}\right) = \ln(1+b)-\ln(b) \\ \frac{1}{\mathrm{e}^x-1} &= b \end{aligned}$$

Thus the entropy obtains the form

$$\begin{aligned} S(b) &= -N_sk_B\left\{\ln\left[1-\mathrm{e}^{-\beta(\Delta-\mu)}\right] - \frac{\Delta-\mu}{k_BT}\frac{1}{\mathrm{e}^{+\beta(\Delta-\mu)}-1}\right\} \\ &= N_sk_B\left\{-\ln(1+b) - [\ln(1+b)-\ln(b)]\,b\right\} \\ &= -N_sk_B\Big\{b\ln(b) - (1+b)\ln(1+b)\Big\} \end{aligned}$$

Thus we can express the entropy as function of the excited states $N = N_sb$.

ENTROPY OF A HARMONIC OSCILLATOR

$$S(b) = -N_sk_B\Big[b\ln(b) - (1+b)\ln(1+b)\Big] \qquad (7.16)$$

Compare this expression to the analogous expression for the two-state system Eq. 7.5 on p. 164.

Note the close similarity to the entropy of the two state system. The only differences are some sign changes.

---

[12] The requirement is that the first excited level is not degenerate or nearly degenerate

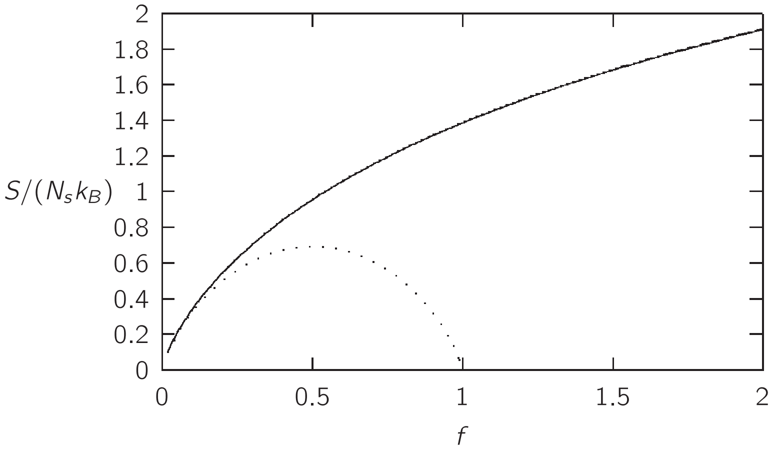


Fig. 7.9: Entropy of a harmonic oscillator and the two-state system as function of the number of excitations. The full line is the entropy of the harmonic oscillator and the dotted line, leading back to zero for the occupation $f = 1$ corresponds to the two-state system. The two entropies become identical for $f << 1$. This is the limit of a dilute gas, where the Fermi and Bose statistics become identical to the Boltzmann statistics.

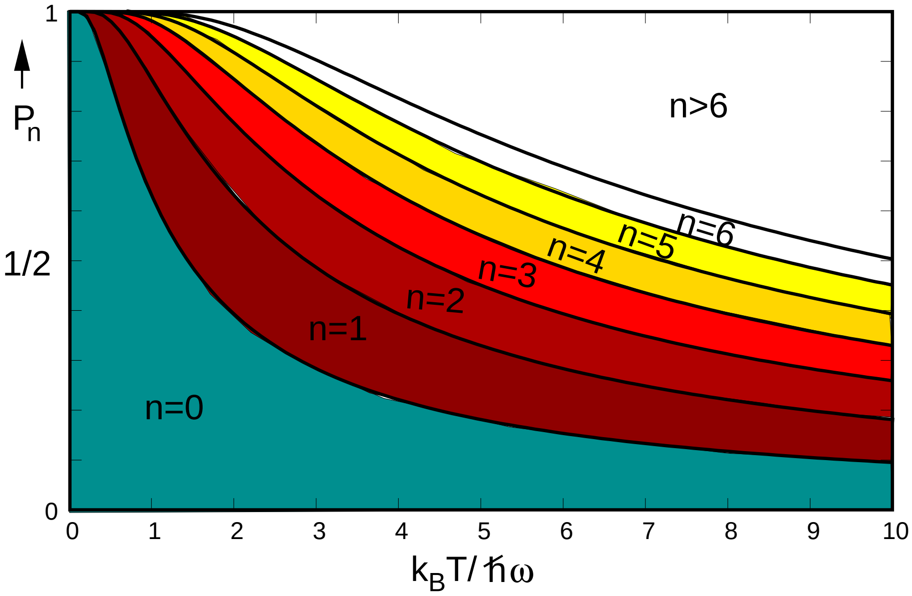


Fig. 7.10: The probability $P_n(T)$ for the occupation of the six lowest states of the harmonic oscillator as function of temperature. Note that the probabilities are cummulative, so that each state of the harmonic oscillator is represented by the region between two lines. The ground state appears with the largest probability for all temperatures. At low temperature, i.e. $k_BT < \hbar\omega$, the harmonic oscillator can be approximated well by a system with only two or three states.

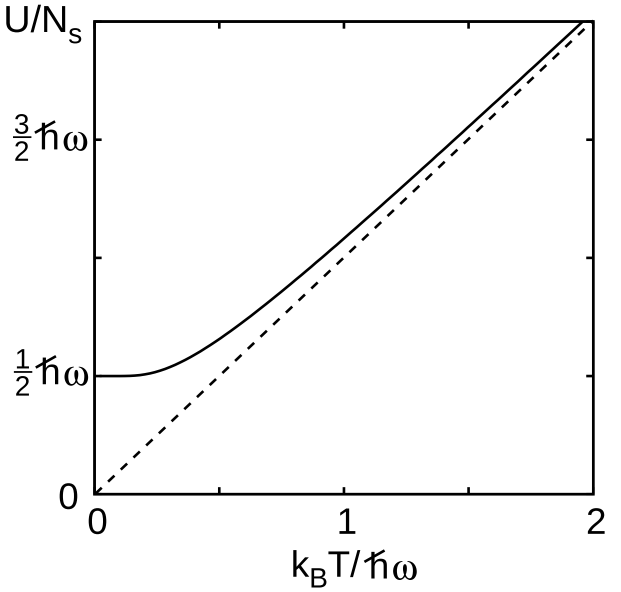


Fig. 7.11: The internal energy $U(T)=\frac{\hbar\omega}{2}\frac{cosh(\hbar\omega/(2k_BT))}{\sinh(\hbar\omega/(2k_BT))}$ as derived from Eq. 7.17 with $\epsilon_0=\frac{1}{2}\hbar\omega$ and $\Delta=\hbar\omega$. At low temperatures the internal energy approaches the zero-point energy $U(T\to 0)=\frac{1}{2}\hbar\omega$, while the high temperature result approaches $U(T\to\infty)=k_BT$, which is identical to the classical limit.

Internal energy

The internal energy can be expressed simply by by counting the number of excited states

$$U=N_s\epsilon_0+N\Delta=N_s\left(\epsilon_0+b\Delta\right)$$

However, let us also consider the standard method starting from the free energy $F(T,\mu)$, which in this case is the grand canonical potential.

$$\begin{aligned}
U(T,\mu)&\overset{\text{Eq. 1.67}}{=}F+TS+\mu N\\
&=\underbrace{\left(N_s\epsilon_0+N_sk_BT\ln\left[1-\mathrm{e}^{-\beta(\Delta-\mu)}\right]\right)}_{F}\\
&\quad\underbrace{-N_sk_BT\left\{\ln\left[1-\mathrm{e}^{-\beta(\Delta-\mu)}\right]-\frac{\Delta-\mu}{k_BT}\frac{1}{\mathrm{e}^{\beta(\Delta-\mu)}-1}\right\}}_{TS}\\
&\quad\underbrace{+\mu N_s\frac{1}{\mathrm{e}^{\beta(\Delta-\mu)}-1}}_{\mu N}\\
&=N_S\epsilon_0+N_s\Delta\underbrace{\frac{1}{\mathrm{e}^{\beta(\Delta-\mu)}-1}}_{b}
\end{aligned}\qquad(7.17)$$

which is identical to what we derived above as expectation value of the energy.

7.3.2 Vibrational energy and entropy

We have seen earlier the free energy of a general ideal gas can be computed from the translational free energy, which corresponds to that of the ideal Boltzmann gas, a rotational and a vibrational contribution. Here we will consider the vibrational contribution, which is identical to that of the harmonic oscillator.

We can approximate the total energy surface of a molecule by a Taylor series about the minimum

$$E(\vec{R}) = E^0 + \frac{1}{2}\sum_{i,j}\frac{\partial^2 E}{\partial R_i \partial R_j}\left(R_i - R_i^0\right)\left(R_j - R_j^0\right) + O\left(|\vec{R} - \vec{R}^0|\right)^3$$

Now we build the dynamical matrix

$$D_{i,j} = \frac{1}{\sqrt{m_i}}\frac{\partial^2 E}{\partial R_i \partial R_j}\frac{1}{\sqrt{m_j}}$$

The eigenvalues d_j of the dynamical matrix provide us with the frequencies $\omega_i = \sqrt{d}_i$. There are up to six eigenvectors with an eigenvalue equal to zero, which just correspond to the translations and rotations. Those are eliminated, since these degrees of freedom are treated differently.

The free energy of the vibrations can simply be composed from the independent harmonic oscillators

$$F_{vib}(T) \overset{Eq.\ 7.14}{=} \sum_{j=1}^{3N-6}\left\{\frac{1}{2}\hbar\omega_j + k_B T \ln[1 - \mathrm{e}^{-\frac{\hbar\omega_j}{k_B T}}]\right\} \tag{7.18}$$

High-Temperature limit/Classical limit

In the high-temperature limit, that is $k_BT >> \hbar\omega$, which is identical with the classical limit$\hbar \rightarrow 0$, we obtain with the help of $\mathrm{e}^x = 1 + x + \frac{1}{2}x^2 + O(x^3)$ and $\ln[1 + x] = x + O(x^2)$,

$$
\begin{aligned}
F_{vib}(T) &\overset{Eq.\ 7.18}{=} \sum_{j=1}^{3N-6} \left\{ \frac{1}{2}\hbar\omega_j + k_BT \ln\left[1 - \mathrm{e}^{-\frac{\hbar\omega_j}{k_BT}}\right] \right\} \\
&\overset{\text{Taylor } \mathrm{e}^x}{=} \sum_{j=1}^{3N-6} \left\{ \frac{1}{2}\hbar\omega_j + k_BT \ln\left[1 - 1 + \frac{\hbar\omega_j}{k_BT} - \frac{1}{2}\left(\frac{\hbar\omega_j}{k_BT}\right)^2 + O\left(\frac{\hbar\omega_j}{k_BT}\right)^3\right] \right\} \\
&= \sum_{j=1}^{3N-6} \left\{ \frac{1}{2}\hbar\omega_j + k_BT \ln\left[\frac{\hbar\omega_j}{k_BT}\left(1 - \frac{1}{2}\left(\frac{\hbar\omega_j}{k_BT}\right) + O\left(\frac{\hbar\omega_j}{k_BT}\right)^2\right)\right] \right\} \\
&= \sum_{j=1}^{3N-6} \left\{ \frac{1}{2}\hbar\omega_j + k_BT \ln\left[\frac{\hbar\omega_j}{k_BT}\right] + k_BT \ln\left[1 - \frac{1}{2}\left(\frac{\hbar\omega_j}{k_BT}\right) + O\left(\frac{\hbar\omega_j}{k_BT}\right)^2\right] \right\} \\
&\overset{Taylor\ \ln}{=} \sum_{j=1}^{3N-6} \left\{ \frac{1}{2}\hbar\omega_j + k_BT \ln\left[\frac{\hbar\omega_j}{k_BT}\right] - k_BT\frac{1}{2}\frac{\hbar\omega_j}{k_BT} + O\left(\frac{\hbar\omega_j}{k_BT}\right)^2 \right\} \\
&= \sum_{j=1}^{3N-6} \left\{ k_BT \ln\left[\frac{\hbar\omega_j}{k_BT}\right] + O\left(\frac{\hbar\omega_j}{k_BT}\right)^2 \right\} \\
&\overset{\ln(a)+\ln(b)=\ln(ab)}{=} k_BT \ln\left[\prod_{j=1}^{3N-6} \frac{\hbar\omega_j}{k_BT}\right] + \sum_{j=1}^{3N-6} O\left(\frac{\hbar\omega_j}{k_BT}\right)^2 \\
&= k_BT \ln\left[\sqrt{\prod_{j=1}^{3N-6} \left(\frac{\hbar}{k_BT}\right)^2 \omega_j^2}\right] + \sum_{j=1}^{3N-6} O\left(\frac{\hbar\omega_j}{k_BT}\right)^2 \\
&= k_BT \ln\left[\left(\frac{\hbar}{k_BT}\right)^{3N-6} \sqrt{\det \mathbf{D}}\right] + \sum_{j=1}^{3N-6} O\left(\frac{\hbar\omega_j}{k_BT}\right)^2
\end{aligned}
\tag{7.19}
$$

Note however, that **D** is the dynamical matrix with the translations and rotations projected out. If translations and rotations are treated as vibrations with zero frequency, the argument of the logarithm vanishes because also the corresponding frequency vanishes, and, as a result, the logarithm, diverges. The proper treatment is to describe the translations as a Boltzmann gas[13], and the rotations should be described as quantum mechanical rotator.

Eq. 7.19 is the classical expression, i.e. for $\hbar \rightarrow 0$ for the vibrational free energy. This expression only holds, when the temperature is sufficiently high, that is $k_BT >> \hbar\omega_j$.

Note that Eq. 7.19 for the free energy already contains the zero-point vibrational energy. A common mistake is the attempt to improve the results by including the zero-point vibrational energies in the ground state energy and add the classical expression for the free energy ontop. As a result the zero-point vibrations are counted twice. This mistake has its origin in an first-order expansion of the exponential in the above derivation, which contains the zero point energy, which cancels exactly with a term from the second-order.

The classical expression for the free energy has the advantage that the result only depends on the determinant of the dynamical matrix and not on the individual eigenvalues. The price is that the approximate expression is limited to sufficiently high temperatures.

---

[13] At very low temperatures and sufficiently high densities, quantum effects may become important and the Boltzmann gas needs to be replaced by a Bose or Fermi gas.

CLASSICAL ENTROPY OF THE HARMONIC OSCILLATOR

The classical entropy of the harmonic oscillator is

$$S=-\frac{dF}{dT}=(3N-6)k_B-k_B\ln\left[\left(\frac{\hbar}{k_BT}\right)^{3N-6}\sqrt{\det[\bar{D}]}\right]+O(\hbar^2) \tag{7.20}$$

Finally, we obtain the internal energy as function of temperature as

$$U(T)=F+TS=(3N-6)k_BT$$

Low-Temperature limit

It may be instructive to recheck whether **Nernst's theorem**, the 3rd law of thermodynamics, is valid for the harmonic oscillator. Nernst's theorem claims that the entropy per system size vanishes at absolute zero temperature.

In the low temperature limit we obtain from the quantum mechanical expression of the free energy of a system of harmonic oscillators, with the help of the expansion $\ln(x)=1+x+O(x^2)$,

$$F_{vib}(T\rightarrow 0)\overset{Eq.~7.18}{=}\sum_{j=1}^{3N-6}\frac{1}{2}\hbar\omega_j-k_BT\sum_{j=1}^{3N-6}\mathrm{e}^{-\frac{\hbar\omega_j}{k_BT}}+O\Big(\Big(\mathrm{e}^{-\frac{\hbar\omega_j}{k_BT}}\Big)^2\Big)$$

This expansion is valid if thermal energies are smaller than the smallest excitation energy, i.e. $k_BT<<\min_j\Big[\hbar\omega_j\Big]$.

We can verify Nernst's theorem by calculating the entropy

$$\begin{aligned}S_{vib}(T\rightarrow 0)&=-k_B\sum_{j=1}^{3N-6}\mathrm{e}^{-\frac{\hbar\omega_j}{k_BT}}-k_BT\sum_{j=1}^{3N-6}\frac{\hbar\omega_j}{k_BT^2}\mathrm{e}^{-\frac{\hbar\omega_j}{k_BT}}\\&=-k_B\sum_{j=1}^{3N-6}(1+\frac{\hbar\omega_j}{k_BT^2})\mathrm{e}^{-\frac{\hbar\omega_j}{k_BT}}\overset{T\rightarrow 0}{\rightarrow}0\end{aligned}$$

Thus we have verified Nernst's theorem for the quantum mechanical expression. Note however, that the classical expression Eq. 7.20 violates Nernst's theorem.

7.4 Summary

Important formulas

An N-dimensional harmonic oscillator has the following internal energy and entropy in the classical limit.

$$\begin{aligned}U^{vib}&=3Nk_BT\\S^{vib,cl}&=k_BT\ln\left[\left(\frac{\hbar}{k_BT}\right)^{3N}\sqrt{\det\mathbf{D}}\right]\\f(\Delta,T,\mu)&=\frac{1}{1+\mathrm{e}^{\beta(\Delta-\mu)}}\\b(\Delta,T,\mu)&=\frac{1}{\mathrm{e}^{\beta(\Delta-\mu)}-1}\end{aligned}$$

Chapter 8

Non-interacting particles

In this section we will describe how our formalism can be applied to many identical particles without interaction. For us it will lead us to terms like the density of states and the Fermi distribution function and the Bose distribution function. Thus we make contact with the terms central to the material in most textbooks about solid state physics. What is explained here is furthermore one possible starting point for quantum field theory.

8.1 Fock space and number representation

For a system with non-interacting particles, the many-particle Hamiltonian can be written as a sum of one-particle Hamiltonians.

$$H_{tot}(\vec{r}_1, \dots, \vec{r}_n, \vec{p}_1, \dots, \vec{p}_n) = \sum_i H_i(\vec{r}_i, \vec{p}_i)$$

where $\vec{r}_i$ and $\vec{p}_i$ are the position and momentum operators acting on the coordinates of the i-th particle.

The eigenstates of the Hamiltonian are composed of product wave functions of the one-particle states. For Fermions we obtain an eigenstate from the anti-symmetrized product of one-particle states, that is a Slater determinant, and for Bosons we take the symmetrized product. For independent particles, the number of particles in a one-particle state is conserved and therefore a good quantum number. Hence, an eigenstate of a system of independent particles can be completely characterized by the number of particles in each one-particle state. This is the so-called **number representation**[1] of many-particle states. An eigenstate can be characterized as $|\Psi\rangle = |n_1, n_2 \dots\rangle$, where every position belongs to a well-defined one-particle state and the numbers n_i denote the number of particles in that state. Due to the Pauli principle the numbers n_i can only be 0 or 1 for Fermions, while they can be any non-negative integer for Bosons. (It is therefore not allowed to drop an unoccupied state from the list.) The Hilbert space is now extended to the **Fock space**. The Fock space is the product of the Hilbert spaces for all possible particle numbers.

For the number representation we first need to choose a complete (orthonormal) basisset of one-particle wave functions $\{|\phi_i\rangle\}$. From the one-particle basis we construct a basisset for many-particle wave functions: A state $|n_1, n_2, \dots\rangle$ corresponds to the symmetrized or anti-symmetrized product wave function built out of n_1 times $\phi_1(r)$, n_2 times $\phi_2(r)$ and so on. This basis contains

1. one-particle states functions

$$|0, \dots, 0, \underbrace{1}_{pos.i}, 0, \dots, \dots\rangle = |\phi_i\rangle$$

---

[1] The German word for number representation is Besetzungszahldarstellung

and

2. two-particle states

$$|0,\dots,0,\underbrace{1}_{pos.i},0,\dots,0,\underbrace{1}_{pos.j},0,\dots\rangle=|\Phi_{i,j}\rangle$$

where the corresponding wave function is defined as

$$\Phi_{i,j}(r,r')=\frac{1}{\sqrt{2}}\left(\phi_i(r)\phi_j(r')\pm\phi_j(r)\phi_i(r')\right)$$

The plus sign in the wave function holds for Bosons and the minus sign for Fermions. For Bosons we can also construct a two particle wave function from a single one-particle orbital

$$|0,\dots,0,\underbrace{2}_{pos.i},0,\dots\rangle=|\Phi_{i,i}\rangle$$

where

$$\Phi_{i,i}(r,r')=\phi_i(r)\phi_i(r')$$

Such a wave function does not exist for Fermions due to the Pauli principle: After anti-symmetrization the wave function would be identical to zero.

3. in a similar fashion we construct many-particle states for any particle number, which consist of symmetrized or anti-symmetrized products of the corresponding one-particle states.

$$|n_1,n_2,n_3\dots\rangle=|\Phi_{\underbrace{1,1,\dots}_{n_1 times},\underbrace{2,2,\dots}_{n_2 times},\underbrace{3,3,\dots}_{n_3 times},\dots}\rangle$$

Note that we need to determine a sign convention. Consider a two particle wave function for Fermions according to the above definition: We obtain $|\Phi_{i,j}\rangle=-|\Phi_{j,i}\rangle$, which implies that the sign of a state in the number representation is not unique. In order to make it unique we need to impose a convention: In the symmetrized or anti-symmetrized product, the term in which all one-particle orbitals appear in the order of their quantum numbers must be positive. For our fermionic two-particle state, this implies that the state in the number representation equals $|\Phi_{i,j}\rangle$ when $i<j$ and it equals $-|\Phi_{i,j}\rangle$ when $j<i$. [2]

It is important to note that the scalar product between states with different particle numbers vanishes. Since they are eigenstates of a hermitian operator, the particle number operator,to different eigenvalues, their scalar product must vanish.

A general many-particle state is represented in this basis as

$$|\Psi\rangle=\sum_{n_1,n_2,\dots}|n_1,n_2,\dots\rangle c_{n_1,n_2,\dots}$$

Note that we can now write a state, which is a superposition of contributions with different particle numbers. Such a state can no more be represented by a wave function in real space.

Similarly the general form of the state operator is

$$\hat\rho=\sum_{n_1,n_2,\dots,m_1,m_2,\dots}|n_1,n_2,\dots\rangle\rho_{n_1,n_2,\dots,m_1,m_2,\dots}\langle m_1,m_2,\dots|$$

The Hamilton operator

$$\hat H=\sum_{n_1,n_2,\dots,m_1,m_2,\dots}|n_1,n_2,\dots\rangle\langle n_1,n_2,\dots|\hat H|m_1,m_2,\dots\rangle\langle m_1,m_2,\dots|$$

---

[2] Note that we need to agree beforehand on the order of the quantum numbers of the one-particle states.

usually does not couple states with different particle numbers[3] Thus the matrix elements between basisfunctions with different number of particles vanish. Thus the matrix elements are the usual matrix elements of N-particle states.

A particular simple representation is obtained for non-interacting particles, where the Hamilton operator is simply the sum of single particle Hamilton operators: Each particle experiences the same Hamiltonian. In this case we build up the number representation from a one-particle basis of eigenstates of the Hamiltonian.

In this representation, the energy of such a many-particle state of non-interacting particles is just the sum of the one-particle energies defined by $H_i|\phi\rangle = |\phi\rangle\epsilon$, that is

$$E|n_1, n_2, \dots\rangle = \Big(\sum_i n_i\epsilon_i\Big)|n_1, n_2, \dots\rangle$$

and the total number of electrons is

$$N|n_1, n_2, \dots\rangle = \Big(\sum_i n_i\Big)|n_1, n_2, \dots\rangle.$$

The index i runs over all one-particle states.

8.2 Excursion to creation and annihilation operators

We may now define a **creation operator** in the following way

$$\hat{c}_i^\dagger = \sum_{n_1,\dots} |\dots, n_i + 1, \dots\rangle\sqrt{n_i + 1}\langle\dots, n_i, \dots|$$

If we apply the creation operator to an N-particle wave function we obtain an $N+1$ particle wave function. The creation operator simply creates one additional particle in the i-th one-particle state. The creation operator here is completely analogous to the creation operator we know from the harmonic oscillator. The only difference is that, in the harmonic oscillator, we create excitations, whereas here we create particles. In second quantization a particle is treated just like any other excitation.

With the help of creation operators, we can construct the entire many-particle basis from a single state, the so-called **vacuum state** $|0\rangle \stackrel{\text{def}}{=} |0, 0, \dots\rangle$. Note that the vacuum state is not zero. It is still normalized to one $\langle 0|0\rangle = 1$. The many-particle basis functions are, expressed by the vacuum state and the creation operators,

$$|n_1, n_2, \dots\rangle = \frac{1}{\sqrt{\prod_i n_i!}}\left(\hat{c}_1^\dagger\right)^{n_1}\left(\hat{c}_2^\dagger\right)^{n_2}\dots|0\rangle$$

The factors had to be included so that the states remain normalized. The proof is left to the reader. Thus we do not need many-particle basis functions any longer. We can describe everything with creation operators and the vacuum state. This leads to a new paradigm: Since the vacuum state is always the same, we do not need to write it down. Instead of states we now deal only with operators.

The hermitian conjugate partner of the creation operator is the **annihilation operator**:

$$\hat{c}_i = \sum_{n_1,\dots} |\dots, n_i - 1, \dots\rangle\sqrt{n_i}\langle\dots, n_i, \dots|$$

The annihilation operator deletes one particle in the i-th one-particle state. The reader may verify that the annihilation operator is indeed the hermitian conjugate of the creation operator. Note also that the factor $\sqrt{n_i}$ ensures that all states have a positive number of particles.

---

[3] Note however, that Hamilton operators exist that do not preserve particle numbers. One example are those describing electron-phonon coupling.

The creation and annihilation operators can be used to construct the **number operator**

$$\hat{n}_i \stackrel{\text{def}}{=} \hat{c}_i^\dagger \hat{c}_i = \sum_{n_1,n_2,\ldots} |\ldots, n_i, \ldots\rangle n_i \langle \ldots, n_i, \ldots|$$

The expectation value of the number operator provides us with the number of particles in a one-particle state.

The number operator is useful because we can write the Hamiltonian for independent particles in a simple fashion. The Hamiltonian for independent particles is the sum of one-particle Hamiltonians acting on the coordinates of the individual particles. We choose the Eigenstates of this one-particle Hamiltonian to set up our number-representation basis for the Fock space. This basisset automatically consists of eigenstates of the many-particle Hamiltonian. Next we define the creation and annihilation operators in this basisset. If the eigenstates of the one-particle hamiltonian are ϵ_i, the many-particle Hamiltonian has the form

$$\hat{H} = \sum_i \epsilon_i \hat{c}_i^\dagger \hat{c}_i$$

This equation results only from the fact that the Hamiltonian is diagonal in a representation of eigenstates of the corresponding one-particle Hamiltonian and that the eigenvalue is given by $E_{n_1,n_2,\ldots} = \sum_i \epsilon_i n_i$.

Similarly we obtain the particle number operator as

$$\hat{N} = \sum_i \hat{c}_i^\dagger \hat{c}_i$$

Without proof I will also show here the form for the Hamiltonian for interacting particles, having an interaction $\hat{W}$

$$\hat{H} = \sum_i \epsilon_i \hat{c}_i^\dagger \hat{c}_i + \frac{1}{2} \sum_{i,j,k,l} W_{i,j,k,l} \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_k \hat{c}_l$$

where $W_{i,j,k,l} = \int d^3r \int d^3r' \; \Phi^*_{i,j}(\vec{r}, \vec{r}') W(\vec{r}, \vec{r}') \Phi_{l,k}(\vec{r}, \vec{r}')$. For an electrostatic interaction between charged particles the interaction has the form $W(\vec{r}, \vec{r}') = \frac{q^2}{4\pi\epsilon_0} \frac{1}{|\vec{r} - \vec{r}'|}$. The interaction term has a simple interpretation. Two particles in the states k and l scatter and leave the collision in the states i and j. Therefore the particles in the states k and l are deleted and two particles in the states i and j are created.

So far I have not discussed the commutator relations, which have been important in the algebraic description of the harmonic oscillator (see ΦSX: Quantum theory). Let us consider the commutator between creation and annihilation operator. We calculate its matrix elements: For identical indices we obtain

$$\begin{aligned}
&\langle \ldots, n_i, \ldots | [\hat{c}_i, \hat{c}_i^\dagger] | \ldots, m_i, \ldots \rangle \\
&= \langle \ldots, n_i, | \hat{c}_i \hat{c}_i^\dagger | \ldots, m_i, \ldots \rangle - \langle \ldots, n_i, \ldots | \hat{c}_i^\dagger \hat{c}_i | \ldots, m_i, \ldots \rangle \\
&= \langle \ldots, n_i, \ldots | \hat{c}_i | \ldots, m_i + 1, \ldots \rangle \sqrt{m_i + 1} - \langle \ldots, n_i, \ldots | \hat{c}_j^\dagger | \ldots, m_i - 1, \ldots \rangle \sqrt{m_i} \\
&= \langle \ldots, n_i, \ldots | \ldots, m_i, \ldots \rangle (m_i + 1) - \langle \ldots, n_i, \ldots | \ldots, m_i, \ldots \rangle m_i \\
&= \delta_{n_i, m_i}
\end{aligned}$$

Similarly we can show that the commutator for different indices vanishes, and we can show that the commutators between two creation operators and two annihilation operators vanish.

Thus we obtain the **commutator relations of creation and annihilation operators.**

$$\left[\hat{c}_i, \hat{c}_j^\dagger\right] = \delta_{i,j} \qquad [\hat{c}_i, \hat{c}_j] = 0 \qquad \left[\hat{c}_i^\dagger, \hat{c}_j^\dagger\right] = 0$$

Extension for Fermions

This formulation has a flaw, because it allows us to create more than one fermion in a single one-particle state. This is not possible due to the requirement that the wave function is antisymmetric with respect to interchange of the particle coordinates. Therefore we have to modify the formulation that we obtain zero if the creation operator acts on a state which has already a particle in that state. Thus we need to impose the condition

$$\left(\hat{c}_i^\dagger\right)^2=0$$

We could now define the creation operator as follows

$$c_i^\dagger=\sum_{n_1,n_2,\ldots}|\ldots,n_i+1,\ldots\rangle\delta_{n_i,0}\langle\ldots,n_i,\ldots|$$

This expression is equivalent to the creation operator for bosons, but it has different commutator relations for $[c_i,c_i^\dagger]$, which can be illustrated for the example of a system with only one one-particle orbital.

$$\begin{aligned}
\begin{pmatrix}\langle0|c^\dagger|0\rangle & \langle0|c^\dagger|1\rangle\\ \langle1|c^\dagger|0\rangle & \langle1|c^\dagger|1\rangle\end{pmatrix}&=\begin{pmatrix}0&0\\1&0\end{pmatrix}\\
\Rightarrow\begin{pmatrix}\langle0|c|0\rangle & \langle0|c|1\rangle\\ \langle1|c|0\rangle & \langle1|c|1\rangle\end{pmatrix}&=\begin{pmatrix}0&1\\0&0\end{pmatrix}\\
\Rightarrow\begin{pmatrix}\langle0|[c,c^\dagger]_-|0\rangle & \langle0|[c,c^\dagger]_-|1\rangle\\ \langle1|[c,c^\dagger]_-|0\rangle & \langle1|[c,c^\dagger]_-|1\rangle\end{pmatrix}&=\begin{pmatrix}1&0\\0&-1\end{pmatrix}\neq\begin{pmatrix}1&0\\0&1\end{pmatrix}
\end{aligned}$$

However if we replace the commutator with the anti-commutator we obtain a relation similar to the result for bosons.

An **anti-commutator** of two operators $\hat{A}$ and $\hat{B}$ is defined as

$$\left[\hat{A},\hat{B}\right]_+\stackrel{\text{def}}{=}\hat{A}\hat{B}+\hat{B}\hat{A}$$

If we want to make the distinction between commutators and anti-commutators explicit, we usually give the commutator a subscript “-”.

We can see that the anti-commutator relations for the fermionic creation and annihilation operators are analogous to the commutator relation for the bosonic creation and annihilation operators:

$$\begin{aligned}
\begin{pmatrix}\langle0|[c,c^\dagger]_+|0\rangle & \langle0|[c,c^\dagger]_+|1\rangle\\ \langle1|[c,c^\dagger]_+|0\rangle & \langle1|[c,c^\dagger]_+|1\rangle\end{pmatrix}&=\begin{pmatrix}1&0\\0&1\end{pmatrix}\\
\begin{pmatrix}\langle0|[c^\dagger,c^\dagger]_+|0\rangle & \langle0|[c^\dagger,c^\dagger]_+|1\rangle\\ \langle1|[c^\dagger,c^\dagger]_+|0\rangle & \langle1|[c^\dagger,c^\dagger]_+|1\rangle\end{pmatrix}&=\begin{pmatrix}0&0\\0&0\end{pmatrix}\\
\begin{pmatrix}\langle0|[c,c]_+|0\rangle & \langle0|[c,c]_+|1\rangle\\ \langle1|[c,c]_+|0\rangle & \langle1|[c,c]_+|1\rangle\end{pmatrix}&=\begin{pmatrix}0&0\\0&0\end{pmatrix}
\end{aligned}$$

An elegant solution for the problem is to redefine the commutator relations for the creation operators. If the creation operators do not commute, we need to make expression which links the number representation states to the creation operators dependent on the order of the creation operators.

===================

Furthermore a wave function $\hat{c}_i^\dagger\hat{c}_j^\dagger|0\rangle$ has a different sign than $\hat{c}_j^\dagger\hat{c}_i^\dagger|0\rangle$, because the two wave functions differ by an interchange of indices. This is valid also for states other than the vacuum state (without proof). Hence we must impose $\hat{c}_i^\dagger \hat{c}_j^\dagger = -\hat{c}_j^\dagger \hat{c}_i^\dagger$, which implies that the anti-commutator $\left[\hat{c}_i^\dagger, \hat{c}_j^\dagger\right]_+ = 0$ vanishes.

Thus, the common approach is to replace the commutator relations by anti-commutator relations. For Fermions the **anti-commutator relations** of creation and annihilation operators have the form

$$\left[\hat{c}_i, \hat{c}_j^\dagger\right]_+ = \delta_{i,j} \qquad [\hat{c}_i, \hat{c}_j]_+ = 0 \qquad \left[\hat{c}_i^\dagger, \hat{c}_j^\dagger\right]_+ = 0$$

Let us go through the arguments one by one. To simplify the discussion let us only consider one orbital. The Fock space has only two states: The vacuum state $|0\rangle$ and the one-particle state $|1\rangle$. Now we can work out the matrix elements.

$$\begin{aligned}
\hat{c}^\dagger|0\rangle = |1\rangle \qquad \hat{c}^\dagger|1\rangle &= \left(\hat{c}^\dagger\right)^2|0\rangle = |0\rangle \\
\Rightarrow \begin{pmatrix} \langle 0|\hat{c}^\dagger|0\rangle & \langle 0|\hat{c}^\dagger|1\rangle \\ \langle 1|\hat{c}^\dagger|0\rangle & \langle 1|\hat{c}^\dagger|1\rangle \end{pmatrix} &= \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \\
\Rightarrow \begin{pmatrix} \langle 0|\hat{c}|0\rangle & \langle 0|\hat{c}|1\rangle \\ \langle 1|\hat{c}|0\rangle & \langle 1|\hat{c}|1\rangle \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \\
\Rightarrow \begin{pmatrix} \langle 0|\hat{c}^\dagger\hat{c}|0\rangle & \langle 0|\hat{c}^\dagger\hat{c}|1\rangle \\ \langle 1|\hat{c}^\dagger\hat{c}|0\rangle & \langle 1|\hat{c}^\dagger\hat{c}|1\rangle \end{pmatrix} &= \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\
\Rightarrow \begin{pmatrix} \langle 0|\hat{c}\hat{c}^\dagger|0\rangle & \langle 0|\hat{c}\hat{c}^\dagger|1\rangle \\ \langle 1|\hat{c}\hat{c}^\dagger|0\rangle & \langle 1|\hat{c}\hat{c}^\dagger|1\rangle \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}
\end{aligned}$$

From here it is easy to verify the anti-commutator relations given above for identical indices i and j.

Now we use the condition $[\hat{c}_i^\dagger, \hat{c}_j^\dagger]_+ = 0$. This implies that we need to introduce a sign convention for the states, which is sensitive to the order in which the particles are created. Let us define

$$|0, \ldots, 0, \underbrace{1}_{pos.i}, 0, \ldots, 0, \underbrace{1}_{pos.j}, 0, \ldots\rangle = \begin{cases} +\hat{c}_i^\dagger \hat{c}_j^\dagger|0\rangle & \text{for} \quad i > j \\ -\hat{c}_i^\dagger \hat{c}_j^\dagger|0\rangle & \text{for} \quad i < j \end{cases}$$

The sign convention of the kets translates into a sign convention for the bras.

$$\langle 0, \ldots, 0, \underbrace{1}_{pos.i}, 0, \ldots, 0, \underbrace{1}_{pos.j}, 0, | = \begin{cases} +\langle 0|\hat{c}_j \hat{c}_i & \text{for} \quad i > j \\ -\langle 0|\hat{c}_j \hat{c}_i & \text{for} \quad i < j \end{cases}$$

If we take the hermitian conjugate of the anti commutator relation $[\hat{c}_i^\dagger, \hat{c}_j^\dagger]_+ = 0$, we obtain the anticommutator relation for the annihilation operators $[\hat{c}_i, \hat{c}_j]_+ = 0$.

Let us now consider Fock space spanned by two one-particle orbitals: It contains the vacuum state $|00\rangle$, the one-particle states $|01\rangle, |10\rangle$ and the two-particle state $|11\rangle$. We investigate the operators $\hat{c}_1^\dagger \hat{c}_2$ and $\hat{c}_2 \hat{c}_1^\dagger$. Its matrix elements vanish except the ones between the ket $|01\rangle$ and the bra $\langle 10|$. Note in the following that $\hat{c}|0\rangle = 0 \neq |0\rangle$.

$$\begin{aligned}
\langle 10|\hat{c}_1^\dagger \hat{c}_2|01\rangle &= \langle 00|\hat{c}_1 \hat{c}_1^\dagger \hat{c}_2 \hat{c}_2^\dagger|00\rangle \overset{[\hat{c}_2, \hat{c}_2^\dagger]_+=1}{=} \langle 00|\hat{c}_1 \hat{c}_1^\dagger|00\rangle - \langle 00|\hat{c}_1 \hat{c}_1^\dagger \hat{c}_2^\dagger \underbrace{c_2|00\rangle}_{=0} \\
&= \langle 01|01\rangle = 1 \\
\langle 10|\hat{c}_2 \hat{c}_1^\dagger|01\rangle &= \langle 00|\hat{c}_1 \hat{c}_2 \hat{c}_1^\dagger \hat{c}_2^\dagger|00\rangle = (-1)^2\langle 11|11\rangle = 1
\end{aligned}$$

Since all other matrix elements vanish, we obtain the anti-commutator relation $[\hat{c}_2, \hat{c}_1^\dagger]_+ = 0$. The reader may verify that the same anti-commutator relation holds if the indices 1 and 2 are interchanged.

Real space representation

The number representation can be built on top of any complete one-particle basis set. We can also transform onto a basisset of position eigenstates. Without proof we list here the main changes.

The creation and annihilation operators in real space are $\hat\Psi^\dagger(\vec r)$ and $\hat\Psi(\vec r)$. the commutator relations for Bosons are

$$[\hat\Psi^\dagger(\vec r, \hat\Psi(\vec r')]_- = \delta(\vec r - \vec r') \qquad [\hat\Psi^\dagger(\vec r, \hat\Psi^\dagger(\vec r')]_- = 0 \qquad [\hat\Psi(\vec r, \hat\Psi(\vec r')]_- = 0$$

For Fermions we obtain instead the anti-commutator relations

$$[\hat\Psi^\dagger(\vec r, \hat\Psi(\vec r')]_+ = \delta(\vec r - \vec r') \qquad [\hat\Psi^\dagger(\vec r, \hat\Psi^\dagger(\vec r')]_+ = 0 \qquad [\hat\Psi(\vec r, \hat\Psi(\vec r')]_+ = 0$$

For Fermions and Bosons the number operator becomes the **particle density operator**

$$\hat n(\vec r) = \hat\Psi^\dagger(\vec r)\hat\Psi(\vec r)$$

The Hamiltonian for particles with charge q and mass m in an external potential $V(r)$ that interact electrostatically has the form

$$\hat H = \int d^3r\, \hat\Psi^\dagger(\vec r)\left[-\frac{\hbar^2}{2m}\vec\nabla^2 + V(\vec r)\right]\hat\Psi(\vec r) + \frac{q^2}{8\pi\epsilon_0}\int d^3r \int d^3r'\, \frac{\hat\Psi^\dagger(\vec r)\hat\Psi^\dagger(\vec r')\hat\Psi(\vec r')\hat\Psi(\vec r)}{|\vec r - \vec r'|}$$

Editor: Check the order of the operators. Thus the main difference from going from the first to the second quantization is to replace the wave function by the real-space creation and annihilation operators. Note that, like in the first quantization, the order of the operators is relevant and not immediately obvious.

Second quantization

Note that we arrived here at the second quantization by starting from the particle concept, since we started from one-particle states, two-particle states, etc. Previously (ΦSX: Quantum mechanics), we arrived at the second quantization by treating the wave function like a particle coordinate that needs to be quantized. In both cases we arrive at exactly the same Theory.

Further background can be found in the book by Haken[15].

8.3 Partition function

After this excursion to the field of second quantization, let us return to the simpler number representation. We investigate non-interacting particles and choose the number representation based on the Eigenstates of the corresponding one-particle Hamilton operator, which has eigenvalues ϵ_i. The many-particle states $|n_1, n_2, \ldots\rangle$ are eigenstates of the many-particle Hamiltonian and the particle number operator. The energy eigenvalue of such a many-particle state $|n_1, n_2, \ldots\rangle$ is then

$$E_{n_1,n_2,\ldots} = \sum_i \epsilon_i n_i$$

The eigenvalues of the particle number operator are

$$N_{n_1,n_2,\ldots} = \sum_i n_i$$

In thermal equilibrium, the density operator has the form

$$\begin{aligned}
\hat\rho^{eq} &\overset{\text{Eq. 6.10}}{=} \frac{1}{Z(T,\mu)} \sum_{n_1,n_2,\ldots,m_1,m_2,\ldots}^{1/\infty} |n_1,n_2,\ldots\rangle\langle n_1,n_2,\ldots|\mathrm{e}^{-\beta(\hat H - \mu\hat N)}|m_1,m_2,\ldots\rangle\langle m_1,m_2,\ldots| \\
&= \frac{1}{Z(T,\mu)} \sum_{n_1,n_2,\ldots}^{1/\infty} |n_1,n_2,\ldots\rangle \mathrm{e}^{-\beta(E_{n_1,n_2,\ldots} - \mu N_{n_1,n_2,\ldots})}\langle n_1,n_2,\ldots|
\end{aligned}$$

The partition function $Z(T,\mu)=\mathrm{Tr}[\mathrm{e}^{-\beta(\hat{H}-\mu\hat{N})}]$ is obtained as sum over all states[4], or alternatively as product of one-particle partition functions.

$$Z(T,\mu)\overset{\text{Eq. 6.9}}{=}\sum_{n_1,n_2,\ldots}^{1/\infty}\exp\Big[-\frac{1}{k_BT}\sum_i(\epsilon_i-\mu)n_i\Big]\overset{\text{non-trivial!}}{=}\prod_{i=1}^{\infty}\sum_{n_i=0}^{1/\infty}\exp\Big[-\frac{1}{k_BT}(\epsilon_i-\mu)n_i\Big]$$

The free energy has a form of a sum of single particle free energies[5]

$$\begin{aligned}F(T,\mu)\overset{\text{Eq. 6.11}}{=}-k_BT\ln Z&=\sum_{i=1}^{\infty}\Big[-k_BT\ln\Big(\sum_{n_i=0}^{1/\infty}\exp[-\frac{1}{k_BT}(\epsilon_i-\mu)n_i]\Big)\Big]\\&=\sum_{i=1}^{\infty}F_i(T,\mu)\end{aligned}$$

with the one-particle free energy per state

$$F_i(T,\mu)=-k_BT\ln[Z_i(\epsilon-\mu,T)]=-k_BT\ln\Big[\sum_{n_i=0}\exp\Big(-\frac{1}{k_BT}(\epsilon_i-\mu)n_i\Big)\Big]\tag{8.1}$$

This equation represents an important conceptual step: Instead of expressing the partition function over all many-particle states, we obtain it now as a weighted sum over one-particle states. This is a dramatic reduction in the complexity of the problem, which is possible because the particles are not interacting.

For Bosons the sum runs from zero to infinity, while for Fermions the sum is limited to the two terms $n_i=0$ and $n_i=1$ due to the Pauli principle. The corresponding one-particle free energies have already been discussed, when we discussed the two-state system and the harmonic oscillator. We obtain

$$F_i^F(T,\mu)\overset{\text{Eqs. 8.1,7.3}}{=}-kT\ln\left[1+\exp(-\frac{\epsilon_i-\mu}{k_BT})\right]\tag{8.2}$$

for Fermions and

$$F_i^B(T,\mu)\overset{\text{Eqs. 8.1,7.14}}{=}+kT\ln\left[1-\exp(-\frac{\epsilon_i-\mu}{k_BT})\right]\tag{8.3}$$

for Bosons.

8.4 One-particle density of states

In the previous section we have seen that the thermodynamic potentials of a system of many non-interacting systems can be expressed by a sum over one-particle states. Furthermore, most sums include terms that only depend on the one-particle energy of each state. That is they are of the form $\sum_i g(\epsilon_i)$. In those cases it is possible to separate the information that depends only on the Hamiltonian from the choice of the ensemble.

The system dependent information can be described by the **density of states** defined as

$$D(\epsilon)=\sum_i\delta(\epsilon-\epsilon_i).\tag{8.4}$$

where the ϵ_i are the eigenvalues of the one-particle Hamiltonian.

---

[4] The reader may want to verify that the sum is identical to the result obtained by optimizing the state operator.

[5] They should better be called single-orbital free energies

Any quantity A, which is a sum over states $A = \sum_i a(\epsilon_i)$ can be represented with the help of the density of states function as

$$A = \sum_i a(\epsilon_i) = \int d\epsilon\; D(\epsilon) a(\epsilon) \; .$$

As an example the Free energies for Fermions and Bosons have the form

$$\begin{aligned} F^F(T,\mu) &\overset{\text{Eq. 8.2}}{=} -kT \int d\epsilon\; D(\epsilon) \ln[1 + \exp(-\frac{\epsilon-\mu}{k_BT})] \\ F^B(T,\mu) &\overset{\text{Eq. 8.3}}{=} +kT \int d\epsilon\; D(\epsilon) \ln[1 - \exp(-\frac{\epsilon-\mu}{k_BT})] \end{aligned}$$

Note that Fermions corresponds directly to a two state model for each one-particle state. The state can be filled by one electron or not, which are the two possible states. That is also why the thermodynamic properties are identical to the two state model

A system of Bosons corresponds directly to a harmonic oscillator for each one-particle energy level. We can fill in arbitrary many Bosons into a one particle state. Each particle then has the energy of the one-particle state. Hence the energy spectrum of the many-particle system corresponds to an infinite ladder of equispaced many-particle states. The main difference is that we do not include here the zero-point vibrational energy. It is left out, because the vacuum is always there and therefore does not affect energy differences.

The zero-point energy plays however an important role in elementary particle physics: According to Einstein's general theory of relativity any form of energy corresponds to a mass, which can produce a gravitational field. Since there are infinitely many possible one-particle states, one can easily arrive at an infinite gravitational field. This is a paradoxon. In order to avoid this problem people speculate about super-symmetric theories for elementary particles. One can argue that Fermions should have a negative zero-point energy, and by exactly pairing Fermions and Bosons one can arrive at a theory, where the vacuum does not have a mass.

In statistical mechanics one can often approximate the density of states by a model density of states. In the following we will investigate the two most important model systems. Free particles with and without mass.

8.4.1 Density of states and dispersion relation

The one-particle density of states can readily be obtained from the dispersion relation $\epsilon(p)$. Independent particles in a constant potential have a conserved momentum $\vec{p} = \hbar\vec{k}$. This implies that they can be described by plane waves

$$\Psi(\vec{r}) = \frac{1}{\sqrt{V}} \mathrm{e}^{i\vec{k}\vec{r}}$$

Note that even electrons in a periodic potential, that is electrons ins a crystal, have a conserved momentum. The wave function is not a plane wave, but a plane wave modulated by some periodic, p dependent function, as seen in the **Bloch theorem**. These dispersion relations are called **band structure** and can be calculated with first-principles calculations. Also lattice vibrations in a crystal can be classified by their wave vector, resulting in a phonon band structure.

If we consider a system with periodic boundary conditions in a a box with side-lengths L_x, L_y, L_z, the states are quantized, so that $k_iL_i = 2\pi n_i$ where n_i is an arbitrary integer. Thus there are only states with $\vec{k} = (\frac{2\pi}{L_x}i, \frac{2\pi}{L_y}j, \frac{2\pi}{L_z}k)$, where i,j,k are arbitrary integers. The volume V of the box is $V = L_xL_yL_z$.

Thus the density of states is

$$D(\epsilon) = \sum_{i,j,k} \delta\Big(\epsilon(\underbrace{\frac{2\pi\hbar}{L_x}i}_{p_x}, \underbrace{\frac{2\pi\hbar}{L_x}j}_{p_y}, \underbrace{\frac{2\pi\hbar}{L_x}k}_{p_z}) - \epsilon\Big)$$

We can attribute to each state a volume in k-space, namely

$$\Delta V_k = \frac{2\pi}{L_x}\frac{2\pi}{L_y}\frac{2\pi}{L_z} = \frac{(2\pi)^3}{V} \tag{8.5}$$

Using the relation $\vec{p} = \hbar\vec{k}$ we can convert this volume into a volume element in momentum space, namely

$$\Delta V_p = \frac{(2\pi\hbar)^3}{V} \tag{8.6}$$

If the size of the box, that is L_x, L_y, L_z, is made very large, the volume element attributed to a single state in momentum space becomes very small. Thus we can replace the sum by an integral, where ΔV_p is represented by d^3p.

$$\begin{aligned} D(\epsilon) &= \frac{V}{(2\pi\hbar)^3}\sum_{i,j,k}\underbrace{\frac{(2\pi\hbar)^3}{V}}_{\rightarrow d^3p}\delta\Big(\epsilon(\vec{p}_{i,j,k})-\epsilon\Big) \\ &\overset{L_i\rightarrow\infty}{=} \frac{V}{(2\pi\hbar)^3}\int d^3p\;\delta\Big(\epsilon(\vec{p})-\epsilon\Big) \end{aligned} \tag{8.7}$$

It is intuitively clear that the expression for the density of states is related to a surface integral over a surface of constant energy. This will be shown in the following.

In order to transform the expression for the density of states into a surface integral in momentum space, it is convenient to introduce the **number of states** $N(\epsilon)$.

The number of states is defined as the number of states that lie below the energy epsilon.

$$N(\epsilon) = \sum_{i,j,k}\theta(\epsilon-\epsilon(p_{i,j,k})) \overset{V\rightarrow\infty}{=} \frac{V}{(2\pi\hbar)^3}\int d^3p\;\theta(\epsilon-\epsilon(\vec{p})) \tag{8.8}$$

where $\theta(x)$ is the Heaviside function $\theta(x)$, which vanishes for $x<0$ and is equal to unity for $x>0$. The Heaviside function is related to the δ-function via

$$\theta(x) = \int_{-\infty}^{x} dx'\;\delta(x') \qquad \Leftrightarrow \qquad \partial_x\theta(x) = \delta(x) \tag{8.9}$$

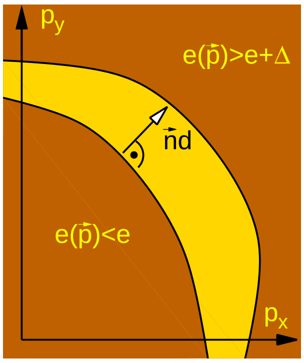
This allows us to relate the number of states to the density of states Thus we obtain

$$\begin{aligned} \partial_\epsilon N(\epsilon) &\overset{\text{Eq. 8.8}}{=} \frac{V}{(2\pi\hbar)^3}\int d^3p\;\partial_\epsilon\theta(\epsilon-\epsilon(\vec{p})) \overset{\text{Eq. 8.9}}{=} \frac{V}{(2\pi\hbar)^3}\int d^3p\;\delta(\epsilon-\epsilon(\vec{p})) \\ &\overset{\text{Eq. 8.7}}{=} D(\epsilon) \end{aligned} \tag{8.10}$$

Here we show how the volume integral with the delta function can be converted into a surface integral.

$$D(\epsilon) \overset{Eq.~8.10}{=} \partial_\epsilon N(\epsilon) = \lim_{\Delta\to 0} \frac{V}{(2\pi\hbar)^3} \int d^3p \, \underbrace{\frac{\theta(\epsilon+\Delta-\epsilon(p)) - \theta(\epsilon-\epsilon(p))}{\Delta}}_{\to\delta(\epsilon-\epsilon(p))}$$

The integral over the difference of the two Heaviside function corresponds to the volume of a volume sheet, which is enclosed by the surfaces defined by $\epsilon(\vec{p}) = \epsilon$ and $\epsilon(\vec{p}) = \epsilon + \Delta$.



Let us calculate the distance of two points on the two surfaces. Let us pick one point $\vec{p}_0$ on the surface defined by $\epsilon(\vec{p}) = \epsilon$. The closest neighbor on the other surface $\vec{p}_1$ lies opposite to $\vec{p}_0$, that is $\vec{p}_1 = \vec{p}_0 + \vec{n} \cdot d$, where d is the distance of the points and $\vec{n} = \frac{1}{|\vec{\nabla}_p\epsilon|}\vec{\nabla}_p\epsilon$ is the normal vector of the surface. $\vec{p}_1$ lies on the other surface and therefore fulfills

$$\epsilon(\vec{p}_1) = \underbrace{\epsilon}_{\epsilon(\vec{p}_0)} + \Delta$$

$$\Rightarrow \epsilon(\vec{p}_0) + \Delta = \epsilon(\vec{p}_0 + \vec{n}\cdot d) \overset{Taylor}{=} \epsilon(\vec{p}_0) + \vec{n}\cdot d\, \vec{\nabla}_p\epsilon + O(d^2)$$

$$\Rightarrow d = \frac{\Delta}{\vec{n}\vec{\nabla}_p\epsilon} = \frac{\Delta}{\frac{\vec{\nabla}_p\epsilon}{|\vec{\nabla}_p\epsilon|}\vec{\nabla}_p\epsilon} = \frac{\Delta}{|\vec{\nabla}_p\epsilon|}$$

Thus we obtain the thickness d of the sheet. Note that the thickness depends on the position on the sheet. The volume element of the sheet can then be written as $d^3p = dA \cdot d$

Thus we can write the volume as

$$D(\epsilon) = \frac{V}{(2\pi\hbar)^3} \oint d^2A_p \, \frac{1}{|\vec{\nabla}_p\epsilon|}$$

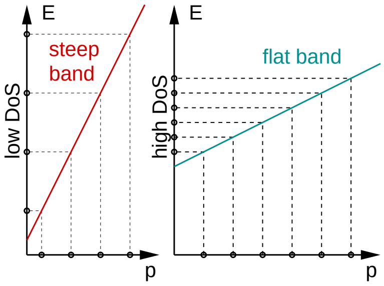
Thus we obtain the density of states from the dispersion relation as

ONE-PARTICLE DENSITY OF STATES PER VOLUME

$$g(\epsilon) \overset{\text{def}}{=} \frac{1}{V} D(\epsilon) = \int \frac{d^3p}{(2\pi\hbar)^3} \delta\left(\epsilon - \epsilon(\vec{p})\right) = \frac{1}{(2\pi\hbar^3)} \oint d^2A_p \, \frac{1}{|\vec{\nabla}_p\epsilon(p)|} \qquad (8.11)$$

Note that $\vec{\nabla}_p\epsilon(p)$ is the velocity (group-velocity) of the particle. This follows from Hamilton's equation $\dot{x}_i = \frac{\partial H(\vec{p},\vec{x})}{\partial p_i}$. It also follows from the expression of the group velocity of a wave packet $v_i = \frac{\partial\omega(\vec{k})}{\partial k_i} = \frac{\partial\epsilon(\vec{p})}{\partial p_i}$.

Thus the density of states is proportional to the area of a constant energy surface and inversely proportional to the mean velocity of the particles at the constant energy surface. Thus, flat bands, which correspond to slow particles have a large contribution to the density of states. Steep bands, which are related to fast particles, contribute little to the density of states at a given energy, but they contribute over a large energy range.



In the following we will calculate the density of states of two model systems the free particle with and without mass.

8.4.2 Free particle density of states with mass

The dispersion relation of a particle with mass is

$$\epsilon(p)=\frac{p^2}{2m}$$

The density of states is obtained exploiting the fact that $|\vec{p}|$ is constant on a surface of constant energy

$$\begin{aligned}
D(\epsilon) &= \frac{V}{(2\pi\hbar)^3}\int dA\frac{1}{|\nabla_p\epsilon|}=\frac{V}{(2\pi\hbar)^3}\underbrace{4\pi p^2}_{\int dA}\frac{1}{|\vec{p}|/m}\\
&\overset{|p|=\sqrt{2m\epsilon}}{=} \frac{V}{(2\pi\hbar)^3}4\pi(\sqrt{2m\epsilon})^2\frac{1}{\sqrt{2m\epsilon}/m}\\
&= 2\pi V\left(\frac{\sqrt{2m}}{2\pi\hbar}\right)^3\sqrt{\epsilon} \qquad (8.12)
\end{aligned}$$

One should remember that the density of states has a square-root behavior in three dimensions. Note however that two-dimensional particles such as a 2-dimensional electron gas has a radically different density of states. A one-dimensional particle, such as an electron in a one-dimensional conductor has a density of states proportional to $\frac{1}{\sqrt{\epsilon}}$, a particle moving in two dimensions has a density of states which is constant.

This model is important to describe the behavior of states at band edges. For example the electron states of a semiconductor at both sides of the Fermi level are the ones relevant for the electric and thermodynamic properties of a semiconductor. When we approximate the band about the minimum to second order, we obtain an approximate dispersion relation

$$\epsilon(p)=\epsilon_c+\frac{1}{2}(\vec{p}-\vec{p}^0)\mathbf{m}^{*-1}(\vec{p}-\vec{p}^0)$$

where ϵ_c is the conduction band minimum, $\vec{p}^0$ is the momentum of the minimum and $\mathbf{m}^*$ is the effective mass tensor. $\mathbf{m}^*$ is nothing but the inverse of the second derivatives of $\epsilon(\vec{p})$ at the band edge. It is important to note that the effective mass can be a tensor. Furthermore there may be several degenerate minima, so that one obtained different types of conduction electrons.

Similarly we can expand the top of the valence band to quadratic order about the minimum. These are the hole bands. The mass is negative. Since it is a missing electron that conducts, the hole has also a negative charge. The hole is conceptually analogous to the antiparticle of the electron, the positron. It is quite common that concepts from elementary particle physics carry over into solid state physics. In the latter however the phenomena can be observed at much lower energies and put into practical use.

8.4.3 Free particle density of states without mass

Mass-less particles are used to describe, for example, the acoustical branch of phonons (lattice vibrations). It can also used to describe electrons in a metal: In thermodynamics only the electrons near the Fermi level are of relevance, so that the band structure $\epsilon(k)$ can be approximated in this region by a linear dispersion relation. Another example where the dispersion relation is linear is light, i.e. photons.

The dispersion relation of a particle without mass is

$$\epsilon(p)=c|\vec{p}|$$

where c is the speed of light, if we discuss photons, or the speed of sound if we discuss phonons, that is lattice vibrations. For metals c is called the Fermi velocity.

The density of states is obtained exploiting the fact that $|\vec{p}|$ is constant on a surface of constant energy

$$\begin{aligned}
D(\epsilon) &= \frac{V}{(2\pi\hbar)^3}\int dA \frac{1}{|\nabla_p\epsilon|} = \frac{V}{(2\pi\hbar)^3}\underbrace{4\pi p^2}_{\int dA}\frac{1}{c} \\
&\overset{|p|=\epsilon/c}{=} \frac{V}{(2\pi\hbar)^3}4\pi(\epsilon/c)^2 c \\
&= \frac{4\pi V}{(2\pi\hbar c)^3}\epsilon^2 \qquad (8.13)
\end{aligned}$$

8.5 The ideal Fermi Gas

The Fermi gas is a model for electrons in a molecule or a solid. Some remarks are appropriate, though, because electrons interact strongly by the Coulomb interaction. Why can we treat them as non-interacting particles?

The reason why the independent particle picture is nearly appropriate is that we consider not electrons themselves, but **quasiparticles**[6], which are made up of electrons and their exchange and correlation hole. Let us consider the electron density $n_2(r_0, r)$ that is seen by a single reference electron at r_0: Because of the Pauli principle, this electron density vanishes at the position of the reference electron. If we build the difference of the total electron density $n(r)$ and the density of the N-1 remaining electrons we obtain the so-called exchange correlation hole $h(r_0, r) = n_2(r_0, r) - n(r)$. The exchange correlation hole is localized close to the reference electron. Its density is just opposite to that of the reference electron. Thus we may define a quasi-particle composed of the reference electron and the exchange correlation hole, which is overall neutral and therefore does not have a long-ranged interaction. There is still an interaction at small distances, which however is strongly reduced because the electrons cannot come close to each other due to the Pauli principle. Hence the electrostatic interaction is strongly reduced compared to what has been expected.

Let us now evaluate the major thermodynamic quantities for the ideal Fermion gas. The Free energy for Fermions is

$$\begin{aligned}
F(T,\mu) &= -k_BT\ln Z(T,\mu) \\
&= \sum_i\Big\{-k_BT\ln\Big[\sum_{n_i=0}^{1}\exp\Big(-\frac{1}{k_BT}(\epsilon-\mu)n_i\Big)\Big]\Big\} \\
&= \int d\epsilon D(\epsilon)\Big[-k_BT\ln[1+\mathrm{e}^{-\beta(\epsilon-\mu)}]\Big]
\end{aligned}$$

The entropy is obtained as

$$S = -\frac{\partial F}{\partial T} = \int d\epsilon D(\epsilon)\Big[k_B\ln[1+\mathrm{e}^{-\beta(\epsilon-\mu)}] + k_B\frac{\beta(\epsilon-\mu)\mathrm{e}^{-\beta(\epsilon-\mu)}}{1+\mathrm{e}^{-\beta(\epsilon-\mu)}}$$

The internal energy is

$$U = F + TS + \mu N = \int d\epsilon D(\epsilon)\epsilon\frac{1}{1+\mathrm{e}^{\beta(\epsilon-\mu)}}$$

---

[6] Quasiparticles is a concept that describes when strongly interacting particles or excitations behave cooperatively like a composite particle, which only weakly interact with each other and the environment. A common example is a polaron, which is a quasiparticle of an electron and a lattice distortion of a polar cristal. As the electron passes through a polar crystal such as rocksalt, it introduces lattice distortions to lower its energy. When the electron moves, the distortions move along with the electron. This so-called polaron behaves like a free particle with a larger mass than the electron.

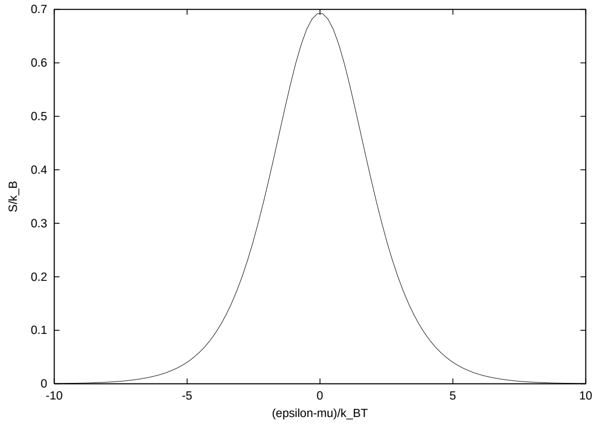


Fig. 8.1: Entropy contribution in units of k_B of a one-particle state for a Fermi gas as function of $(\epsilon-\mu)/k_BT$. States far from the chemical potential do not contribute since they are either with certainty full or empty. States at the chemical potential contribute with $k_B\ln[2]$, because both states, with and without a particle, are accessible.

Here we introduce the occupation numbers

$$f(\epsilon)=\frac{1}{\mathrm{e}^{\beta(\epsilon-\mu)}+1} \tag{8.14}$$

which is also called the **Fermi distribution function**.

Using the Fermi distribution function, Eq. 8.14, we obtain

$$U[f(\epsilon)]=\int d\epsilon\; D(\epsilon)\cdot\epsilon\, f(\epsilon) \tag{8.15}$$

$$F[f(\epsilon)]=\int d\epsilon\; D(\epsilon)\cdot\Big[k_BT\ln[1-f(\epsilon)]\Big] \tag{8.16}$$

$$S[f(\epsilon)]=\int d\epsilon\; D(\epsilon)\cdot k_B\Big[(1-f(\epsilon)\ln[1-f(\epsilon)]+k_Bf(\epsilon)\ln[f(\epsilon)]\Big] \tag{8.17}$$

where the dependence on temperature and chemical potential is included indirectly via the Fermi distribution given in Eq. 8.14.

The free energy can be directly considered as a functional of the occupations $f(\epsilon)$. If we minimize the free energy with respect to the occupations, we obtain directly the Fermi distribution function. This is the so-called Mermin functional used in electronic structure calculations.

$$\begin{aligned}F([f(\epsilon)],T,\mu)&=U+TS\\&=\int d\epsilon\; D(\epsilon)\left[(\epsilon-\mu)f(\epsilon)+k_BT\left(f(\epsilon)\ln[f(\epsilon)]+(1-f(\epsilon))\ln[1-f(\epsilon)]\right)\right]\end{aligned}$$

The minium principle yields

$$\begin{aligned}dF([f(\epsilon)])&=\int d\epsilon\; D(\epsilon)\left[(\epsilon-\mu)+k_BT\left(\ln[f(\epsilon)]+1-\ln[1-f(\epsilon)]-1\right)\right]df(\epsilon)\overset{!}{=}0\\&\Rightarrow 0=(\epsilon-\mu)+k_BT\left(\ln[f(\epsilon)]+1-\ln[1-f(\epsilon)]-1\right)\\&\Rightarrow 0=(\epsilon-\mu)+k_BT\ln\left[\frac{f(\epsilon)}{1-f(\epsilon)}\right]\\\Rightarrow \mathrm{e}^{-\beta(\epsilon-\mu)}&=\frac{f(\epsilon)}{1-f(\epsilon)}\qquad\Rightarrow\qquad f(\epsilon)=\frac{1}{1+\mathrm{e}^{\beta(\epsilon-\mu)}}\end{aligned}$$

Editor: This may be visualized by showing the contribution to F at each energy.

Let us have a closer look at the Fermi distribution function. For $\epsilon-\mu << -k_BT$ we obtain an occupation equal to one. For $\epsilon-\mu >> k_BT$ we obtain $\mathrm{e}^{-\beta(\epsilon-\mu)}$.

We analyze the behavior of the Fermi distribution by considering the limits at infinity and near the chemical potential.

- For high energies, i.e. $\epsilon-\mu >> k_BT$, we obtain

  $$\frac{1}{\mathrm{e}^{\beta(\epsilon-\mu)}+1}=\mathrm{e}^{-\beta(\epsilon-\mu)}+O(\mathrm{e}^{\beta(\epsilon-\mu)})^2$$

  This is just the Boltzmann factor for an excitation from the chemical potential. These states are empty, with an occasional electron thermally excited into high energy region.

- For energies much smaller than the chemical potential, i.e. $\mu-\epsilon >> k_BT$, we obtain

  $$\frac{1}{\mathrm{e}^{\beta(\epsilon-\mu)}+1}=1-\mathrm{e}^{\beta(\epsilon-\mu)}+O(\mathrm{e}^{\beta(\epsilon-\mu)})^2$$

  The states are occupied with the exception of electrons, that are thermally excited into the unoccupied states. We could also say that some holes are excited from the chemical potential into these low energy states.

- Near the chemical potential we make a Taylor expansion in the energy.

  $$\frac{1}{\mathrm{e}^{\beta(\epsilon-\mu)}+1}=\frac{1}{2}+\frac{\epsilon-\mu}{2k_BT}+O(\epsilon-\mu)^2$$

  Thus the transition from occupied to unoccupied takes place in an energy region of about $2k_BT$ near the chemical potential. At low temperatures, the transition is abrupt.

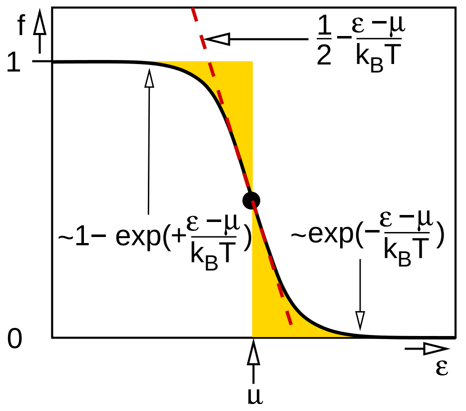


Fig. 8.2: Fermi distribution function.

8.6 Application: semiconductor statistics

8.6.1 Intrinsic semiconductor

Let us consider the charge carriers in a semiconductor. The conductivity is proportional to the number of free charge carriers, that is electrons in the conduction band and holes in the valence band. If the charge carrier concentration is by the semiconductor itself and not by the presence of dopants, we call the semiconductor **intrinsic**.

First we have to define the system: We need to know the density of states. The density of states is obtained from the dispersion relation. The dispersion relation of a semiconductor is characterized by a band gap E_g in the range of about one electron volt. The band gap in silicon is 1.14 eV wide. Near the band edges, the dispersion $\epsilon(k)$ can be approximated by parabolas. This implies that the conduction electrons behave like free particles with a positive effective mass m^*.

$$\begin{aligned}
\epsilon(k) &= \epsilon_0 + \frac{1}{2}\frac{d^2\epsilon(k)}{dk^2}(k-k_0)^2 + O(k-k_0)^3 \\
&= \epsilon_0 + \frac{(\hbar k - \hbar k_0)^2}{2m^*} + O(k-k_0)^3 \\
m^* &= \hbar^2\Big[\frac{d^2\epsilon(k)}{dk^2}|_{k_0}\Big]^{-1}
\end{aligned}$$

Editor: Mention that effective mass is a tensor! Provide a link to Ashcroft-Mermin[?] p569 where ϵ_0 is the energy of the band edge. Do not confuse this symbol with the vacuum permittivity. k_0 is the position of the band edge in reciprocal space. Usually, there are several bands close to the band edges, so that each band contributes with their own effective masses. In reality there are several hole bands and sometimes several electron bands. Furthermore, the effective mass is a tensor, that is the masses depend on the direction into which the electron or hole travels. The hole band is a parabola with negative curvature $E(k) = -\frac{1}{2}m_h^*\vec{k}^2$.

Let us count the number of conduction electrons and those of the holes in each branch of the dispersion relation. σ denotes a particular branch of the dispersion relation.

$$\begin{aligned}
N_{e,\sigma} &= \int_{E_g}^{\infty} d\epsilon\ D_\sigma(\epsilon)f(\epsilon-\mu) \\
N_{h,\sigma} &= \int_{-\infty}^{0} d\epsilon\ D_\sigma(\epsilon)[1-f(\epsilon-\mu)] = \int_{-\infty}^{0} d\epsilon\ D(\epsilon)f(-\epsilon+\mu)] \\
&= \int_{0}^{\infty} d\epsilon\ D(-\epsilon)f(\epsilon+\mu)]
\end{aligned}$$

We see here that the holes behave like electrons, if we simply change the sign of the energy scale. Holes behave like antiparticles of the electrons, where the rest energy m_0c^2 is replaced by one-half of the band gap.

In the following we will try to evaluate the number of charge carriers as function of temperature. Using the free particle density of states we obtain the number of carriers in a given branch of the dispersion relation. In order to simplify the discussion we have shifted the energy-zero to the band edge.

$$\begin{aligned}
N &= \int_0^\infty d\epsilon\ D(\epsilon)f(\epsilon-\mu) = \frac{2\pi V\sqrt{2m}^3}{(2\pi\hbar)^3}\int_0^\infty d\epsilon \frac{\sqrt{\epsilon}}{1+\mathrm{e}^{\beta(\epsilon-\mu)}} \\
&= 2\pi V\Big(\frac{2m^*k_BT}{(2\pi\hbar)^2}\Big)^{\frac{3}{2}}\Big(\frac{\sqrt{\pi}}{2}\Big)\Big(\sum_{n=1}^{\infty}(-1)^{n+1}n^{-\frac{3}{2}}\mathrm{e}^{\frac{n\mu}{k_BT}}\Big) \qquad (8.18)
\end{aligned}$$

Here we simplify the expression

$$N = \int_0^\infty d\epsilon\ D(\epsilon)f(\epsilon-\mu) \overset{Eq.\ 8.12}{=} \frac{2\pi V\sqrt{2m}^3}{(2\pi\hbar)^3}\int_0^\infty d\epsilon\ \frac{\sqrt{\epsilon}}{1+\mathrm{e}^{\beta(\epsilon-\mu)}}$$

used in Eq. 8.18.
We need to solve an integral of the form

$$\int dx \frac{\sqrt{x}}{1+\mathrm{e}^{x-a}}$$

with $a = \mu/(k_BT)$ *cont'd*

cont'd

Since I do not know an analytical solution, let me try an expansion in a small quantity. We already used the assumption that the chemical potential lies far from the band edges. Thus we can consistently assume that $y = \mathrm{e}^{-(x+a)}$ is a small quantity.

We rewrite the integral as

$$\int dx\ \sqrt{x}\mathrm{e}^{-(x-a)}\frac{1}{1+\mathrm{e}^{-(x-a)}}$$

I use the Taylor series of $\frac{1}{1+y}$

$$\frac{1}{1+y} = \sum_{n=0}^{\infty}\frac{1}{n!}\underbrace{\frac{d^n(1+y)^{-1}}{dy^n}\Big|_{y=0}}_{n!(-1)^n} y^n = \sum_{n=0}^{\infty}(-y)^n$$

and insert $y = 1 + \mathrm{e}^{-x+a}$ so that

$$\begin{aligned}\int dx\sqrt{x}\mathrm{e}^{-(x-a)}\frac{1}{1+\mathrm{e}^{-(x-a)}} &= \sum_{n=0}(-1)^n\int dx\ \sqrt{x}\mathrm{e}^{-(n+1)(x-a)}\\ &= \sum_{n=1}(-1)^{n+1}\int dx\ \sqrt{x}\mathrm{e}^{-n(x-a)}\\ &= \sum_{n=1}(-1)^{n+1}\mathrm{e}^{na}\int dx\ \sqrt{x}\mathrm{e}^{-nx}\end{aligned}$$

Another variable transform $z = nx$ leads to

$$\int dx\frac{\sqrt{x}}{1+\mathrm{e}^{x-a}} = \sum_{i=1}^{\infty}\frac{(-1)^{n+1}\mathrm{e}^{na}}{n^{\frac{3}{2}}}\int_0^{\infty} dz\ \sqrt{z}\mathrm{e}^{-z}$$

cont'd

cont'd

Now we need to resolve the integral in parenthesis: We use a variable transform $y = \sqrt{z}$ so that $z = y^2$ and $dz = 2ydy$ and look up the final integral

$$\int_0^{\infty} dz\ \sqrt{z}\mathrm{e}^{-z} = 2\int_0^{\infty} dy y^2\mathrm{e}^{-y^2} = \frac{\sqrt{\pi}}{2}$$

We obtain

$$\int dx\frac{\sqrt{x}}{1+\mathrm{e}^{x-a}} = (\frac{\sqrt{\pi}}{2})\sum_{i=1}^{\infty}\frac{(-1)^{n+1}\mathrm{e}^{na}}{n^{\frac{3}{2}}}$$

cont'd

cont'd
Thus we obtain

$$
\begin{aligned}
N &= \frac{2\pi\sqrt{2m^*}^3 V}{(2\pi\hbar)^3}\int_0^\infty d\epsilon\sqrt{\epsilon}\frac{1}{1+\mathrm{e}^{\beta(\epsilon-\mu)}} \\
&= \frac{2\pi\sqrt{2m^*}^3 V}{(2\pi\hbar)^3}\sqrt{k_BT}^3\int_0^\infty dx\frac{\sqrt{x}}{1+\mathrm{e}^{x-\beta\mu}} \\
&= 2\pi V\Big(\frac{2m^*k_BT}{(2\pi\hbar)^2}\Big)^{\frac{3}{2}}(\frac{\sqrt{\pi}}{2})\Big(\sum_{n=1}^{\infty}(-1)^{n+1}n^{-\frac{3}{2}}\mathrm{e}^{\frac{n\mu}{k_BT}}\Big) \\
&= V\Big(\frac{m^*k_BT}{2\pi\hbar^2}\Big)^{\frac{3}{2}}\Big[\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{\sqrt{n^3}}\Big(\mathrm{e}^{\frac{\mu}{k_BT}}\Big)^n\Big]
\end{aligned}
$$

In the beginning we placed the zero of the energy scale at the conduction band. To obtain a more general expression we introduce again the conduction band edge ϵ_c for electrons and the valence band edge ϵ_v for the holes. Furthermore we evaluate the spatial density n of the charge carriers instead of their total number.

$$
n_e(T,\mu) = \frac{N_e(T,\mu}{V} = \Big(\frac{m_e^*k_BT}{2\pi\hbar^2}\Big)^{\frac{3}{2}}\Big[\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{\sqrt{n^3}}\Big(\mathrm{e}^{-(\frac{\epsilon_c-\mu)}{k_BT}}\Big)^n\Big] \tag{8.19}
$$

$$
n_h(T,\mu) = \frac{N_h(T,\mu}{V} = \Big(\frac{m_h^*k_BT}{2\pi\hbar^2}\Big)^{\frac{3}{2}}\Big[\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{\sqrt{n^3}}\Big(\mathrm{e}^{-(\frac{\mu-\epsilon_c)}{k_BT}}\Big)^n\Big] \tag{8.20}
$$

This result is still exact within the effective mass approximation, that is the approximation of parabolic bands.

We need to sum this result over all branches of the dispersion relation, which are labeled by σ. The band structure often exhibits several degenerate minima of the conduction band and valence band maxima. Thus we must sum the carrier concentrations over all relevant branches of the dispersion relation. Each branch of the dispersion relation may have its own effective mass. This is especially the case for the valence band top, where two heavy holes and one light hole contribute. First and foremost, we must not forget **spin degeneracy**. There are states for spin-up and states for spin-down electrons, which are degenerate in the absence of magnetic fields. Instead of summing over these bands we may also consider the spin degeneracy by a factor two, if we exclude the explicit sum over branches related only by spin reversal. Secondly we must sum over all minima of the conduction band and all maxima of the valence band.

The effective mass approximation is only valid, if the charge carriers are concentrated close to the band edges. This is the case if the chemical potential lies far from the band edges that is $\frac{\epsilon_c-\mu}{k_BT} >> 1$ and $\frac{\mu-\epsilon_v}{k_BT} >> 1$. With this approximation however only the first term in the series expansion of Eqs. 8.19 and 8.20 contributes appreciably. Thus we we simplify the expression assuming that $\frac{\epsilon_c-\mu}{k_BT} >> 1$ and $\frac{\mu-\epsilon_v}{k_BT} >> 1$, which allows to ignore the higher powers of the series expansion.

We obtain the approximate result for the carrier concentrations.

$$
n_e(T,\mu) \approx P_e(T)\mathrm{e}^{\frac{-(\epsilon_c-\mu)}{k_BT}} \tag{8.21}
$$

$$
n_h(T,\mu) \approx P_h(T)\mathrm{e}^{\frac{-(\mu-\epsilon_v)}{k_BT}} \tag{8.22}
$$

with $P(T)$ defined as

$$
\begin{aligned}
P_e(T) &= \sum_\sigma\Big(\frac{m_{e,\sigma}^*k_BT}{2\pi\hbar^2}\Big)^{\frac{3}{2}} \\
P_h(T) &= \sum_\sigma\Big(\frac{m_{h,\sigma}^*k_BT}{2\pi\hbar^2}\Big)^{\frac{3}{2}}
\end{aligned} \tag{8.23}
$$

This expression has the form of a Boltzmann factor: Thermal energies lift electrons from the chemical potential to the conduction band edge. This picture is oversimplified of course, because it does not include the pre factor $P(T)$, which behaves as if there were a single set of degenerate states for the conduction band edge, with a temperature dependent degeneracy $VP(T)$.

Determination of the chemical potential

We still need to know the chemical potential as function of temperature, to know the number of electrons and holes at any temperature. The chemical potential follows from the requirement that the number of electrons must be equal to the number of holes, because otherwise the system would be charged.

$$\begin{aligned}
n_e(T,\mu) &= n_h(T,\mu)\\
P_e(T)\mathrm{e}^{-\beta(\epsilon_c-\mu)} &= P_h(T)\mathrm{e}^{-\beta(\mu-\epsilon_v)}\\
k_BT\ln[P_e(T)]-\epsilon_c+\mu &= k_BT\ln[P_h(T)]-\mu+\epsilon_v\\
\mu(T) &= \frac{\epsilon_c+\epsilon_v}{2}+k_BT\ln[\frac{P_h(T)}{P_e(T)}]\\
&= \frac{\epsilon_c+\epsilon_v}{2}+\frac{k_BT}{2}\ln\Big[\frac{\sum_{\sigma=1}^{n_h}m_{h,\sigma}^{\frac{3}{2}}}{\sum_{\sigma=1}^{n_e}m_{e,\sigma}^{\frac{3}{2}}}\Big]
\end{aligned}$$

The chemical potential at zero temperature is in the middle of the band gap but at somewhat larger temperatures it changes linearly with temperature towards that band with the smaller density of states in order to enforce charge neutrality. The chemical potential at zero Kelvin is usually called the Fermi Level or Fermi niveau.

8.6.2 Extrinsic semiconductor

Now we consider dopant atoms. If the carrier concentration is dominated by the presence of dopant atoms, the semiconductor is called **extrinsic**

Dopant atoms are atoms that introduce shallow levels, that is, levels that are either close to the valence band edge or close to the conduction band edge. Dopant atoms are usually atoms with a slightly increased or reduced atomic number of valence electrons compared to those of the host crystal. In the case of silicon typical dopant atoms would be phosphorous and boron.

In silicon four electrons are consumed by bond formation. Phosphorous, however, has five valence electrons. The surplus electron of phosphorous is donated to the valence band. This is the reason why phosphorous is called a **shallow donor**. The phosphorous left behind is positively charged. The surplus electron is attracted to the positive phosphorous atom and will form a bound state, which is similar to the ground state of hydrogen atom, with the main difference that we need to use the effective mass of the conduction electron and the dielectric constant of silicon.

Similarly a boron atom has only three valence electrons. Thus it captures an electron and thus creates a hole. this is why boron is called a **shallow acceptor**.

The binding energy of an electron in a hydrogen atom is

$$\epsilon=-\frac{1}{2}\frac{m_e e^4}{(4\pi\epsilon_0\hbar)^2}$$

For a dopant we need to replace the mass by the effective mass and the dielectric constant by the dielectric constant of silicon.

The longitudinal electron effective mass of silicon is 0.19 m_e and the transversal effective mass is 0.98 m_e. The light hole mass is 0.16 m_e and the heavy hole mass is 0.49 m_e. The relative dielectric constant is $\epsilon_r = 11.9$. Thus we estimate the binding energy for electrons and holes as

$$E_B(electrons) = -\frac{1}{2}\frac{m^*/m_e}{\epsilon_r}\frac{m_e e^4}{(4\pi\epsilon_0\hbar)^2} = 0.5 * \frac{0.19}{11.9^2}H = 18.25 - 94.16\ meV$$
$$E_B(holes) == -\frac{1}{2}\frac{m^*/m_e}{\epsilon_r}\frac{m_e e^4}{(4\pi\epsilon_0\hbar)^2} = 0.5\frac{0.16}{11.9^2}H = 15.37 - 47.08\ meV$$

Since the effective mass is much smaller than that of a free electron and the dielectric constant is much larger, the binding energy is extremely small, namely in the range of 10 meV. (The band gap of silicon is 1.1 eV wide.) This binding energy can easily be overcome by thermal energies of by an electric field, so that we can consider the additional electrons and holes as free charge carriers.

Dopant	Valence	Type	Energy level
B	3-valent	acceptor	$\epsilon = \epsilon_v + 0.046$ eV
P	5-valent	donor	$\epsilon = \epsilon_c - 0.044$ eV

Table 8.1: Dopant atoms in silicon

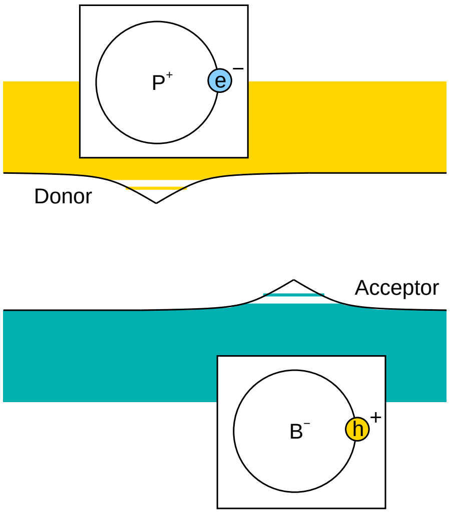


Fig. 8.3: Schematic representation of the band structure of a semiconductor with dopant atoms. The energy is shown as function of a spatial coordinate. The insets show the Bohr representation of the corresponding defects.

Our goal is now to determine the concentration of charge carriers, that is of electrons and holes. Together with the mobility the number of carriers determines the conductivity of the material.

Before we begin with the derivation, let me define a useful quantity, the intrinsic carrier concentration n_i.

Definition 8.1 *INTRINSIC CARRIER CONCENTRATION*

$$n_i \overset{def}{=} \sqrt{n_e n_h} \overset{Eqs.\ 8.21,8.22}{=} \sqrt{P_e(T)P_h(T)}\mathrm{e}^{-\beta E_g/2} \tag{8.24}$$

Interestingly the intrinsic carrier concentration is independent of the chemical potential.

Like in the intrinsic case also in the extrinsic semiconductor, the carrier concentrations are given by Eqs. 8.21,8.22. However, we need to determine the chemical potential under the influence of the dopants. The chemical potential is determined by the charge neutrality condition

$$-en_e + en_h + eN_D - eN_A = 0 \tag{8.25}$$

where e is the elementary charge, n_e and n_h are the electron and hole densities and N_D and N_A are the densities of donor and acceptor atoms. A donor atom contributes a positive charge, because a donor atom, being five-valent, gives away one electron into the conduction band, leaving behind a positive ion. Similarly an acceptor atom leaves behind a positive ion.

From the charge neutrality condition Eq. 8.25, we obtain

$$n_e - \underbrace{n_h}_{n_i^2/n_e} - (N_D - N_A) = 0$$

We express the hole density by the intrinsic carrier concentration Eq. 8.24 and the electron density Eq. 8.21.

$$\begin{aligned} n_e^2 - (N_D - N_A)n_e - n_i^2 = 0 \\ \Rightarrow n_e = \frac{N_D - N_A}{2} \pm \sqrt{\left(\frac{N_D - N_A}{2}\right)^2 + n_i^2} \end{aligned} \tag{8.26}$$

One of the two solutions can be excluded because it leads to a negative electron concentration. This leads us to the final result for the electron concentration.

$$n_e = \frac{N_D - N_A}{2} + \sqrt{\left(\frac{N_D - N_A}{2}\right)^2 + n_i^2} \tag{8.27}$$

We obtain the hole concentration again from the charge neutrality condition Eq. 8.25

$$n_h \overset{Eq.\ 8.25}{=} n_e - (N_D - N_A) \overset{Eq.\ 8.27}{=} -\frac{N_D - N_A}{2} + \sqrt{\left(\frac{N_D - N_A}{2}\right)^2 + n_i^2} \tag{8.28}$$

This, together with the expression for the electron concentration Eq. 8.27, is our final result. Note that the intrinsic carrier concentration can be determined without the knowledge of the chemical potential.

Note that unexpectedly, it was not necessary to determine the chemical potential in order to work out the carrier concentrations. If desired, however, we can also determine the chemical potential as follows:

$$\begin{aligned} n_e &\overset{Eq.\ 8.21}{=} P_e(T)\mathrm{e}^{-\beta(\epsilon_c - \mu)} \overset{Eq.\ 8.27}{=} \frac{N_D - N_A}{2} + \sqrt{\left(\frac{N_D - N_A}{2}\right)^2 + n_i^2} \\ \Rightarrow \mu &= \epsilon_c + k_B T \ln\left[\frac{\frac{N_D - N_A}{2} + \sqrt{\left(\frac{N_D - N_A}{2}\right)^2 + n_i^2}}{P_e(T)}\right] \end{aligned}$$

This expression is suitable if the donors dominate over the acceptors and if furthermore, the intrinsic carrier concentration is small compared to the donor concentration.

If the intrinsic carrier concentration dominates over the dopant concentration, we can express the chemical potential in a more convenient form.

$$\begin{aligned}
\mu &= \epsilon_c + k_BT\ln\left[\frac{N_D-N_A}{2n_i}+\sqrt{1+\left(\frac{N_D-N_A}{2n_i}\right)^2}\right]+k_BT\ln\left[\frac{n_i}{P_e}\right] \\
&\overset{Eq.\ 8.24}{=} \epsilon_c + k_BT\ln\left[\frac{N_D-N_A}{2n_i}+\sqrt{1+\left(\frac{N_D-N_A}{2n_i}\right)^2}\right]+k_BT\ln\left[\sqrt{\frac{P_h}{P_e}}\right]-\frac{1}{2}E_g \\
&= \underbrace{\frac{\epsilon_c+\epsilon_v}{2}+k_BT\ln\left[\sqrt{\frac{P_h}{P_e}}\right]}_{\mu_{intrinsic}} + k_BT\ln\left[\frac{N_D-N_A}{2n_i}+\sqrt{1+\left(\frac{N_D-N_A}{2n_i}\right)^2}\right]
\end{aligned}$$

Let us rest here and discuss our finding, namely Eq. 8.27 and Eq. 8.28:

- when the intrinsic carrier concentration is negligible compared to the dopant concentration, or more precisely compared to $|N_D - N_A|$, we obtain that only one of the carriers, either electrons and holes, are present. Their concentration exactly compensates the charge of the dopants. This is the case in the low temperature limit.

  When the donors dominate, the electron concentration equals $n_e = N_D - N_A$ and the hole concentration vanishes. When the acceptors dominate, the hole concentration equals $n_h = N_A - N_D$ and the electron concentration vanishes.

  Note however that the basic assumption of the expression for the carrier concentration, namely that the chemical potential lies far from the band edges (compared to k_BT) is violated in this case. Nevertheless the final conclusion remains valid.

- when the intrinsic carrier concentration dominates, which is the case at high temperatures, we can neglect the dopant concentration. Thus we obtain the high-temperature result of the intrinsic semiconductor, namely

  $$n_e = n_h = n_i$$

  which is the charge neutrality condition of the intrinsic semiconductor.

8.6.3 Carrier depletion in a p-n junction

The function of a transistor is to switch a electric current from source to drain on and off, by applying a small voltage to the gate. This function is achieved by properly combining diodes. A diode consists of two semiconductors with opposite doping. It is therefore also called a p-n junction. A semiconductor is called n-type if it doped by donors and therefore conducts by electrons and it is called p-doped when it is doped by acceptors and therefore conducts via holes. (n stands for negative and p stands for positive carriers.)

In and p-n junction the carrier distribution is no more homogeneous. Electrons and holes flow across the interface until equilibrium is achieved and the chemical potential is the same everywhere in the sample. As charges are moved about electric potentials build up.

The problem is described by the condition of local equilibrium, Poisson equation and global charge neutrality:

$$\begin{aligned}
\rho(x) &= -eP_e(T,x)\mathrm{e}^{-\beta(\epsilon_c-e\Phi(x)-\mu)}+eP_h(T,x)\mathrm{e}^{-\beta(\mu-\epsilon_v+e\Phi(x))} \\
&\quad +eN_D(x)-eN_A(x) \\
\nabla^2\Phi(x) &= -\frac{\rho}{\epsilon_0\epsilon_r} \\
\int_{-\infty}^{\infty} dx\rho(x) &= 0
\end{aligned}$$

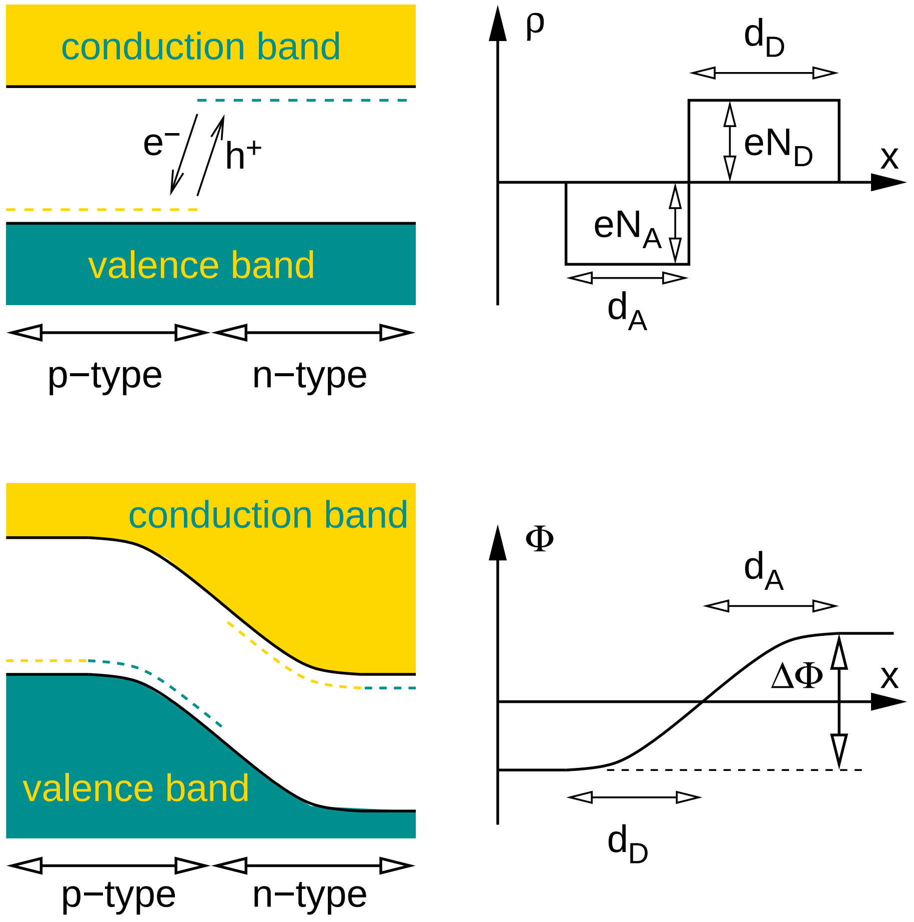


Fig. 8.4: Illustration of the dipole at a p-n junction. Upper left: band diagram without dipole. Electrons in the n-type material on the right-hand side lower their energy, by occupying the acceptors in the p-type material on the left-hand side. Upper right: charge dipole layer. Lower left: electrostatic potential of dipole layer. Lower right: band diagram with dipole. The Fermi level is constant throughout the p-n junction. The band edges are bent due to the electrostatic potential resulting from the dipole layer. All states above the Fermi level are occupied and all states below the Fermi level are empty.

The number of charge carriers is defined locally by thermal equilibrium for a given value of the chemical potential. The electrons have the charge $-e$ and holes have the charge the charge $+e$. We ignore the binding energy of the shallow donors and acceptors, which is valid at sufficiently high temperatures, that is if k_BT is larger than the binding energy. This allows us to count the electrons located in the donors levels to the conduction band and the number of holes in the acceptor levels to the valence band.

If we start with a vanishing electrical potential, electrons will flow from the n-type region to the p-type region and holes will cross the interface in the opposite region. The charge flow will set up a charge dipole, which consists of a **depletion region**, a region with fully ionized dopants and no charge carriers. The charge dipole produces a step-like potential that shifts the band structure of the p-type and n-type regions relative to each other until they are aligned such that the charge neutrality level for each semiconductor matches with the global chemical potential. The charge neutrality level is that position of the chemical potential that would result in *local* charge neutrality.

In order to simplify the problem we make the assumption that the transition region from a neutral semiconductor to a fully ionized situation is much abrupt compared to the extent of the depletion region. In that case, the density is simply a step function

$$\rho(x)=\begin{cases}0 & \text{for} \quad x<-d_A\\ -eN_A & \text{for} \quad -d_A<x<0\\ +eN_D & \text{for} \quad 0<x<d_D\\ 0 & \text{for} \quad d_D<x\end{cases}$$

We assumed that the interface is located at $x=0$, with the p-type material on the left and the n-type material on the right.

The relative sizes of the depletion regions are determined by the global charge neutrality condition.

$$N_Ad_A=N_Dd_D \Rightarrow d_D=\frac{N_A}{N_D}d_A$$

Now we evaluate the Poisson equation by integrating the Poisson equation twice

$$\Phi(x)=\begin{array}{ll}\Phi(-\infty) & \text{for} \quad x<-d_A\\ \Phi(-\infty)+\frac{eN_A}{2\epsilon_0\epsilon_r}(x+d_A)^2 & \text{for} \quad -d_A<x<0\\ \Phi(+\infty)-\frac{eN_D}{2\epsilon_0\epsilon_r}(x-d_D)^2 \quad \text{for} \quad 0<x<d_D & \\ e\Phi(+\infty) & \text{for} \quad d_D<x\end{array}$$

The potential step is obtained from matching the solutions at $x=0$.

$$\begin{aligned}\Delta\Phi=\Phi(+\infty)-\Phi(-\infty)&=\frac{eN_A}{2\epsilon_0\epsilon_r}d_A^2+\frac{eN_D}{2\epsilon_0\epsilon_r}d_D^2\\ &=\frac{e}{2\epsilon_0\epsilon_r}\Big(N_Ad_A^2+N_Dd_D^2\Big)=\frac{e}{2\epsilon_0\epsilon_r}\Big(N_A+N_D(\frac{N_A}{N_D})^2\Big)d_A^2\\ &=\frac{e}{2\epsilon_0\epsilon_r}\frac{N_A}{N_D}(N_D+N_A)d_A^2\end{aligned}$$

Hence the size of the depletion layer is given as a function of the potential step as

$$d_A=\sqrt{\frac{2\epsilon_0\epsilon_r}{e}\frac{N_D/N_A}{N_A+N_D}\Delta\Phi}$$

$$\begin{aligned}d_A+d_D&=(1+\frac{N_A}{N_D})\sqrt{\frac{2\epsilon_0\epsilon_r}{e}\frac{N_D/N_A}{N_A+N_D}\Delta\Phi}\\ &=\sqrt{\frac{2\epsilon_0\epsilon_r}{e}\frac{N_A+N_D}{N_AN_D}\Delta\Phi}\end{aligned}$$

Without applied potential the potential step must compensate exactly the difference of the charge neutrality levels. At low temperatures and small binding energies this is just the band gap E_g. What happens if we apply an electric field? In that case the charge neutrality levels of the two semiconductors are themselves shifted by the voltage. The thickness of the depletion layer at an applied voltage is obtained by setting $\Delta\Phi = E_g + eV$ and obtain

$$\begin{aligned} d_A + d_D &= \sqrt{\frac{2\epsilon_0\epsilon_r}{e}\frac{N_A+N_D}{N_DN_A}(E_g - eV)} \\ &= \Big(d_A(V=0) + d_D(V=0)\Big)\sqrt{1-\frac{eV}{E_g}} \end{aligned}$$

If the the voltage just offsets the band gap, that is $eV = E_g$, the depletion layer vanishes, and remains zero if the potential is increased further.

Due to its lack of charge carriers, the depletion layer is a barrier to electric current. Thus the thickness of the depletion layer is a measure for the current, which allows us to estimate a current voltage characteristics. It shows that the p-n-junction acts as a one-way street for electrons.

A transistor is build from two p-n junctions pointing towards each other. Thus we have a p-n-p junction or a n-p-n junction. The two outer materials are called **source** and **drain**, and the oppositely doped semiconductor in the middle is called the **channel**. In such a structure at least one p-n junction blocks the voltage from source to drain. This is the off-state of a transistor. To switch the current on, we can apply a voltage to the channel, so that both junctions become conducting. If a voltage is applied between source and drain, a current can flow. The voltage is applied to the **gate** of the transistor, which is a metal contact separated from the channel by a small insulating oxide, the gate oxide.

The structure gate/ gate oxide/ channel can be thought of a capacitor, which induces the required charge carriers in the channel, if a voltage is applied between gate and channel. Transistors build in thus way are called MOSFETs Metal-oxide-semiconductor field-effect transistors. Another design, the bipolar transistor, which is described in most textbooks, has been used mostly until the early nineties.

8.7 The ideal Boson Gas

Bosons are particles which have a wave function that is symmetric under particle exchange. Here we have no limitation on the number of particles that can occupy a single one-particle state. We proceed analogously to the Fermion case, but now sum over all occupation numbers from zero to infinity.

For Bosons we obtain:

$$\begin{aligned} F(T,\mu) &= -k_BT\ln Z(T,\mu) \\ &= \int d\epsilon\; D(\epsilon)\Big[-k_BT\ln[\sum_{n_i=0}^{\infty} \mathrm{e}^{-\beta(\epsilon-\mu)n_i}]\Big] \\ &= \int d\epsilon D(\epsilon)\;\Big[-k_BT\ln[\frac{1}{1-\mathrm{e}^{-\beta(\epsilon-\mu)}}]\Big] \\ &= \int d\epsilon D(\epsilon)\;\Big[k_BT\ln[1-\mathrm{e}^{-\beta(\epsilon-\mu)}]\Big] \end{aligned}$$

Note here that the expression for bosons is, apart from two sign changes, identical to the expression for Fermions.

The entropy is obtained as

$$S = -\frac{\partial F}{\partial T} = \int d\epsilon D(\epsilon)\Big[-k_B\ln[1-\mathrm{e}^{-\beta(\epsilon-\mu)}] - k_B\frac{\beta(\epsilon-\mu)\mathrm{e}^{-\beta(\epsilon-\mu)}}{1-\mathrm{e}^{-\beta(\epsilon-\mu)}}\Big]$$

The internal energy is

$$U = F + TS + \mu N = \int d\epsilon\ D(\epsilon)\epsilon \frac{1}{e^{\beta(\epsilon-\mu)} - 1}$$

Here we introduce the **Bose distribution**

$$b(\epsilon) = \frac{1}{e^{\beta(\epsilon-\mu)} - 1} \tag{8.29}$$

Using the occupation numbers $b(\epsilon)$ from Eq. 8.29, we obtain

$$\begin{aligned}
U[b(\epsilon)] &= \int d\epsilon D(\epsilon) \cdot \epsilon\ b(\epsilon) \\
F[b(\epsilon)] &= \int d\epsilon D(\epsilon) \cdot \Big(-k_B T \ln[1 + b(\epsilon)]\Big) \\
S[b(\epsilon)] &= \int d\epsilon D(\epsilon) \cdot k_B \Big[(1 + b(\epsilon)) \ln[1 + b(\epsilon)] - k_B b(\epsilon) \ln[b(\epsilon)]\Big]
\end{aligned}$$

Let us look at the Bose distribution function: For $\epsilon >> \mu + k_B T$ we obtain $f \approx e^{-\beta(\epsilon-\mu)}$. For $\epsilon - \mu << k_B T$ we obtain $f \approx k_B T/(\epsilon - \mu)$. From the pole at $\epsilon = \mu$ we can immediately tell that the chemical potential must always be negative, in order to avoid an infinite occupation.

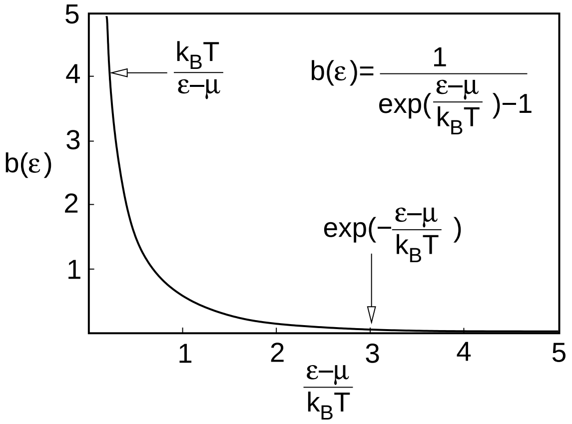


Fig. 8.5: Bose distribution function as function of $(\epsilon - \mu)/k_B T$.

8.7.1 Bose-Einstein Condensation

Bosons exhibit a phase transition at low temperatures, which is called **Bose-Einstein condensation**. This phase transition is unusual, because noninteracting systems usually do not undergo phase transitions.

Bosons can condense at very low energy into a coherent macroscopic quantum state, where a macroscopic number of particles occupies the ground state of the one-particle spectrum. A well-known example for a macroscopic quantum state is laser-light, even though this state is not the result of a Bose-Einstein condensation.

An example for a system undergoing a Bose-Einstein condensation is ^{4}He.[7] ^{4}He contains an even number of Fermions, two protons and two neutrons and two electrons. Hence it has even spin and

---

[7] The superscript in ^{4}He indicates the number of nucleons. ^{4}He consists of an even number of fermions, namely 2 protons, 2 neutrons and two electrons. On the other hand ^{3}He is itself a Fermion, because it consists of an odd number of Fermions, namely 2 protons, 1 neutron and two electrons.

is a boson. At a temperature of 2.17 K it undergoes a Bose-Einstein condensation and it becomes superfluid. A **superfluid** has zerio viscosity, zero entropy, and infinite thermal conductivity. The most striking observation of a superfluid is that it leaks out of every container. It forms a thin so-called **Rollin film** at the walls, which allows it to creap up the walls even against the gravitational force. The superfluid state does not exhibit any friction. This reminds of superconductor, where the resistivity drops to zero. In the latter case electrons pair up forming Cooper pairs. The latter are bosons and can condense into a kind of Bose-Einstein condensate.

Let us now dig into the origin of the Bose-Einstein condensation. Let us consider the occupation of one-particle states for bosons with a conserved particle number.[8]

$$N=\int d\epsilon\; D(\epsilon)b(\epsilon)=\int d\epsilon\; D(\epsilon)\frac{1}{\mathrm{e}^{\beta(\epsilon-\mu)}-1}$$

Now we need to adjust the chemical potential until the correct number of particles is obtained. We will see that this can only be done above a specific temperature, which will be identified with the transition temperature for the Bose Einstein condensation.

The occupation number has a pole at $\epsilon=\mu$.

$$\frac{1}{\mathrm{e}^{\beta(\epsilon-\mu)}-1}=\frac{k_BT}{\epsilon-\mu}+O(\epsilon-\mu)^2$$

Since the particle number is finite, the chemical potential must lie in a region of vanishing density of states, that is equal or below zero. The particle number increases with increasing chemical potential. As we reduce the temperature, the chemical potential must approach zero from below, so that the number of particles remains constant.

An upper bound for the number of particles is obtained with a chemical potential of zero.

$$\begin{aligned}
N\le N(T,\mu=0)&=\frac{2\pi V}{(2\pi\hbar)^3}(2m)^{\frac{3}{2}}\int d\epsilon\;\frac{\sqrt{\epsilon}}{\mathrm{e}^{\beta\epsilon}-1}\\
&=\frac{2\pi V}{(2\pi\hbar)^3}(2m)^{\frac{3}{2}}(k_BT)^{\frac{3}{2}}\int dx\;\frac{\sqrt{x}}{\mathrm{e}^{x}-1}\\
&=V\Big(\frac{mk_BT}{2\pi\hbar^2}\Big)^{\frac{3}{2}}\underbrace{\frac{2}{\sqrt{\pi}}\int dx\;\frac{\sqrt{x}}{\mathrm{e}^{x}-1}}_{C}\\
&=CV\Big(\frac{mk_BT}{2\pi\hbar^2}\Big)^{\frac{3}{2}}=C\frac{V}{\lambda_T^3} \qquad (8.30)\\
C&=\frac{2}{\sqrt{\pi}}\int dx\frac{\sqrt{x}}{\mathrm{e}^{x}-1}\approx 2.612
\end{aligned}$$

λ_T is the thermal de-Broglie wavelength. The constant $C=\zeta(\frac{3}{2})=2.612$ is related to the Riemann zeta function $\zeta(x)$. See Book by Agarwal and Eisner[16] p.119 for more details.

---

[8] There is no conservation law for the particle number of photons. Photons do not undergo a Bose-Einstein condensation, which is because of the missing particle number conservation.

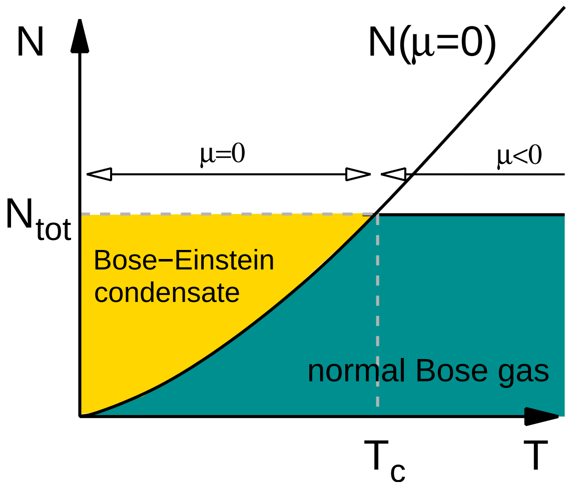


Fig. 8.6: Demonstration of the number of particles in the normal state (dark region) and the condensate (yellow region) as function of temperature. N_0 is the total number of particles in the system. $N(\mu = 0)$ shows the maximum number of particles that can be accomodated in the normal state. Below the critical temperature T_c the chemical potential is pinned at the ground state energy.

The consequences of Eq. 8.30 are demonstrated in Fig. 8.6: The maximum particle number in the continuum of states decreases with $T^{\frac{3}{2}}$. At a certain temperature T_c the chemical potential approaches zero, because only then all particles can be described by the integral given above. Above the critical temperature the chemical potential shifts to negative values in order to place exactly the right number of states into the continuum of states. Below the critical temperature, however, it seems as if we cannot place all particles into the available states. This seems to be contradictory. Where did the remaining particles go?

The origin for the problem is that the integral for the number of particles is discontinuous at $\mu = 0$. For positive μ the integral is infinite, and below the integral has a finite value that drops to zero as the temperature vanishes. The integral therefore has an undefined value, and the correct result depends on the precise fashion we approach the value $\mu = 0$. The problem is that we approached two limits in the wrong order. So far we have performed the thermodynamic limit ($V \to \infty$) first, which resulted in a continuous density of states, and then we approached the low-temperature limit $\mu \to 0$. This order has to be reversed.

Let us consider the physical situation at $T = 0$. At zero temperature, all bosons are in the lowest one-particle state. This many-particle state is the bosonic ground state. As long as the system is finite, that is as long as we are not in the thermodynamic limit $V \to \infty$, the energy of the ground state ϵ_0 of the one-particle spectrum will be finite and positive. As long as the first excitation energy is much larger than thermal energies $\epsilon_1 - \epsilon_0 >> k_BT$, and we can determine the chemical potential from the ground-state energy alone.

$$N \overset{k_BT<<\epsilon_1-\epsilon_0}{\approx} b(\epsilon_0) = \frac{1}{e^{\beta(\epsilon_0-\mu)} - 1}$$

$$\Rightarrow \qquad \mu \overset{T\to 0}{\to} \epsilon_0 - k_BT\ln(1+\frac{1}{N}) \overset{N\to\infty}{\to} \epsilon_0 - \frac{k_BT}{N}$$

Thus it is always possible to place all particle into the available states. The chemical potential, however, becomes positive, even though only slightly.[9] Had we already done the thermodynamic limit and worked with the continuous density of states a positive chemical potential would immediately lead to an infinite particle number, because the pole is not integrable, and to negative occupations.

---

[9] The chemical potential becomes positive, when thermal energies k_BT are smaller than the ground state energy $E_0 = N\epsilon_0$ of the many-particle system.

Let us now heat the system up just a tiny little bit, so that only a very small number of particles get excited. This implies that the first excitation energy $\epsilon_1 - \epsilon_0$ is still much larger than thermal energies, so that $\mathrm{e}^{\beta(\epsilon_i-\mu)} >> 1$ except for the ground state, i.e. $i \neq 0$. We obtain

$$\begin{aligned}
N &= \frac{1}{\mathrm{e}^{\beta(\epsilon_0-\mu)}-1} + \sum_{i\neq 0} \frac{1}{\mathrm{e}^{\beta(\epsilon_i-\mu)}-1} \\
&\approx \frac{1}{\mathrm{e}^{\beta(\epsilon_0-\mu)}-1} + \sum_{i\neq 0} \mathrm{e}^{-\beta(\epsilon_i-\mu)} \\
&= \frac{1}{\mathrm{e}^{\beta(\epsilon_0-\mu)}-1} + \mathrm{e}^{\beta\mu} \sum_{i\neq 0} \mathrm{e}^{-\beta\epsilon_i} \\
&= \frac{1}{\mathrm{e}^{\beta(\epsilon_0-\mu)}-1} + \mathrm{e}^{\beta\mu} \int_{\epsilon>0} d\epsilon\; D(\epsilon)\mathrm{e}^{-\beta\epsilon} \\
&\approx \frac{1}{\mathrm{e}^{\beta(\epsilon_0-\mu)}-1} + \underbrace{\mathrm{e}^{\beta\mu}}_{\to 1} \int_{\epsilon>0} d\epsilon\; D(\epsilon)\underbrace{\frac{1}{\mathrm{e}^{\beta\epsilon}-1}}_{b(\epsilon,\mu=0)} \\
&\overset{\text{Eq. 8.30}}{=} \underbrace{\frac{1}{\mathrm{e}^{\beta(\epsilon_0-\mu)}-1}}_{N_0(t)} + \underbrace{CV\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}}_{N_{norm}(T)}
\end{aligned} \tag{8.31}$$

We have divided the system into two subsystems: One subsystem is the Bose condensate: The Bose condensate describes all particles that reside in the ground state of the one particle spectrum. The Bose condensate contains $N_0(T)$ particles. The remaining particles form a normal Bose gas with $N_{norm}(T)$ particles. The normal Bose gas is formed by all particles in excited one-particle states. We have removed the factor $\mathrm{e}^{beta\mu}$ for the normal state, because it approaches one in the thermodynamic limit, which now can be done for the normal state without difficulty[10] . This the so-called **two-liquid model**.

In our approximate expression Eq. 8.31 the chemical potential is adjusted such that the number of particles in the Bose condensate just compensates the number of particles that cannot be placed into the continuum. If we denote the number of particles in the ground state as $N_0(\mu, T)$ and the number of particles in the continuum by $N_{norm}(\mu, T)$, we obtain

$$\begin{aligned}
&N_0(T) = N_{tot}(1-(\frac{T}{T_c})^{\frac{3}{2}}) \quad \text{for} \quad T < T_c \\
&N_0(T) = 0 \quad \text{for} \quad T > T_c
\end{aligned}$$

The number of states in the ground state is set to zero above the critical temperature, because then the integral is well defined since the chemical potential differs from zero.

What is the critical temperature?

$$\begin{aligned}
\frac{N_{tot}}{V} &= C\Big(\frac{mk_BT_c}{2\pi\hbar^2}\Big)^{\frac{3}{2}} \\
\Rightarrow \qquad T_c &= \frac{2\pi\hbar^2}{mk_B}\Big(\frac{N}{CV}\Big)^{\frac{2}{3}}
\end{aligned}$$

Since liquid ^{4}He has a given density at low temperature this defines a unique transition temperature about 1.74 K, which is only slightly pressure dependent.

What is the ground state? Its wave function is the simple product of single particle wave functions. Because all contributing one-particle orbitals are identical, the wave function is automatically

---

[10] In the dilute limit, the Bose gas becomes identical to the Boltzmann gas. Because the Boltzmann factor does not have a singilarity, we can perform the thermodynamic limit for all states except the ground state and replace the density of states of the finite system with the continuous density of states of the infinite system. Note, that the ground state has zero weight in the integral for the normal state.

symmetrized.

$$\Psi(\vec{r}_1, \dots, \vec{r}_N) = \phi_0(\vec{r}_1) \dots \phi_0(\vec{r}_N)$$

Since there is a macroscopic number of particles in a single one-particle state. Thus we are able to observe the behavior of a quantum mechanical wave function directly.

Particles in the ground state, that is the Bose-Einstein condensate, flow without friction. If a particle scatters, it is scattered out of the condensate into the continuum. The ground state wave function, however, is unaffected. The particle that is lost will be refilled, as the system approaches thermal equilibrium. This is the origin of **superfluidity**. The Bose Einstein condensate and the normal fluid (the continuum states) coexist. One of them flows without dissipation and the other behaves like a regular fluid with friction. As we lower the temperature more of the normal fluid is converted into the supra-fluid phase.

8.7.2 Application: Black-Body radiation

Consider a metallic box at a given temperature. The walls of the box are in thermal equilibrium with the photons inside. The spectral intensity distribution of photons in thermal equilibrium is completely independent of the shape of the box or the material of its walls. It is simply a property of the photon field. The intensity distribution of the electromagnetic radiation therefore only depends only on the temperature. We can use the spectral distribution of the light coming out of a black body, to determine the temperature of the box.

This method would provide us with a simple means to measure the temperature right at the surface of the sun. Even further one can determine the temperature at the surface of a star, by analysing the light emmitted by it. Quite spectacularly we can make conclusions on the expansion of the universe, by comparing the background radiation observed today with the temperature at which the universe became transparent for the first time. Initially electrons and protons formed a plasma, which is not transparent. Only when electrons and protons formed hydrogen atoms with a sufficiently low concentration of free electrons, at about 3000 K, light could propagate freely, and it does so until today. However the red shift of the expansion of the universe shifted the spectrum to a temperature of only 2.7 K.

Closer to earth, the spectal distribution also allows us to determine, without contact, the temperature in an oven that is so hot that inserting any apparatus for the measurement would cause serious materials problems.

We start out with the dispersion relation of the photons.

$$\omega(k) = c|k| \Rightarrow \epsilon(k) = \hbar c|k|$$

where c is the speed of light. The density of states for a linear dispersion relation is obtained from Eq. 8.13 on p. 195. That expression must be multiplied by two, in order to account for the two polarization directions of photons.

$$D(\epsilon) = 2 \cdot \frac{4\pi V}{(2\pi\hbar c)^3}\epsilon^2 \tag{8.32}$$

Our goal is the spectral intensity of the black-body radiation. The **intensity** I is the energy flux, and is given by the product of energy density and velocity.

$$I = c\frac{U}{V}$$

where U is the internal energy and V is the corresponding volume.

Let us therefore determine the internal energy as

$$U = \int d\epsilon \; D(\epsilon)\epsilon\frac{1}{e^{\beta\epsilon} - 1} = \int d\omega \; \underbrace{\hbar D(\hbar\omega)\hbar\omega\frac{1}{e^{\beta\hbar\omega} - 1}}_{u(\omega)}$$

Here we have set the chemical potential to zero, because there is no particle conservation law for photons. The photons can be created and absorbed by the walls.

$$\begin{aligned}
I &= \frac{c}{V}U = \int d\omega\, \frac{c}{V}\hbar D(\hbar\omega)\frac{\hbar\omega}{e^{\beta\hbar\omega}-1} \\
&\overset{\text{Eq. 8.32}}{=} \int d\omega\, \frac{c}{V}\hbar\cdot 2\cdot\frac{4\pi V}{(2\pi\hbar c)^3}\frac{(\hbar\omega)^3}{e^{\beta\hbar\omega}-1} \\
&= \int d\omega\, 8\pi\hbar c\left(\frac{k_BT}{2\pi\hbar c}\right)^3\frac{(\beta\hbar\omega)^3}{e^{\beta\hbar\omega}-1}
\end{aligned}$$

Thus we obtain **Planck's law**[17]

PLANCK'S RADIATION LAW

The spectral intensity distribution of a black body is

$$\frac{dI}{d\omega} = 8\pi\hbar c\left(\frac{k_BT}{2\pi\hbar c}\right)^3\frac{(\beta\hbar\omega)^3}{e^{\beta\hbar\omega}-1} \tag{8.33}$$

The intensity is the energy flux density. Note that $dI/d\omega$ is not a derivative of the intensity. The quantity is rather the contribution dI to the intensity from a window $d\omega$ of angular frequency. The intensity distribution in wave lengths is given in Eq. C.6 on p. 237.

The maximum of the function $x^3/(e^x-1)$, which determines the shape of the Intensity distribution Eq. 8.33, lies at ≈2.82. The frequency with maximum intensity changes proportionally with the temperature.

WIEN'S DISPLACEMENT LAW

Wien's displacement law states law that the maximum of the intensity distribution of a black body shifts linearly with the intensity. The proportionality of the photon energy at the intensity maximum in frequencies and the temperature is given as

$$\hbar\omega_{max} \approx 2.82k_BT \tag{8.34}$$

The maximum of the intensity distribution in wave-lengths is given in Eq. C.7 on p. 237. Note that the maxima of the intensity distributions in angular frequency and in wave length differ.

Gamov proposed on theoretical grounds that the universe should have a certain remaining **background radiation**, which is left over from the big bang. As the universe expands, this temperature has cooled to temperature of 2.7K, as has been measured by Penzias and Wilson in 1965 by exploiting the spectral density of the background radiation.

In the long-wavelength limit $\omega \to 0$ we obtain $u(\omega) \sim \omega^2$ is called the **law of Rayleigh and Jeans**. If this behavior is extrapolated to high frequencies, we find that the energy of the photon gas is infinite. This is the so-called **uv-catastrophe**. The paradox of the UV catastrophe was one of the motivations for quantum theory.

The law of Rayleigh Jeans had been derived on the basis of the equipartition theorem of classical mechanics, which says that each classical degree of freedom has an average kinetic energy of $\frac{1}{2}k_BT$. This implies that each harmonic oscillator has a mean energy of $E_{kin} + E_{pot} = k_BT$. The modes of the electromagnetic spectrum are just such oscillators.

Thus we obtain an internal energy of

$$U = \int d\omega\, \hbar D(\hbar\omega)\frac{1}{2}k_BT = \int d\omega\, \frac{4\pi V k_BT}{(2\pi\hbar c)^3}\hbar\omega^2$$

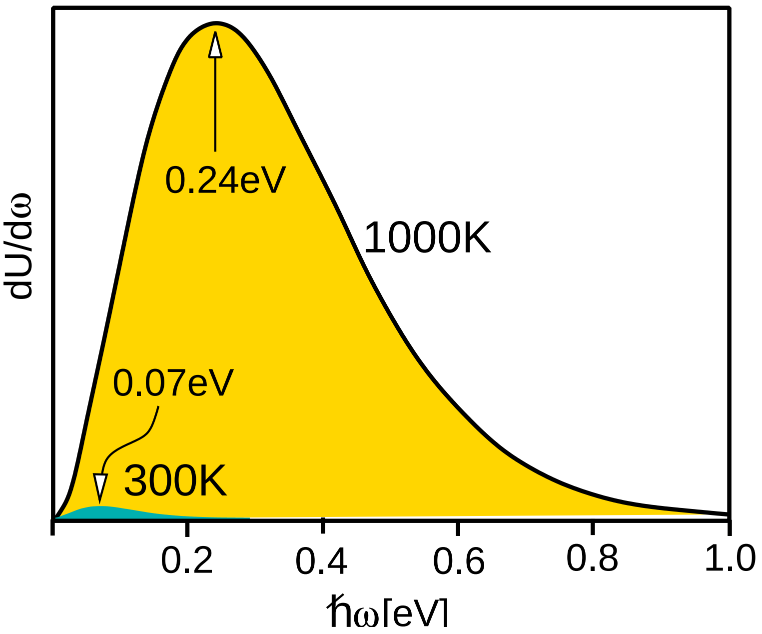


Fig. 8.7: Planck's Law. Spectral intensity for the blackbody radiation for 300 K and 1000 K

Clearly the result is infinite irrespective of temperature, which is a nonsensical result. Note that this infinity even occurs without considering the zero point energy of the phonons, which themselves would contribute another infinite energy. This energy however would not be directly observable.

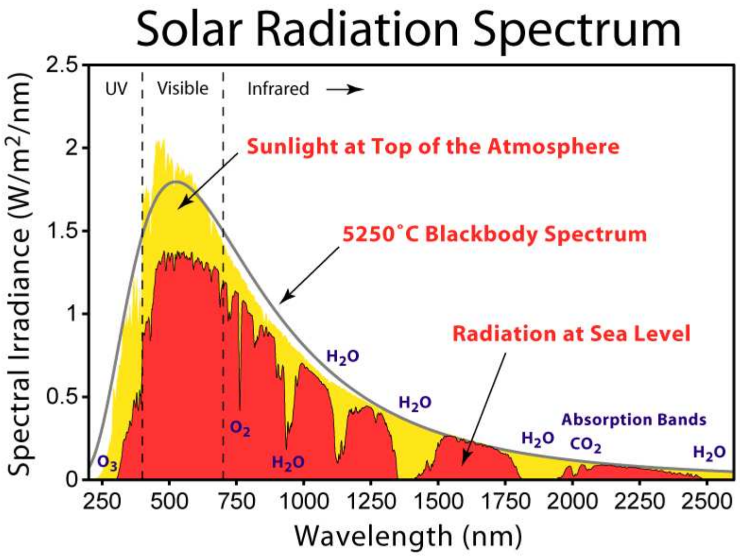


Fig. 8.8: Solar irradiance spectrum above atmosphere and at surface. The intensity is given here as function of wave length, so that the high-energy radiation is at the very left. Planck's radition law in terms of wave lengths is given in Eq. C.6. Image prepared by R.A. Rhode as part of the Global Warming Art project[18] .

8.8 Comparison of Bosons and Fermions

While Bosons and Fermions behave radically different, their thermodynamic expressions look very similar. This has the advantage that there is less to remember but it comes with the price that they are easily confused.

Bosons	**Fermions**
Free energy (Grand canonical potential)	Free energy (Grand canonical potential)
$F(T,\mu)=\int d\epsilon\; D(\epsilon)\Big[-k_BT\ln\Big(1+\mathrm{e}^{-\beta(\epsilon-\mu)}\Big)\Big]$	$F(T,\mu)=\int d\epsilon\; D(\epsilon)\Big[-k_BT\ln\Big(1+\mathrm{e}^{-\beta(\epsilon-\mu)}\Big)\Big]$
Bose distribution	Fermi distribution
$b(\epsilon,T,\mu)=\frac{1}{\mathrm{e}^{\beta(\epsilon-\mu)}-1}$	$f(\epsilon,T,\mu)=\frac{1}{\mathrm{e}^{\beta(\epsilon-\mu)}+1}$
Thermodynamic functions expressed by $b(\epsilon,\mu,T)$	Thermodynamic functions expressed by $f(\epsilon,\mu,T)$
$F=\int d\epsilon\; D(\epsilon)\Big[-k_BT\ln\Big(1+b(\epsilon)\Big)\Big]$	$F=\int d\epsilon\; D(\epsilon)\Big[+k_BT\ln\Big(1-f(\epsilon)\Big)\Big]$
$S=\int d\epsilon D(\epsilon)\Big[-k_B\Big(b\ln(b)-(1+b)\ln(1+b)\Big)\Big]$	$S=\int d\epsilon D(\epsilon)\Big[-k_B\Big(f\ln(f)+(1-f)\ln(1-f)\Big)\Big]$
$U=F+TS+\mu N=\int d\epsilon\; D(\epsilon)b(\epsilon)\epsilon$	$U=F+TS+\mu N=\int d\epsilon\; D(\epsilon)f(\epsilon)\epsilon$

The similarity is expressed by the fact that they become idential in the dilute high-temperature case. This implies that $\epsilon-\mu >> k_BT$ for all relevant energy levels. In that case both, Fermions and Bosons, behave like classical particles, and can be described like the Boltzmann gas. The explanation for this observation is straight-forward: If the particles never come close, they do not experience if they are identical with their neighbors or not.

If the temperature is high, the Boltzmann factor $\mathrm{e}^{\beta(\epsilon-\mu)}$ is small. In that case

$$b(\beta(\epsilon-\mu))\rightarrow \mathrm{e}^{-\beta(\epsilon-\mu)}\leftarrow f(\beta(\epsilon-\mu))$$

Up to first order in the Boltzmann factor we obtain the Free energy as

$$F(T,\mu)=\int d\epsilon\; D(\epsilon)\Big[-k_BT\mathrm{e}^{-\beta(\epsilon-\mu)}\Big]$$

From this expression we may derive the grand-canonical potential of the Boltzmann gas Eq. 2.11, which is is obtained by insertion of the density of states Eq. 8.12 into the above equation. Along the way we use the integral evaluated in a footnote.[11]

$$
\begin{aligned}
F &= \int d\epsilon\; D(\epsilon)\Big[-k_BT\mathrm{e}^{-\beta(\epsilon-\mu)}\Big] \\
&\overset{\text{Eq. 8.12}}{=} \int d\epsilon\; 2\pi V\left(\frac{\sqrt{2m}}{2\pi\hbar}\right)^3\sqrt{\epsilon}\Big[-k_BT\mathrm{e}^{-\beta(\epsilon-\mu)}\Big] \\
&= -k_BT\;2\pi V\left(\frac{\sqrt{2m}}{2\pi\hbar}\right)^3\int d\epsilon\;\sqrt{\epsilon}\;\mathrm{e}^{-\beta(\epsilon-\mu)} \\
&= -k_BT2\pi V\left(\frac{\sqrt{2m}}{2\pi\hbar}\right)^3(k_BT)^{\frac{3}{2}}\underbrace{\left\{\int_0^\infty dx\;\sqrt{x}\;\mathrm{e}^{-x}\right\}}_{\frac{1}{2}\sqrt{\pi}}\mathrm{e}^{\beta\mu} \\
&= -k_BT\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}V\mathrm{e}^{\beta\mu}
\end{aligned}
$$

which is identical to Eq. 8.12.

---

[11] We solve the integral by a variable substitution $x=y^2$with $dx=2ydy$ so that $\sqrt{x}dx=2y^2dy$.

$$\int_0^\infty dx\;\sqrt{x}\mathrm{e}^{-x}=2\int_0^\infty dy\;y^2\mathrm{e}^{-y^2}=2\int_0^\infty dy\;y^2\mathrm{e}^{-y^2}=\frac{1}{2}\sqrt{\pi}$$

The integral of $y^2\mathrm{e}^{-y^2}$ can be related to the integral over a three dimensional Gauss function

$$4\pi\int_0^\infty r^2\mathrm{e}^{-r^2}=\int_\infty d^3r\mathrm{e}^{-\vec{r}^2}=\left(\int_{-\infty}^\infty dx\;\mathrm{e}^{-x^2}\right)^3=\pi^{\frac{3}{2}}$$

so that

$$\int_0^\infty y^2\mathrm{e}^{-y^2}=\frac{1}{4}\sqrt{\pi}$$

The integral of a Gauss function must be known. It is not easily derived.

Appendix A

Probability and entropy

A.1 Entropy and number of yes-no questions

An interpretation for the entropy is that $S/(k_B \ln[2])$ is approximately the number of yes-no questions, that need to be asked to completely determine the state of the the system. The underlying reason is that the entropy of a yes-no question has a maximum value of $k_B \ln(2)$ and the slope for about equal probabilities vanishes. Thus, the answer to a yes-no question reduces the entropy by approximately $k_B \ln(2)$. Hence $S/k_B \ln(2)$ is approximately the number of yes-no questions that need to be asked in an optimum strategy.

The reader, who followed through the derivation of the entropy and who is comfortable with the above argument may skip the remainder of the section and pass on to the section on the maximum entropy principle. In the following I will try to familiarize the reader with the underlying arguments:

Let us imagine a system with $N = 2^n$ states. We divide all possible states into two groups with equal number of states and ask in which group of states the system is in. Now we repeat this procedure again with the group that contains the correct state, until only one state is left. After Q questions we know the state and therefore, we have reduced the entropy to zero. We obtain the initial entropy as $S = k \ln[2^n] = n(k \ln[2])$. Therefore $S/(k \ln[2]) = n$. Information theory regards the answer to a yes-now question with equal probabilities as the unit for information, and therefore the value of k is chosen as $k = 1/\ln[2]$ in that field. In thermodynamics the coefficient k is chosen to link an energy scale and a temperature scale.

Let us go through the exercise once more in more detail. This will lead to a more handwaving[1] definition of the entropy of a system with equal probability, namely $S = k_B \ln[N]$. Imagine we have a coin in one of N pots. Each of the pots has a number i between 1 and N. Now we need to identify the pot with the coin by a series of yes-no questions. We may ask questions of the type: Is the coin in one of the pots with a number $1 \leq i \leq m$? [2] If the answer is yes, we can restrict the search to those m points. If the answer is no, we can restrict the search to the $N - m$ pots with a number i larger than m. Once we have specified a group of pots that contain the coin, we again divide this group into two subsets by choosing another number m, ask if it is in one of the pots with number larger than m. By increasingly restricting the set of pots, we can single out the right pot.

Depending on the experiment, we may use more or less questions to find the right pot. Therefore, we average the number of questions with the probabilities, to find out an average number of question that we need to ask.

In the following we will work out the number $Q(N)$ of yes/no questions as function of the number N of events. The result is summarized in table. A.1 and it will be represented graphically in fig. A.1.

We denote the average number of questions required to find the coin in N pots by $Q(N)$. For

---

[1] A handwaving argument is one lacking rigor.

[2] More generally, we may specify a group of pots by specifying the numbers written on the pots in the group and ask if the coin is in one of the specified pots.

N	Number of questions $Q(N)$
1	0
2	1
3	$1+\frac{2}{3}Q(2)+\frac{1}{3}Q(1)=1+\frac{2}{3}=\frac{5}{3}$
4	$1+\frac{1}{2}Q(2)+\frac{1}{2}Q(2)=2$
5	$1+\frac{3}{5}Q(3)+\frac{2}{5}Q(2)=2+\frac{2}{5}=\frac{12}{5}$
6	$1+\frac{1}{2}Q(3)+\frac{1}{2}Q(3)=2+\frac{2}{3}=\frac{8}{3}$
7	$1+\frac{4}{7}Q(4)+\frac{3}{7}Q(3)=2+\frac{6}{7}=\frac{20}{7}$
8	$1+\frac{4}{8}Q(4)+\frac{4}{8}Q(4)=3$
9	$1+\frac{5}{9}Q(5)+\frac{4}{9}Q(4)=3+2/9=\frac{29}{9}$
10	$1+Q(5)=3+\frac{2}{5}=\frac{17}{5}$
N	$\min_i 1+\frac{i}{N}Q(i)+\frac{N-i}{N}Q(n-i)$

Table A.1: Determination of the minimum number $Q(N)$ of questions as function of the number N of possible outcomes.

$N=2$ we need only one question to determine the right pot. Therefore $Q(2)=1$. For N=3, we can only divide the pots only in a subset of two pots and the remaining one pot. Depending on the outcome of the first question, we may have to find the coin in the set of two pots or in just one pot. The probability that we are left with two pots is $\frac{2}{3}$ and the probability that we have just one pot is $\frac{1}{3}$. Therefore we need to ask two questions in two out of three cases and only one question in one out of three cases. Hence we obtain on average $\frac{2}{3}Q(2)+\frac{1}{3}Q(1)=2/3$ further questions. Adding $Q(2)=1$ for the first question we obtain $Q(3)=5/3$.

For $N=4$ there we can choose one of two different strategies. (1) We may select the first pot and ask if the coin is there is there. This yields $Q(4)=1+\frac{1}{4}Q(1)+\frac{3}{4}Q(3)=2+\frac{1}{4}$. (2) The second strategy is to select two pots in the first question. We are left with the choice of two pots regardless of the answer, which requires one additional question. This yields $Q(4)=1+\frac{1}{2}Q(2)+\frac{1}{2}Q(2)=2$. The second strategy is favorable, because it yields the correct answer on average with fewer questions. Since the rule is to always select the most efficient strategy, we select the second choice, that is $Q(4)=2$.

In general, the first question divides the pots into two groups, where one contains m pots and the other $N-m$ pots. Since we start with small problems we have already worked out the number of questions required to find the coin in each of the groups. The probability that the coin is in the group with m pots is $\frac{m}{N}$ and the probability that it is in the second group is $\frac{N-m}{N}$. Including the first equation we need $1+\frac{m}{N}Q(m)+\frac{N-m}{N}Q(N-m)$. The optimum strategy is the one that minimizes the number of questions, that is

$$Q(N)=\min_m\Big[1+\frac{m}{N}Q(i)+\frac{N-m}{N}Q(N-m)\Big]$$

The optimum strategy is to divide the pots always into two subsets, with nearly the same number of pots. For even N or large N, where $1/N$ can be ignored, we can choose $m=N/2$ and obtain the recursion relation.

$$Q(N)=1+\frac{1}{2}Q(N/2)+\frac{1}{2}Q(N/2)=1+Q(N/2)$$

The Ansatz[3] $Q(N) = c\ln[N]$ fulfills the equation, if $c = 1/\ln[2]$. Thus, we obtain

$$Q(N) \approx \frac{1}{\ln[2]}\ln[N]$$

The approximate sign indicates that this formula is only exact if N is a power of two. In Fig. A.1 we compare $Q(N)$ with $S(\frac{1}{N}, \ldots, \frac{1}{N})/(k\ln[2])$. We see that the result is surprisingly accurate even for small N. The individual numbers are always above the theoretical curve, because we could not always choose subsets with equal probabilities. Thus, the entropy reduction for most question has been less that $k\ln[2]$ so that we required more questions.

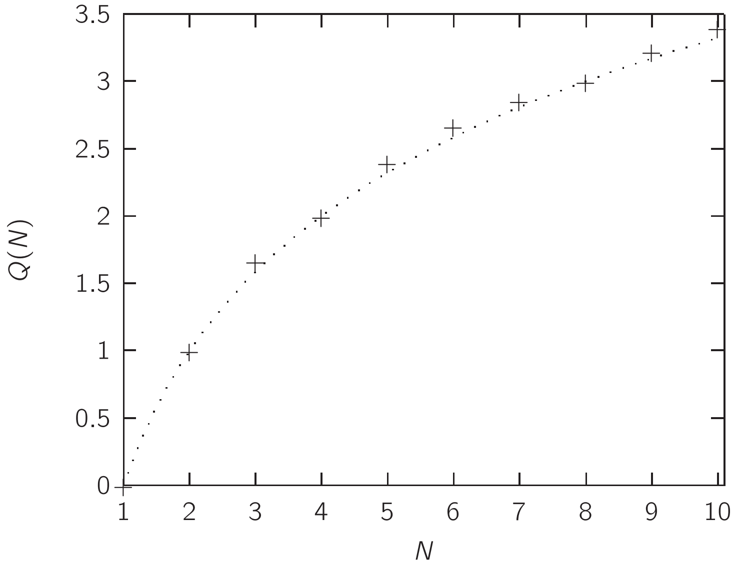


Fig. A.1: The number of questions versus the number of possibilities, required to determine a situation with equal a-priori probabilities. The full line is given by the theoretical expression $\ln[N]/\ln[2]$

If we multiply $Q(N)$ with the factor $k_B\ln[2]$ as specified before we obtain the expression for the entropy $S = k_B\ln[N]$.

---

[3] The idea that leads to this Ansatz is the following: One must know that $\ln[a \cdot b] = \ln[a] + \ln[b]$. Thus, $\ln[N/2] = \ln[N] - \ln[2]$. Thus, with this Ansatz $Q(N) = \ln[N]$, we expect to obtain an equation, namely $\ln[N] = 1 + \ln[N] - \ln[2]$, which we can try to resolve for $Q[N]$. Unfortunately, $\ln[N]$ drops out of that equation, and the result is always false. A remedy to that problem would be to change the Ansatz by multiplying the logarithm with a factor c. This factor would not also act on $\ln[2]$ so that we can adapt it to make the resulting equation true. Therefore we start with the Ansatz $Q[N] = c\ln[2]$.

Appendix B

Basic quantum mechanics

B.1 Position operator and position eigenstates

Let us assume that only know bra's and kets but no wave functions. In the following we will define wave functions from the abstract bracket notation. Let us denote the position eigenstates as $|x\rangle$. The position eigenstates is the link between kets and the corresponding wave functions. We require that the unity operator is expressed by the position eigenstates as

$$\hat{1} = \int_x |x\rangle\langle x| \tag{B.1}$$

Given a ket $|\psi\rangle$, the corresponding wave function $\psi(x)$ is defined by the scalar product

$$\psi(x) \overset{\text{def}}{=} \langle x|\psi\rangle$$

Given a wave function $\psi(x)$,we obtain the corresponding ket by

$$|\psi\rangle = \underbrace{\int dx\; |x\rangle\langle x|}_{=\hat{1}}\psi\rangle = \int dx\; |x\rangle\psi(x)$$

What is the wave function $\psi_{x_0}(x)$ corresponding to a position eigenstate $|x_0\rangle$? We obtain the solution by multiplication with the unit operator

$$|x_0\rangle = \int dx\; |x\rangle\; \underbrace{\langle x|x_0\rangle}_{=\psi_{x_0}(x)}$$

This equation can only be valid if

$$\psi_{x_0}(x) = \langle x|x_0\rangle = \delta(x - x_0)$$

Now we can define the position operator $\hat{x}$ by its eigenvalue equation

$$\hat{x}|x\rangle = |x\rangle x$$

Its explicit form in terms of position eigenstates is obtained as

$$\int dx\; \underbrace{|x\rangle x}_{=\hat{x}|x\rangle}\langle x| = \int dx \hat{x}|x\rangle\langle x| = \hat{x}\underbrace{\int dx|x\rangle\langle x|}_{=\hat{1}} = \hat{x}$$

$$\Rightarrow \hat{x} = \int dx\; |x\rangle x\langle x|$$

Momentum operator and momentum eigenstates

Next we define momentum eigenstates

$$\langle x|p\rangle = \mathrm{e}^{ipx/\hbar}$$

as eigenstates of the translation operator in real space.

Again, we start from the eigenfunctions to define the position operator

$$\frac{\hbar}{i}\partial_x\langle x|p\rangle = \frac{\hbar}{i}\partial_x \mathrm{e}^{ipx/\hbar} = p\langle x|p\rangle$$

$$\overset{\int dx\,|x\rangle\cdot}{\Rightarrow} \quad \underbrace{\int dx\,|x\rangle\frac{\hbar}{i}\partial_x\langle x|}_{\hat{p}}\,p\rangle = \underbrace{\int dx\,|x\rangle\langle x|}_{=\hat{1}}\,|p\rangle p = |p\rangle p$$

Thus we find that the momentum operator defined as

$$\hat{p} = \int dx\,|x\rangle\frac{\hbar}{i}\partial_x\langle x|$$

produces the correct eigenstates and eigenvalues.

Evaluating the matrix elements $\langle p|p'\rangle$ is non-trivial, because the wave functions are not square integrable. We obtain the result through a limit. We introduce periodic boundary conditions to discretize the momentum spectrum. With periodic boundary conditions we allow only states that are periodic with a period L.

$$\underbrace{\langle x+L|p\rangle = \langle x|p\rangle}_{\text{periodicity}} \qquad \Rightarrow \qquad p_n = \frac{2\pi\hbar}{L}n$$

Now we perform the integral over one periodic unit

$$\int_0^L dx\,\langle p_i|x\rangle\langle x|p_j\rangle = \int_0^L dx\,\mathrm{e}^{i(p_j-p_i)x/\hbar} = \left\{\begin{matrix} L & \text{for} & i=j \\ 0 & \text{for} & i\neq j\end{matrix}\right\}$$

$$= L\delta_{i,j} = \frac{2\pi\hbar}{\Delta p}\delta_{i,j}$$

where $\delta p = p_{i+1} - p_i = \frac{2\pi\hbar}{L}$.

Now we can perform the limit $L \to \infty$ and obtain[1]

$$\langle p|p'\rangle = \int_{-\infty}^{\infty} dx\,\langle p|x\rangle\langle x|p'\rangle = 2\pi\hbar\delta(p-p')$$

Finally, we need to represent the unit operator in terms of momentum eigenstates. Also here it is advantageous to start from a discrete spectrum. For any complete, but not necessarily orthonormal, basis $\{|u_i\rangle\}$, the unit operator has the form

$$\hat{1} = \sum_{i,j}|u_i\rangle S_{i,j}\langle u_j|$$

where $\mathbf{S}$ is the inverse of the overlap matrix. The overlap matrix has elements $\langle u_i|u_j\rangle$.

$$\delta_{i,j} = \sum_k S_{i,k}\langle p_k|p_j\rangle = \sum_k S_{i,k}\frac{2\pi\hbar}{\delta p}\delta_{k,j} = S_{i,j}\frac{2\pi\hbar}{\Delta p}$$

$$\Rightarrow S_{i,j} = \frac{\Delta p}{2\pi\hbar}\delta_{i,j}$$

---

[1] we need to show that the integral over the delta function is equal to one:

$$1 = \sum_i \delta_{i,j} = \sum_i \Delta p\frac{1}{\Delta p}\delta_{i,j} \to \int dp'\frac{\delta_{p',j}}{\Delta p} = \int dp\,\delta(p'-p_j)$$

Thus we can use our momentum eigenstates and obtain

$$\hat{1} = \sum_{i,j} |p_i\rangle \frac{\Delta p}{2\pi\hbar} \delta_{i,j} \langle p_j| = \sum_i \frac{\Delta p}{2\pi\hbar} |p_i\rangle\langle p_i| = \int \frac{dp}{2\pi\hbar}\, |p\rangle\langle p| \tag{B.2}$$

Thus the sum over states normalized states is transformed into an integral over states with a probability density equal to one.

Note that an operator may have a spectrum that contains discrete eigenvalues and continuous intervals of eigenvalues.

Hamilton operator

The Hamilton operator is obtained from the classical Hamilton function by replacing the position argument with the position operator and the momentum argument with the momentum operator.

$$\hat{H} \overset{\text{def}}{=} H(\hat{p}, \hat{x})$$

The function of an operator is defined by its power-series expansion. If the Hamilton function has the power series expansion of the Hamilton function is

$$H(p, x) = \sum_{i,j} a_{i,j} p^i x^j$$

the Hamilton operator[2] has the form

$$\begin{aligned}
\hat{H} = \sum_{i,j} a_{i,j} \hat{p}^i \hat{x}^j &= \sum_{i,j} a_{i,j} \left(\int dx\, |x\rangle \frac{\hbar}{i} \partial_x \langle x| \right)^i \left(\int dx\, |x\rangle x \langle x| \right)^j &\text{(B.3)}\\
&= \int dx\, |x\rangle \sum_{i,j} a_{i,j} \left(\frac{\hbar}{i} \partial_x \right)^i (x)^j \langle x| &\text{(B.4)}\\
&= \int dx\, |x\rangle H(\frac{\hbar}{i}\partial_x, x) \langle x| &\text{(B.5)}
\end{aligned}$$

---

[2] The momenta and positions must be ordered such that the resulting operator is hermitian

Appendix C

Background material

C.1 Rotational energy levels

Editor:This is not read see Atkins Physical chemistry, p554 and Atkins Molecular quantum mechanics

For a general molecule we calculate the moment of inertia as

$$I_{i,j} = \sum_{k=1}^{N} m_k r_{k,i} r_{k,j}$$

The rotational energy is

$$E = \frac{1}{2} \vec{L} \mathbf{I}^{-1} \vec{L}$$

This can be simplified by transforming on the principal axes of the moment of inertia, which has eigenvalues I_1, I_2, I_3. Now the energy can be written as

$$E = \sum_{i=1}^{3} \frac{L_i^2}{2I_i}$$

From the angular momentum we obtain

$$
\begin{aligned}
L_z|\ell,m\rangle &= |\ell,m\rangle\hbar m\\
L^2|\ell,m\rangle &= |\ell,m\rangle\hbar^2\ell(\ell+1)\\
L_+|\ell,m\rangle &= |\ell,m+1\rangle\hbar\sqrt{(\ell-m)(\ell+m+1)}\\
L_-|\ell,m\rangle &= |\ell,m-1\rangle\hbar\sqrt{(\ell+m)(\ell-m+1)}\\
L_- &= L_x - iL_y\\
L_+ &= L_x + iL_y\\
L_z^2|\ell,m\rangle &= |\ell,m\rangle\hbar^2 m^2\\
L_x^2+L_y^2|\ell,m\rangle &= \hbar^2|\ell,m\rangle\left[\ell(\ell+1)-m^2\right]\\
L_x^2 &= \frac{1}{4}(L_++L_-)^2 = \frac{1}{4}(L_+^2+L_+L_-+L_-L_++L_-^2)\\
&= \frac{1}{4}(L_+^2+L^2-L_z^2-\hbar L_z+L_-^2)\\
&= \frac{1}{4}\left(L^2-L_z^2-\hbar L_z+(L_+^2+L_-^2)\right)\\
L_y^2 &= \frac{1}{4}(L_+-L_-)^2 = \frac{1}{4}(L_+^2-L_+L_--L_-L_++L_-^2)\\
&= \frac{1}{4}(L_+^2-L^2+L_z^2+\hbar L_z+L_-^2)\\
&= -\frac{1}{4}\left(L^2-L_z^2-\hbar L_z-(L_+^2+L_-^2)\right)
\end{aligned}
$$

C.2 Equations of state from the equilibrium probabilities

Here we derive the equations of state from the form of the equilibrium probabilities.

Here we derive Eqs. 1.67, 1.65 and Eq. 1.66 from the microscopic states.
Proof:

- First we show how the expectation values can be obtained from the free energy.

$$
\begin{aligned}
\frac{\partial F(T,\{f_k\})}{\partial f_i} &\overset{\text{Eq. 1.61}}{=} \frac{\partial}{\partial f_i}\left(-k_BT\ln[Z(T,\{f_k\})]\right) = -k_BT\frac{1}{Z(T,\{f_k\})}\frac{\partial}{\partial f_i}Z(T,\{f_k\})\\
&\overset{\text{Eq. 1.61}}{=} -k_BT\underbrace{\mathrm{e}^{+\frac{1}{k_BT}F}}_{\frac{1}{Z(T,\{f_k\})}}\sum_{\vec{n}}\left(-\frac{1}{k_BT}X_{i,\vec{n}}\right)\mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}+\sum_i f_iX_{i,\vec{n}})}\\
&= \sum_{\vec{n}}X_{\vec{n},i}\underbrace{\mathrm{e}^{-\frac{1}{k_BT}(E_{\vec{n}}+\sum_i f_iX_{i,\vec{n}}-F(T,\{f_k\}))}}_{=P_{\vec{n}}(T,f_i)}\\
&= \sum_{\vec{n}}P_{\vec{n}}X_{i,\vec{n}} = X_i(T,\{f_k\})
\end{aligned}
$$

cont'd...

- Now we calculate the temperature derivative of the free energy

$$
\begin{aligned}
\frac{\partial F(T,\{f_k\})}{\partial T} &\overset{\text{Eq. 1.61}}{=} \frac{\partial}{\partial T}\left(-k_BT\ln[Z(T,\{f_k\})]\right)\\
&= \underbrace{-k_B\ln[Z(T,\{f_k\})]}_{F/T} - k_BT\frac{1}{Z(T,\{f_k\})}\frac{\partial}{\partial T}Z(T,\{f_k\})\\
&= \frac{1}{T}F - k_BT\underbrace{\mathrm{e}^{+\frac{1}{k_BT}F}}_{\frac{1}{Z(T,\{f_k\})}}\\
&\quad\cdot\sum_{\vec n}\left(\frac{1}{k_BT^2}\left(E_{\vec n}+\sum_i f_iX_{\vec n,i}\right)\right)\mathrm{e}^{-\frac{1}{k_BT}\left(E_{\vec n}+\sum_i f_iX_{\vec n,i}\right)}\\
&= \frac{1}{T}F - \frac{1}{T}\sum_{\vec n}\left(E_{\vec n}+\sum_i f_iX_{\vec n,i}\right)\underbrace{\mathrm{e}^{-\frac{1}{k_BT}(E_{\vec n}+\sum_i f_iX_{\vec n,i}-F(T,\{f_k\}))}}_{=P_{\vec n}(T,\{f_k\})}\\
&= \frac{1}{T}\left(F(T,\{f_k\}) - U(T,\{f_k\}) - \sum_i f_iX_i(T,\{f_k\})\right)
\end{aligned}
$$

cont'd...

- The entropy as function of the intensive variables is

$$
\begin{aligned}
S(T,\{f_k\}) &\overset{\text{Eq. 1.12}}{=} -k_B\sum_{\vec n}P_{\vec n}(T,\{f_k\})\ln[P_{\vec n}(T,\{f_k\})]\\
&= \frac{1}{T}\sum_{\vec n}P_{\vec n}(T,\{f_k\})\left(E_{\vec n}+\sum_{i=1}^n f_iX_{i,\vec n}-F(T,\{f_k\})\right)\\
&= \frac{1}{T}\Big(\sum_{\vec n}P_{\vec n}(T,\{f_k\})E_{\vec n}+\sum_{i=1}^n f_i\sum_{\vec n}P_{\vec n}(T,\{f_k\})X_{i,\vec n}\\
&\quad -F(T,\{f_k\})\underbrace{\sum_{\vec n}P_{\vec n}(T,\{f_k\})\Big)}_{=1}\\
&= \frac{1}{T}\left(U(T,\{f_k\})+\sum_{i=1}^n f_iX_i(T,\{f_k\})-F(T,\{f_k\})\right)
\end{aligned}
$$

The expression for the entropy is identical with the temperature derivative of the free energy. Using one of these expressions we can then evaluate the internal energy

C.3 Configurational entropy

C.3.1 Number of arrangements of N distinguishable particles on N positions

Here we determine the number $r(N)$ of ways N distinguishable particles can be arranged on N lattice sites.

- We can choose from N atoms and place them on the first position.
- For each such choice we can choose from $N-1$ atoms for the second position. Hence there are $N(N-1)$ ways to place the first two atoms.
- For the third position we can choose from $N-2$ atoms, so that we can arrange the first three atoms in $N(N-1)(N-2)$ ways on the first three positions.
- We continue this and obtain $N(N-1)(N-2)\dots(N-i+1)$ for the i-th position.
- Once we arrive at $i=N$, we obtain $N\cdot(N-1)\dots2\cdot1=N!$.

Thus we obtain the result

$$r(N)=N! \tag{C.1}$$

C.3.2 Number of arrangements of indistinguishable particles

Consider M_s lattice sites on which N atoms can be placed. Every position can be occupied by at most one atom. The atoms shall be indistinguishable, so that one configuration is determined only by the positions that are occupied: A permutation of two atoms does not produce a new configuration.

1. The first atom can be placed on any of the M_s lattice sites. Therefore, we obtain $q(1)=M_s$
2. The second atom can be placed on any of the M_s-1 lattice sites. Thus we obtain $M_s(M_s-1)$ configurations. However, each configuration has been counted twice: For every configuration with the first atom on site i and the second atom on site j, there is also one with the first atom on site j and the second on site i. Wen we account for double counting we obtain $q(2,M_s)=\frac{M_s(M_s-1)}{2}$
3. The third atom can be placed M_s-2 sites. Again we need to account for double counting. For any three occupied sites we obtained the following six occupations (123); (132); (213); (231); (312); (321), where the numbers refer to the first, second and third atom and the position in the triple refers to the three occupied lattice sites. There are six possibilities and therefore $q(3,M_s)=\frac{M_s(M_s-1)(M_s-2)}{6}$. Note that the divisor is simply $r(3)$ as defined in Eq. C.1.
4. For the N-th atom we obtain $M_s(M_s-1)\dots(M_s-N+1)=\frac{M_s!}{(M_s-N)!}$ possibilities for distinguishable atoms. The number of ways the M atoms have been arranged for a given set of position is $M!$, as obtained in eq. C.1.

Thus we obtain the configurational degeneracy as

$$q(N,M_s)=\frac{M_s!}{N!(M_s-N)!} \tag{C.2}$$

C.3.3 Configurational entropy

If all configuration can occur with the same a-priori probability, the entropy is given as $S_{conf}=k_B\ln[q(N,M_s)]$ by the configurational degeneracy given in Eq. C.2. Thus we obtain

$$S_{conf}(N,M_s)=k_B\ln[\frac{M_s!}{N!(M_s-N)!}] \tag{C.3}$$

For large numbers, $M_s >> 1; N >> 1; M_s - N >> 1$ we can use Stirling's formula, Eq. D.1. Furthermore we introduce the average occupation $\theta = N/M_s$. We obtain

$$\begin{aligned}
S_{conf} &= k_B \ln[\frac{M_s!}{N!(M_s-N)!}] \\
&\approx k_B\Big(\underbrace{M_s\ln[M_s] - M_s}_{\approx\ln[M!]}\underbrace{-N\ln[N] + N}_{\approx-\ln[N!]}\underbrace{-(M_s-N)\ln[M_s-N] + (M_s-N)}_{\approx-\ln[(M_s-N)!]}\Big) \\
&= k_B\Big(M_s\ln[M_s] - N\ln[N] - (M_s-N)\ln[M_s-N]\Big) \\
&= M_s k_B\Big(\ln[M_s] - \theta\underbrace{\ln[M_s\theta]}_{=\ln[M_s]+\ln[\theta]} - (1-\theta)\underbrace{\ln[M_s(1-\theta)]}_{=\ln[M_s]+\ln[1-\theta]}\Big) \\
&= M_s k_B\Big(-\theta\ln[\theta] - (1-\theta)\ln[(1-\theta)]\Big)
\end{aligned}$$

Thus we obtain the following expression for the configurational entropy in the limit of large numbers

$$S_{conf}(\theta) = -M_s k_B\Big(\theta\ln[\theta] + (1-\theta)\ln[(1-\theta)]\Big) \tag{C.4}$$

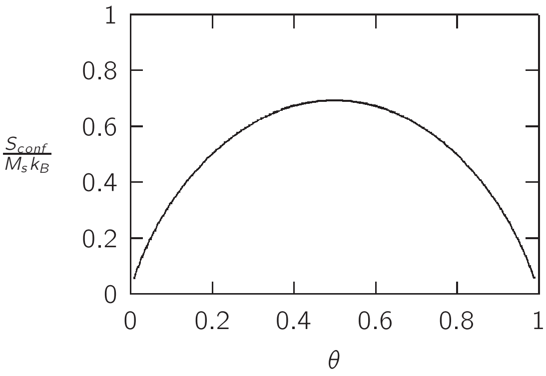


Fig. C.1: Configurational entropy as function of occupation θ

C.4 Free energy of formation and defect concentration

Consider a defect in a host lattice. The energy required to create a defect is the formation energy $E_F = E[N=1] - E[N=0]$, where $E[M_s=1]$ is the total energy of the system with one defect and $E[N=0]$ is the total energy of the defect, free host.

I the defects are sufficiently dilute, so that there is no interaction between defects, we can compose the Free energy for a given number of defects from the total energy $M_s\theta E_F$ and the entropies. We divide the entropy into the configurational entropy and a vibrational entropy Ns_{vib}. s_{vib} is the vibrational entropy per atom.

$$\begin{aligned}
F(T,N) &= -k_BT\ln[Z] = -k_BT\ln[\frac{M_s!}{N!(M_s-N)!}\mathrm{e}^{-\frac{N(E_F-Ts_{vib})}{k_BT}}] \\
&= -\underbrace{k_BT\ln[\frac{M_s!}{N!(M_s-N)!}]}_{=TS_{conf}} + N(E_F - Ts_{vib}) \\
&= M_s\Big(\theta E_F - \theta Ts_{vib} + k_BT\Big(\theta\ln[\theta] + (1-\theta)\ln[1-\theta]\Big)
\end{aligned}$$

From the free energy we can derive the chemical potential, which can be equated with that of the particle bath to obtain the concentration of defects.

$$\begin{aligned}\mu &= \frac{dF}{dN} = \frac{dF}{d\theta}\frac{d\theta}{dN}\\ &= E_F + k_BT\left(\ln[\theta] + 1 - \ln[1-\theta] - 1\right) - Ts_{vib}\\ &= E_F + k_BT\ln[\frac{\theta}{1-\theta}] - Ts_{vib}\end{aligned}$$

We set this chemical potential equal to that of the particle reservoir, and obtain

$$\begin{aligned}\mu_{bath} = \mu &= E_F + k_BT\ln[\frac{\theta}{1-\theta}] - Ts_{vib}\\ \frac{\theta}{1-\theta} &= \mathrm{e}^{-\frac{E_F-\mu_{bath}-Ts_{vib}}{k_BT}}\\ \theta &= \left[1+\exp(\frac{E_F-\mu_{bath}-Ts_{vib}}{k_BT})\right]^{-1}\end{aligned} \tag{C.5}$$

The result is shown in Fig. C.2

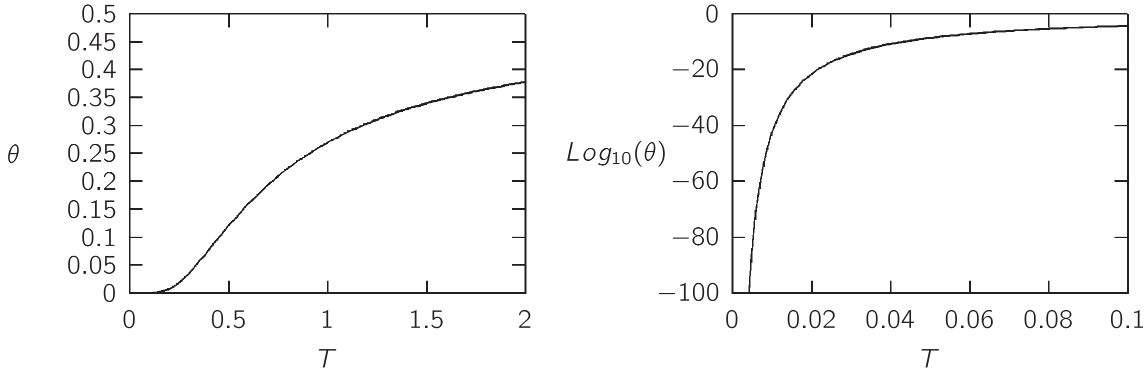


Fig. C.2: Top: Occupation as function of temperature. Bottom: Decadic logarithm of the above. $(E_F - \mu_{bath} - Ts_{vib})/k_B$ has been set to one.

C.4.1 Concentration if the heat bath is a gas

The chemical potential of a gas is

$$\mu = k_BT\ln\left[\frac{p\lambda_T^3}{k_BT}\right] \qquad \text{with} \qquad \lambda_T = \sqrt{\frac{2\pi\hbar}{mk_BT}}$$

We insert this into Eq.C.5 and obtain

$$\begin{aligned}\theta &= \left[1+\exp(\frac{E_F-\mu_{bath}-Ts_{vib}}{k_BT})\right]^{-1}\\ &= \left[1+\frac{k_BT}{p\lambda_T^3}\exp(\frac{E_F}{k_BT})\exp(-\frac{s_{vib}}{k_B})\right]^{-1}\end{aligned}$$

In the limit of small partial pressures, that is small coverage, we obtain

$$\theta(p << 1) = \frac{p\lambda_T^3}{k_BT}\exp(-\frac{E_F}{k_BT})\exp(\frac{s_{vib}}{k_B})$$

The concentration depends strongly on s_{vib}. The vibrational entropy is for each vibrational mode is

$$s_{vib} = k_B \ln[1 - \mathrm{e}^{-\hbar\omega/(k_B T)}]$$

If one atom is adsorbed, there are three additional vibrational modes so that the entropy above must be multiplied by three. In general we denote the number additional vibrational modes by d.

$$\theta = \left[1 + \frac{k_B T}{p\lambda_T^3} \cdot \frac{1}{(1 - \mathrm{e}^{-\hbar\omega/(k_B T)})^d} \exp(\frac{E_F}{k_B T})\right]^{-1}$$

At low temperatures, that is $k_B T << \hbar\omega$ the exponential containing the vibrational frequencies becomes small compared to unity and drops out. At high temperatures we obtain

$$\theta(k_B T >> \hbar\omega) = \left[1 + \frac{k_B T}{p\lambda_T^3} \cdot \left(\frac{k_B T}{\hbar\omega}\right)^d \exp(\frac{E_F}{k_B T})\right]^{-1}$$

C.5 Concentration of oxygen vacancies in an oxide

C.5.1 Vacancies

First we need to determine the entropy for vacancies in a lattice with M sites. We need to count the number of possibilities to distribute N vacancies onto M lattice sites.

For one vacancy there are clearly M possibilities, hence $S(N = 1) = k_B \ln[M]$. For 2 vacancies we can place the first onto one of M lattice sites, and the second onto $M - 1$ lattice sites. However, we created each vacancy distribution twice. (The first on place A and the second on place B or the first on site B and the second on place A.) Therefore the entropy is $S(N = 2) = k_B \ln[M(M - 1)/2]$. Continuing this result we obtain for a general number of vacancies:

$$S(N) = k_B \ln[\frac{M!}{(M - N)!N!}]$$

We use Stirling's Formula $\ln[n!] = n \ln[n] - n + O(1/N)$ and obtain

$$\begin{aligned}
S(N) &= k_B \left(M \ln[M] - M - (M - N) \ln[M - N] + M - N - N \ln[N] + N\right) \\
&= k_B \left(M \ln[M] - (M - N) \ln[M - N] - N \ln[N]\right) \\
&= k_B \left(M \ln[\frac{M}{M - N}] + N \ln[\frac{M - N}{N}]\right)
\end{aligned}$$

Now we introduce the concentration $c = N/M$

$$\begin{aligned}
S(N) &= M k_B \left(\ln[\frac{1}{1 - c}] + c \ln[\frac{1 - c}{c}]\right) \\
&= M k_B \left(-\ln[1 - c] + c \ln[1 - c] - c \ln[c]\right) \\
&= -M k_B \left((1 - c) \ln[1 - c] + c \ln[c]\right)
\end{aligned}$$

Note that the entropy is similar to the Entropy for electrons.

The energy of a system is equal to N times the formation energy of a vacancy, namely $E_F = E(N + 1) - E(N)$.

$$E(N) = N E_F = M c E_F$$

Combining energy and entropy we obtain the free energy

$$F(N) = E(N) - T S(N) = M\left[c E_F + k_B T \left((1 - c) \ln[1 - c] + c \ln[c]\right)\right]$$

The chemical potential of oxygen atoms, which is the negative of the chemical potential for a vacancy. It can be calculated from the derivative of the free energy.

$$
\begin{aligned}
-\mu_O &= \frac{\partial F(T,N)}{\partial N} \\
&= M\frac{dc}{dN}\frac{\partial}{\partial c}\left[cE_F + k_BT\left((1-c)\ln[1-c] + c\ln[c]\right)\right] \\
&= E_F + k_BT\left(-\ln[1-c] - \frac{1-c}{1-c} + \ln[c] + \frac{c}{c}\right) \\
&= E_F + k_BT\ln[\frac{c}{1-c}]
\end{aligned}
$$

Thus we obtain the chemical potential required to produce a given vacancy concentration.

$$
\begin{aligned}
\mu_O &= -E_F - k_BT\ln[\frac{c}{1-c}] \\
c &= \frac{1}{1+\mathrm{e}^{\frac{\mu_O+E_F}{k_BT}}}
\end{aligned}
$$

C.5.2 Ideal Gas

$$
\begin{aligned}
F(T,V,N) &= -Nk_BT + \ln[\frac{V}{N\lambda_T^3}] \\
\lambda_T &= \sqrt{\frac{2\pi\hbar^2}{mk_BT}}
\end{aligned}
$$

where m is the mass of the molecule, not the mass of the atom.

The Gibbs Free energy

$$
G(T,p,N) = F(T,V,N) + pV \quad \text{, where} \qquad p = \frac{\partial F(T,V,N)}{\partial V}.
$$

The chemical potential is the derivative of the Gibbs free energy with respect to particle number

$$
\begin{aligned}
\mu|_{p,T} &= \frac{\partial G}{\partial N} = \left[\frac{\partial F}{\partial V} + p\right]\frac{\partial V}{\partial N} + \frac{\partial F}{\partial N} \\
&= -k_BT\left[1 + \ln\left[\frac{V}{N\lambda_T^3}\right]\right] - Nk_BT\left[-\frac{1}{N}\right] \\
&= -k_BT\left[1 + \ln\left[\frac{V}{N\lambda_T^3}\right] - 1\right] \\
&= -k_BT\ln\left[\frac{V}{N\lambda_T^3}\right] \\
&= +k_BT\ln\left[\frac{p\lambda_T^3}{k_BT}\right]
\end{aligned}
$$

In the last step, we used the ideal gas law $pV = Nk_BT$ to express the particle density by the partial pressure. (The ideal Gas law follows from $p = \partial F/\partial V$).

C.5.3 Rotator

$$
\begin{aligned}
E_{\ell,m} &= \frac{\hbar^2\ell(\ell+1)}{2m^* r^2} \\
Z_1 &= \sum_{\ell,m} \mathrm{e}^{-\beta E_{\ell,m}} \\
&= \sum_{\ell} (2\ell+1)\mathrm{e}^{-\frac{\hbar^2\beta}{2m^* r^2}\ell(\ell+1)} \\
&\approx \int_{-\frac{1}{2}}^{\infty} dx (2x+1)\mathrm{e}^{-\sigma x(x+1)} \\
&= \int_{-\frac{1}{2}}^{\infty} dx \frac{-1}{\sigma}\frac{d}{dx}\mathrm{e}^{-\sigma x(x+1)} \\
&= \frac{1}{\sigma}\mathrm{e}^{\sigma/4} \\
\sigma &= \frac{\hbar^2\beta}{2m^* r^2}
\end{aligned}
$$

The integrand has a zero at $x = -\frac{1}{2}$. Starting the integral at this point is introduces a smaller error than starting from zero. The approximation is accurate for high temperatures.

$$
\begin{aligned}
F(T,N) &= -k_B T \ln[Z_N] = -Nk_B T \ln[Z_1] = -Nk_B T \ln\left[\frac{1}{\sigma}\mathrm{e}^{\sigma/4}\right] \\
&= Nk_B T \ln\left[\sigma \mathrm{e}^{-\sigma/4}\right]
\end{aligned}
$$

The contribution to the chemical potential from the rotation is therefore

$$
\begin{aligned}
\mu_{rot} = \frac{\partial F}{\partial N} &= k_B T \\
In&\left[\sigma \mathrm{e}^{-\sigma/4}\right]
\end{aligned}
$$

Note that the mass is the effective mass of the molecule $\frac{1}{m^*} = \frac{1}{m_1} + \frac{1}{m_2}$.

C.5.4 Oxidation

$$
\begin{aligned}
O_2 &\leftrightarrow 2O \\
\Rightarrow E[O_2] &= 2E[O] + 2\mu_O \\
\Rightarrow \mu_O &= \frac{1}{2}E[O_2] - E[O] \\
SrTiO_3 &\leftrightarrow SrTiO_{3-x} + xO \\
\Rightarrow E[SrTiO_3] &= E[SrTiO_{3-x}] + xE[O] + x\mu_O \\
\Rightarrow \mu_O &= \frac{E[SrTiO_3] - E[SrTiO_{3-x}]}{x} - E[O] = -E_F - E_O
\end{aligned}
$$

The energy is to be taken the free energy.

The energy to remove an oxygen atom from the gas is

$$
\mu_O = \frac{1}{2}E[O_2] - E[O] + \frac{k_B T}{2}\ln\left[\frac{p\lambda_T^3}{k_B T}\right]
$$

where λ_T need to be evaluated with the molecular mass (and not the atomic mass).

The energy to remove an oxygen from the oxide is

$$\mu_O = -E_F - E_O - k_B T \ln\left[\frac{c}{1-c}\right]$$

In thermal equilibrium the two chemical potentials must be identical

$$\begin{aligned}
-E_F - E_O - k_B T \ln\left[\frac{c}{1-c}\right] &= \frac{1}{2}E[O_2] - E[O] + \frac{k_B T}{2} \ln\left[\frac{p\lambda_T^3}{k_B T}\right] \\
\mathrm{e}^{-\frac{2E_F + E[O_2]}{k_B T}} &= \left(\frac{c}{1-c}\right)^2 \frac{p\lambda_T^3}{k_B T} \\
p &= \left(\frac{1-c}{c}\right)^2 \left(\frac{k_B T}{\lambda_T^3}\right) \mathrm{e}^{-\frac{2E_F + E[O_2]}{k_B T}} \\
c &= \left[1 + \sqrt{\frac{p\lambda_T^3}{k_B T}\mathrm{e}^{\frac{E_F + E[O_2]/2}{k_B T}}}\right]^{-1}
\end{aligned}$$

C.6 Origin of the exchange interaction

Editorial Remark: This should go into an appendix The interaction between spins in the Ising model is not of magnetic nature as one might naively expect. The magnetic interaction is long ranged and much smaller than the exchange interactions that play a role in the Ising model.

Let us investigate the exchange interaction. Consider two sites with one electron on each site. The basis set is built up of a single spatial orbital on each site, ϕ and ψ. Each orbital is a two component spinor, where the upper component corresponds to the spin-up contribution and the lower component is the spin-down contribution so that we can build four one-particle orbitals from two spatial wave functions.

$$\begin{aligned}
|\uparrow, 0\rangle &= \begin{pmatrix} \phi(r) \\ 0 \end{pmatrix} = \phi(r)\alpha \\
|\downarrow, 0\rangle &= \begin{pmatrix} 0 \\ \phi(r)\rangle \end{pmatrix} = \phi(r)\beta \\
|0, \uparrow\rangle &= \begin{pmatrix} \psi(r) \\ 0\rangle \end{pmatrix} = \psi(r)\alpha \\
|0, \downarrow\rangle &= \begin{pmatrix} 0 \\ \psi(r)\rangle \end{pmatrix} = \psi(r)\beta
\end{aligned}$$

where $\alpha = (1, 0)$ and $\beta = (0, 1)$.

Out of the one-particle wave functions we build two-particle Slater determinants of the type $|\uparrow, \downarrow\rangle = \frac{1}{\sqrt{2}}\Big(|\uparrow, 0\rangle \times |0, \downarrow\rangle - |0, \downarrow\rangle \times |\uparrow, 0\rangle\Big)$. Out of the 16 product states that can be build from the four one-particle states, 6 are excluded because they differ only by a sign change, 4 are excluded because they involve two identical one-particle states, which is prohibited due to Pauli principle. We are left with 6 states:

$$\begin{aligned}
|\uparrow\downarrow,0\rangle &= \phi(r_1)\phi(r_2)\frac{\alpha_1\beta_2-\beta_1\alpha_2}{\sqrt{2}}\\
|\uparrow,\uparrow\rangle, &= \frac{1}{\sqrt{2}}\Big(\phi(r_1)\psi(r_2)-\psi(r_1)\phi(r_2)\Big)\alpha_1\alpha_2\\
|\uparrow,\downarrow\rangle, &= \frac{1}{\sqrt{2}}\Big(\phi(r_1)\psi(r_2)\alpha_1\beta_2-\psi(r_1)\phi(r_2)\beta_1\alpha_2\Big)\\
|\downarrow,\uparrow\rangle, &= \frac{1}{\sqrt{2}}\Big(\phi(r_1)\psi(r_2)\beta_1\alpha_2-\psi(r_1)\phi(r_2)\alpha_1\beta_2\Big)\\
|\downarrow,\downarrow\rangle, &= \frac{1}{\sqrt{2}}\Big(\phi(r_1)\psi(r_2)-\psi(r_1)\phi(r_2)\Big)\beta_1\beta_2\\
|0,\uparrow\downarrow\rangle &= \psi(r_1)\psi(r_2)\frac{\alpha_1\beta_2-\beta_1\alpha_2}{\sqrt{2}}
\end{aligned}$$

Two states, namely $|\uparrow\downarrow,0\rangle$ and $|0,\uparrow\downarrow\rangle$, correspond to charge transfer states, where one electron has been transferred from one site to another. We exclude them with the argument that the Coulomb repulsion makes these states very unfavorable.

When we evaluate the expectation value of the Hamiltonian. Note here that the operator yielding the charge density is proportional to the unity operator in spin space.

$$\begin{aligned}
H &= \sum_i \frac{-\hbar^2}{2m}\nabla_i^2 + V(r_i)\begin{pmatrix}1&0\\0&1\end{pmatrix}\\
&+\frac{e^2}{4\pi\epsilon}\frac{1}{|r_1-r_2|}\begin{pmatrix}|r_1\rangle\langle r_1| & 0\\ 0 & |r_1\rangle\langle r_1|\end{pmatrix}\begin{pmatrix}|r_2\rangle\langle r_2| & 0\\ 0 & |r_2\rangle\langle r_2|\end{pmatrix}
\end{aligned}$$

we need four variables

$$\begin{aligned}
\epsilon &= \int d^3r\,\phi_i^*(r)\Big[\frac{-\hbar^2}{2m}\nabla^2+V(r)\Big]\phi_i(r)\\
U &= \int dr\int dr'\,\frac{\phi_1^*(r)\phi(r)\phi_1^*(r)\phi(r)}{|r-r'|}\\
K &= \int dr\int dr'\,\frac{\phi_1^*(r)\phi(r)\phi_2^*(r)\psi(r)}{|r-r'|}\\
J &= \int dr\int dr'\,\frac{\phi_1^*(r)\psi(r)\phi_2^*(r)\phi(r)}{|r-r'|}
\end{aligned}$$

In order to diagonalize the Hamiltonian, we form angular momentum eigenstates

$$\begin{aligned}
|\ell=0,m=0\rangle &= \frac{1}{\sqrt{2}}\Big(|\uparrow,\downarrow\rangle-|\downarrow\uparrow\rangle\Big)\\
|\ell=1,m=1\rangle &= |\uparrow,\uparrow\rangle\\
|\ell=1,m=0\rangle &= \frac{1}{\sqrt{2}}\Big(|\uparrow,\downarrow\rangle+|\downarrow\uparrow\rangle\Big)\\
|\ell=1,m=-1\rangle &= |\downarrow,\downarrow\rangle
\end{aligned}$$

The energies of the eigenstates can then be evaluated as the diagonal elements of the Hamiltonian

$$\begin{pmatrix}\langle\uparrow\uparrow|\\ \langle\uparrow\downarrow|\\ \langle\downarrow\uparrow|\\ \langle\downarrow\downarrow|\end{pmatrix} H \begin{pmatrix}|\uparrow\uparrow\rangle\\ |\uparrow\downarrow\rangle\\ |\downarrow\uparrow\rangle\\ |\downarrow\downarrow\rangle\end{pmatrix} = \begin{pmatrix}K-J & 0 & 0 & 0\\ 0 & K & -J & 0\\ 0 & -J & K & 0\\ 0 & 0 & 0 & K-J\end{pmatrix}$$

We obtain the eigenstates as

$$\begin{aligned}
H|\uparrow,\uparrow\rangle &= K-J|\uparrow,\uparrow\rangle\\
H\frac{1}{\sqrt{2}}\Big(|\uparrow,\downarrow\rangle+|\downarrow,\uparrow\rangle\Big) &= (K-J)\frac{1}{\sqrt{2}}\Big(|\uparrow,\downarrow\rangle+|\downarrow,\uparrow\rangle\Big)\\
H\frac{1}{\sqrt{2}}\Big(|\uparrow,\downarrow\rangle-|\downarrow,\uparrow\rangle\Big) &= (K+J)\frac{1}{\sqrt{2}}\Big(|\uparrow,\downarrow\rangle-|\downarrow,\uparrow\rangle\Big)\\
H|\downarrow,\downarrow\rangle &= K-J|\downarrow,\downarrow\rangle
\end{aligned}$$

The Coulomb integral appears on all diagonal elements, because all states have equal charge density, and only spins are interchanged.

The two states

$$\begin{aligned}
\frac{1}{\sqrt{2}}\Big(|\uparrow,\downarrow\rangle-|\downarrow,\uparrow\rangle\Big) &= \frac{1}{2}\Big(\phi(r_1)\psi(r_2)+\psi(r_1)\phi(r_2)\Big)\Big(\alpha_1\beta_2-\beta_1\alpha_2\Big)\\
\frac{1}{\sqrt{2}}\Big(|\uparrow,\downarrow\rangle+|\downarrow,\uparrow\rangle\Big) &= \frac{1}{2}\Big(\phi(r_1)\psi(r_2)-\psi(r_1)\phi(r_2)\Big)\Big(\alpha_1\beta_2+\beta_1\alpha_2\Big)
\end{aligned}$$

are the singlet state with $S_{z,1}+S_{z,2}=0$ and the triplet state $S_{z,1}+S_{z,2}=\hbar$.

Thus I can write the Hamiltonian as

$$bH = K - J\sigma_1\sigma_2$$

The coupling among the spins is the exchange coupling J.

C.7 Planck law and Wien displacement law in wavelengths

Often Planck's law is expressed as intensity per wave-length interval. In order to make contact with the more common equations we do the translation explicitely.

We begin with **Planck's radiation law** Eq. 8.33 on p. 213 in section 8.7.2.

$$\frac{dI}{d\omega} \overset{\text{Eq. 8.33}}{=} 8\pi\hbar c\left(\frac{k_BT}{2\pi\hbar c}\right)^3\frac{(\beta\hbar\omega)^3}{e^{\beta\hbar\omega}-1}$$

Note that this quantity is not the derivative of the intensity with respect to angular frequency – the intensity does not depend on frequency –, but the contribution from a angular frequency interval of width $d\omega$ centered at ω to the Intensity I.

The angular frequency is related to the wave-length by

$$\omega = c|k| = c\frac{2\pi}{\lambda}$$

Thus we can relate a wave-length interval $\Delta\lambda$ to an interval $\Delta\omega$ in angular frequency by

$$\Delta\omega = \left|\frac{d\omega(\lambda)}{d\lambda}\right|\Delta\lambda = \frac{2\pi c}{\lambda^2}\Delta\lambda$$

This yields

$$
\begin{aligned}
\frac{dI}{d\lambda} &= \frac{dI}{d\omega}\frac{d\omega}{d\lambda} \\
&= \left[8\pi\hbar c\left(\frac{k_BT}{2\pi\hbar c}\right)^3 \frac{(\beta\hbar\omega)^3}{e^{\beta\hbar\omega}-1}\right]\left[\frac{2\pi c}{\lambda^2}\right] \\
&= \left[8\pi\hbar c\left(\frac{k_BT}{2\pi\hbar c}\right)^3 \frac{(2\pi\hbar c/(k_BT\lambda)^3}{e^{2\pi\hbar c/(k_BT\lambda)}-1}\right]\left[\frac{2\pi c}{\lambda^2}\right] \\
&= 16\pi^2\hbar c^2\left(\frac{k_BT}{2\pi\hbar c}\right)^3 \frac{1}{\lambda^2}\frac{(2\pi\hbar c/(k_BT\lambda)^3}{e^{2\pi\hbar c/(k_BT\lambda)}-1} \\
&= \frac{(4\pi)^2\hbar c^2}{\lambda^5}\frac{1}{e^{2\pi\hbar c/(k_BT\lambda)}-1}
\end{aligned}
$$

This is Planck's radiation law expressed in wave lengths.

PLANCK'S RADIATION LAW

The intensity distribution in a black body with temperature T as function of wavelength is

$$
\frac{dI}{d\lambda} = \frac{(4\pi)^2\hbar c^2}{\lambda^5}\frac{1}{e^{2\pi\hbar c/(k_BT\lambda)}-1} \tag{C.6}
$$

Let us now turn to Wien's displacement law. Note that the shape of the intensity distribution differs in frequency and in wavelength space. Therefore it is not sufficient to take the position of the maximum and convert the frequency into a wave length. We need to consider that the shape changes, when we multiply with $\frac{d\omega}{d\lambda}$.

We can bring the intensity into a form

$$
\frac{dI}{d\lambda} = \frac{(4\pi)^2\hbar c^2}{(2\pi\hbar c)^5}(k_BT)^5\frac{(2\pi\hbar c/(k_BT\lambda))^5}{e^{2\pi\hbar c/(k_BT\lambda)}-1}
$$

We need to determine the maximum of the function

$$
f(x) = \frac{x^5}{e^x-1}
$$

which lies at $x_{max} = 4.965114231$.

Next we resolve for x

$$
\begin{aligned}
&& x_{max} &= \frac{2\pi\hbar}{k_BT\lambda_{max}} \\
\Rightarrow && \lambda_{max}T &= \frac{2\pi\hbar}{k_Bx_{max}} \stackrel{\text{def}}{=} b
\end{aligned}
$$

WIEN'S DISPLACEMENT LAW

The wave length with the maximum intensity contribution is inverse proportional to the temperature.

$$
\lambda_{max} = \frac{b}{T}
$$

The constant b is called **Wien's displacement constant** and has the value[19] $b = 2.8977686(51) \times 10^{-3}$ mK.

Appendix D

Mathematical formulas

D.1 Stirling's formula

Stirling's formula

$$\ln[n!] \approx n\ln[n] - n \tag{D.1}$$

and its more accurate version

$$\ln[n!] \approx \left(n + \frac{1}{2}\right)\ln[n] - n + \ln[2\pi] \tag{D.2}$$

is very useful in statistical mechanics because the factorial $n! \overset{\text{def}}{=} 1 \cdot 2 \cdots n$ occurs in many combinatorial problems and the logarithm is required to evaluate entropies.

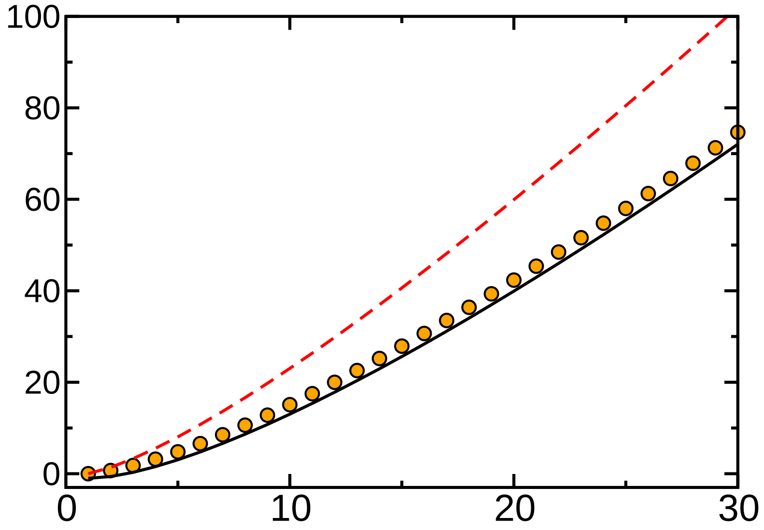


Fig. D.1: Comparison of stirling's formula with exact result. The circles represent $\ln[n!]$. The full line is stirling's formula $n\ln[n] - n$. The dashed line is $n\ln[n]$.

As seen by comparison with Eq. D.2, the error of the approximation Eq. D.1 in Stirling's formula increases like $\ln[n]$ with n. Why could a formula be useful, that does not converge to the correct result at all? It is useful, if the quantity of interest contains $\frac{1}{n}\ln[n!]$ so that we can use

$$\frac{1}{n}\ln[n!] \rightarrow \frac{1}{n}\left[n\ln[n] - n\right] \qquad \text{for} \qquad n \rightarrow \infty$$

Most problems in thermodynamics “effectively” are of this form.

D.1.1 Proof of Stirling’s formula

Here another proof using the saddle point approximation:

The proof is based on a representation of the factorial by an integral: Consider the integral, which we solve by repeated partial integration

$$\begin{aligned}\int_0^\infty dt\; t^n \mathrm{e}^{-t} &= \int_0^\infty dt\; \left[nt^{n-1}\mathrm{e}^{-t} - \partial_t\left(t^n \mathrm{e}^{-t}\right)\right]\\ &\overset{n\geq 0}{=} n\left(\int_0^\infty dt\; t^{n-1}\mathrm{e}^{-t}\right)\\ &\overset{n-1\geq 0}{=} n(n-1)\left(\int_0^\infty dt\; t^{n-2}\mathrm{e}^{-t}\right)\\ &\overset{n-i\geq 0}{=} n(n-1)\dots(n-i)\left(\int_0^\infty dt\; t^{n-i-1}\mathrm{e}^{-t}\right)\\ &\overset{1\geq 0}{=} n(n-1)\dots 1\underbrace{\left(\int_0^\infty dt\; t^{0}\mathrm{e}^{-t}\right)}_{=1}\\ &= n(n-1)\dots 1 = n!\end{aligned}$$

Now we bring the integral into a more convenient form

$$n! = \int_0^\infty dt\; t^n \mathrm{e}^{-t} = \int_0^\infty dt\; \mathrm{e}^{n\ln(t)-t}$$

The maximum of the function in the exponent has its maximum at $t = n$. We introduce now a variable transform $s = nt$ so that the position of the maximum is independent of n.

$$\begin{aligned}n! &\overset{t=ns}{=} n\int_0^\infty ds\; \mathrm{e}^{n\ln[ns]-ns}\\ &\overset{\ln[ab]=\ln[a]+\ln[b]}{=} n\int_0^\infty ds\; \mathrm{e}^{n(\ln[s]-s+\ln[n])}\end{aligned}$$

Furthermore we separate the value of the maximum out of the integral

$$n! = n\mathrm{e}^{n\ln[n]-n}\int_0^\infty ds\; \mathrm{e}^{n(\ln[s]-s+1)} \tag{D.3}$$

Now we use the so-called **saddle-point approximation**: For an integral of the form $\int_a^b dx\; \mathrm{e}^{nf(x)}$ with a sufficiently large n, only a narrow region near the maximum of $f(x)$ contributes appreciably to the integral. In this region, we can replace $f(x)$ by its Taylor expansion up to second order about the maximum x_0. Within this approximation the integrand is converted into a Gaussian. Since the Gaussian contributes appreciably only in a narrow region around the maximum we may also change the bounds of the integral to infinity. Thus we obtain

$$\int_a^b dx\; \mathrm{e}^{nf(x)} \approx \int_{-\infty}^{\infty} dx\; \mathrm{e}^{n\left(f(x_0)+\frac{1}{2}\frac{d^2f}{dx^2}\Big|_{x_0}(x-x_0)^2\right)}$$

This integral can be solved analytically using $\int_{-\infty}^{\infty} dx\, \mathrm{e}^{-x^2} = \sqrt{\pi}$, i.e. Eq. D.6.

$$
\begin{aligned}
\int_a^b dx\, \mathrm{e}^{nf(x)} &\approx \int_{-\infty}^{\infty} dx\, \mathrm{e}^{nf(x_0)+\frac{1}{2}n\frac{d^2f}{dx^2}\big|_{x_0}(x-x_0)^2} \\
&= \mathrm{e}^{nf(x_0)} \int_{-\infty}^{\infty} dx\, \mathrm{e}^{\frac{1}{2}n\frac{d^2f}{dx^2}\big|_{x_0}x^2} \\
&= \mathrm{e}^{nf(x_0)} \sqrt{\frac{1}{-\frac{n}{2}\frac{d^2f}{dx^2}\big|_{x_0}}} \underbrace{\int_{-\infty}^{\infty} dx \mathrm{e}^{-x^2}}_{\sqrt{\pi}} \\
&= \sqrt{\frac{2\pi}{-n\frac{d^2f}{dx^2}\big|_{x_0}}}\, \mathrm{e}^{nf(x_0)}
\end{aligned}
\tag{D.4}
$$

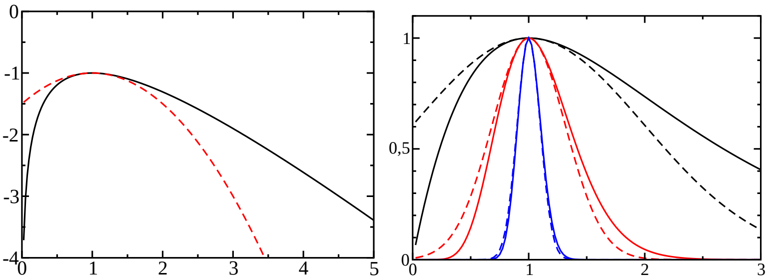


Fig. D.2: Left: The function $\ln[x] - x$ and its Taylor expansion to second order about $x = 1$. Right: the function $exp(n(x\ln[x] - x + 1))$ and its Gaussian approximation (dashed) for n=1,10,100. Note that the accuracy of the Gaussian approximation becomes better with increasing n.

Using $f(x) = \ln(s) - s + 1$ we obtain the position of the maximum at $x_0 = 1$, and $f(1) = 0$, $\partial_x f(x)|_1 = 0$, and $\partial_x^2 f(x)\big|_1 = -1$.

$$
\ln(s) - s + 1 \overset{Taylor}{=} -\frac{1}{2}(s-1)^2 + O(s-1)^3
$$

Thus within the saddlepoint approximation we obtain Eq. D.3

$$
n! = n\mathrm{e}^{n\ln[n]-n} \int_0^{\infty} ds\, \mathrm{e}^{n(\ln[s]-s+1)} \overset{Eq.\ D.4}{\approx} \mathrm{e}^{n\ln[n]-n}\sqrt{2\pi n}
\tag{D.5}
$$

Now we can take the logarithm of the factorial and obtain:

$$
\begin{aligned}
\ln[n!] &= n\ln[n] - n + \frac{1}{2}\ln[n] + \frac{1}{2}\ln[2\pi] \\
&= \left(n + \frac{1}{2}\right)\ln[n] - n + \frac{1}{2}\ln[2\pi]
\end{aligned}
$$

which is the desired result Eq. D.2, from which Eq. D.1 can be readily obtained by ignoring all terms that grow slower with n than n.

D.1.2 Another derivation of Stirling's formula

In the following another derivation of Stirling's formula is shown. We do not follow the original derivation, but use a few, increasingly accurate, approximations. The reason for this is that it demonstrates a few, rather versatile, concepts for numerical integration.

The simple formula

$$\ln[n!] = \ln[1] + \ln[2] + \ldots + \ln[n] = \sum_{i=1}^{n} \ln[i]$$

Now we convert the sum into an integral.

$$\ln[n!] \approx \int_1^n dx\ \ln[x] = \Big[x\ln[x] - x\Big]_1^n = n\ln[n] - n + 1 \approx n\ln[n] - n$$

and obtain the simple result Eq. D.1. This derivation is very simple to memorize.

Correction 1

The critical step Eq. D.6 has been the conversion of a sum into an integral. It is not clear yet if that is legitimate.

Therefore, we show below how the integral can be made more accurate by approximating the integral by piecewise line segments and later on by piece-wise parabolic segments, etc. These calculations will provide us with increasingly better representations of $\ln[n!]$.

We start with the integral, which we approximate by piecewise straight line segments. For each line segment the integral is the average of the function values at the boundaries of the segment multiplied with the size of the interval.

$$\begin{aligned}
\int_1^n dx \ln[x] &\approx \sum_{i=1}^{n-1} \frac{1}{2}\left(\ln(i) + \ln(i+1)\right) \\
&= \frac{1}{2}\ln[1] + \ln[2] + \ln[3] + \ldots + \ln[n-1] + \frac{1}{2}\ln[n] \\
&= \underbrace{\left(\sum_{i=1}^{n} \ln[i]\right)}_{=\ln[n!]} - \frac{1}{2}\underbrace{\ln[1]}_{=0} - \frac{1}{2}\ln[n] \\
\Rightarrow \ln[n!] = \sum_{i=1}^{n} \ln[i] &\approx \left(\int_1^n dx \ln[x]\right) + \frac{1}{2}\ln[n] \\
&= [x\ln[x] - x]_1^n + \frac{1}{2}\ln[n] \\
&= \left(n\ln[n] - n - 1\underbrace{\ln[1]}_{=0} + 1\right) + \frac{1}{2}\ln[n] \\
&= n\ln[n] - n + \frac{1}{2}\ln[n] + 1
\end{aligned}$$

Thus we obtain already a better approximation for $\ln[n!]$, namely

$$\ln[n!] \approx n\ln[n] - n + \frac{1}{2}\ln[n] + 1$$

This result is already closer to Eq. D.2.

Correction 2

Let us estimate the error by calculating a correction for the integral.

$$\int_1^n dx\; \ln[x] \approx \underbrace{\sum_{i=1}^{n-1} \frac{\ln(i)+\ln(i+1)}{2}}_{\ln(n!)-\frac{1}{2}\ln(n)} + \Delta$$

Note that Δ enters with a negative sign in our formula for $\ln[n!]$

$$\ln(n!) = \int_1^n dx\; \ln[x] + \frac{1}{2}\ln(n) - \Delta$$

Instead of approximating the integrand by line-segments, we approximate it now by piece-wise parabolas. The parabola shall be determined by the values at the boundaries of the interval and the curvature in the middle of the interval.

Let us describe the parabola for a interval extending from x_1 to x_1+1 by

$$g(x) = a + b(x-x_1) + c(x-x_1)^2$$

Then the line segment $f(x)$ has the form

$$f(x) = a + (b+c)(x-x_1)$$

Thus the correction for the interval is

$$\begin{aligned}
\int_{x_1}^{x_1+1} dx\; (g(x)-f(x)) &= \int_{x_1}^{x_1+1} dx\; \big(a + b(x-x_1) + c(x-x_1)^2 - a - (b+c)(x-x_1)\big) \\
&= \int_{x_1}^{x_1+1} dx\; \big(c(x-x_1)^2 - c(x-x_1)\big) \\
&= c\int_0^1 dx\; \big(x^2 - x\big) = c\left[\frac{1}{3}x^3 - \frac{1}{2}x^2\right]_0^1 = -\frac{c}{6}
\end{aligned}$$

The value of c is obtained by equating the second derivative of g with that of the logarithm.

$$\begin{aligned}
\left.\frac{d^2g(x)}{dx^2}\right|_{x_1+\frac{1}{2}} &= \left.\frac{d^2\ln(x)}{dx^2}\right|_{x_1+\frac{1}{2}} \\
\Rightarrow 2c &= -\frac{1}{(x_1+\frac{1}{2})^2} \qquad \Rightarrow \qquad c = \frac{-1}{2(x_1+\frac{1}{2})^2}
\end{aligned}$$

Thus we obtain the correction for the interval as

$$\int_{x_1}^{x_1+1} dx\; (g(x)-f(x)) = \frac{1}{4(x_1+\frac{1}{2})^2}$$

The total error is then obtained by summing the correction for all intervals

$$\Delta = \sum_{i=1}^{n-1} \frac{1}{12(i+\frac{1}{2})^2}$$

We can approximate the sum again by an integral, by approximating the integrand by a set of line segments.

$$
\begin{aligned}
\frac{1}{12}\int_1^{n-1} dx \frac{1}{(x+\frac{1}{2})^2} &\approx \left(\sum_{i=1}^{n-1}\frac{1}{12(i+\frac{1}{2})^2}\right) - \frac{1}{2}\left(\frac{1}{12(1+\frac{1}{2})^2}+\frac{1}{12(n-1+\frac{1}{2})^2}\right)\\
\left(\sum_{i=1}^{n-1}\frac{1}{12(i+\frac{1}{2})^2}\right) &\approx \frac{1}{12}\int_1^{n-1} dx\, \frac{1}{(x+\frac{1}{2})^2} + \frac{1}{2}\left(\frac{1}{12(1+\frac{1}{2})^2}+\frac{1}{12(n-1+\frac{1}{2})^2}\right)\\
&= \frac{1}{12}\left[-\frac{1}{x+\frac{1}{2}}\right]_1^{n-1} + \frac{1}{6}\left(\frac{1}{9}+\frac{1}{(2n-1)^2}\right)\\
&= -\frac{1}{6(2n-1)} + \frac{1}{18} + \frac{1}{54} + \frac{1}{6(2n-1)^2}\\
&= \frac{2}{27} - \frac{1}{6(2n-1)} + \frac{1}{6(2n-1)^2}
\end{aligned}
$$

Thus we obtain an error estimate Δ

$$
\Delta \approx \frac{2}{27} - \frac{1}{6(2n-1)} + \frac{1}{6(2n-1)^2}
$$

We can see that the error contains a constant. Similarly if we would try to correct the piecewise parabolic interpolation with higher polynomials and convert the sum into an integral, we would always obtain a constant term from the correction of the lower boundary of the integral. Thus the constant is no more reliable. The terms depending on n on the other hand disappear for large n.

Our best estimate for $\ln[n!]$ is so far

$$
\begin{aligned}
\ln[n!] &\approx n\ln[n] - n + \frac{1}{2}\ln[n] + 1 - \Delta\\
&= (n+\frac{1}{2})\ln[n] - n + \frac{25}{27} + \frac{1}{2(2n-1)} - \frac{1}{2(2n-1)^2}
\end{aligned}
$$

Editor: The proof shown seems to be identical with the Euler-McLaurin Formula. See Arfken und Weber p360.

D.2 The geometric series

$$
\sum_{i=0}^{N-1} q^i = 1 + q + q^2 + \ldots + q^{N-1} = \frac{1-q^N}{1-q}
$$

The equation is proven by multiplication with $1-q$ and comparing each power in q.

$$
\begin{aligned}
(1-q)\sum_{i=0}^{N-1} q^i = 1 \quad &+q + q^2 \ldots + q^{N-1}\\
&-q - q^2 - \ldots - q^{N-1} \; -q^N = 1 - q^N
\end{aligned}
$$

The infinite series converges for $|q| < 1$ and has the value

$$
\sum_{i=0}^{N-1} q^i = \frac{1}{1-q}
$$

D.3 Integral of a Gauss function

Here the identity

$$\int_{-\infty}^{\infty} dx\ \mathrm{e}^{-x^2} = \sqrt{\pi} \tag{D.6}$$

We start out forming the product of two such integrations

$$C^2 = \int_{-\infty}^{\infty} dx\ \mathrm{e}^{-x^2} \times \int_{-\infty}^{\infty} dy\ \mathrm{e}^{-y^2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dx\ dy\ \mathrm{e}^{-(x^2+y^2)}$$

Now we transform to planar polar coordinates (r, ϕ) and use that the angular integral is 2π and that $dxdy = r\ drd\phi$.

$$C^2 = 2\pi \int_0^{\infty} dr\ r\mathrm{e}^{-r^2} = 2\pi\frac{1}{-2}\int_0^{\infty} dr\ \frac{d\mathrm{e}^{-r^2}}{dr} = \pi$$

which proves the above relation

D.4 The integral

Here the equation

$$\int_0^{\infty} dx\sqrt{x}\mathrm{e}^{-x} = \frac{\sqrt{\pi}}{2}$$

is derived.

We first perform a variable transform to $y = \sqrt{x}$.

$$\begin{aligned}
I = \int dx\sqrt{x}\mathrm{e}^{-x} &= 2\int dy y^2\mathrm{e}^{-y^2} \\
&= \int dy y(2y\mathrm{e}^{-y^2}) \\
&= -\int dy y\frac{d}{dy}\mathrm{e}^{-y^2} \\
&= -\int dy\Big[\frac{d}{dy}(y\mathrm{e}^{-y^2}) - (\frac{d}{dy}y)\mathrm{e}^{-y^2}\Big] \\
&= \int dy\mathrm{e}^{-y^2} = \frac{\sqrt{\pi}}{2}
\end{aligned}$$

In the last step we used the integral of a Gauss function as derived in a previous appendix.

D.5 Legendre transform

D.6 Euler equation

The Euler equation says that

$$U(S, V, N) = TS - PV + \mu N \tag{D.7}$$

The Euler equation implies that the total Legendre transform of the internal energy vanishes.

Let us consider a homogeneous system divide it in pieces. In this case the extensive variables scale with the size of the system,

$$U(\lambda S, \lambda V, \lambda N) = \lambda U(S, V, N)$$

while the intensive variables remain constant.

Let us differentiate with respect toλ

$$\frac{\partial U}{\partial S}S+\frac{\partial U}{\partial V}V+\frac{\partial U}{\partial N}N=U(S,V,N)$$

By inserting the definitions of the intensive variables $T=\frac{\partial U}{\partial S}$, $p=-\frac{\partial U}{\partial V}$, $\mu=-\frac{\partial U}{\partial N}$. the Euler equation

$$TS-PV+\mu N=U(S,V,N)$$

is obtained.

D.7 Free energy of the ideal Boltzmann gas

Here we derive the Helmholtz free energy from the partition function $Z(T,V,N)$ derived in the section on the Boltzmann gas. It will then lead to the Sackur-Tetrode equation. However we need to employ here the Stirling's formula, which shows that the entropy of the (TVN ensemble is identical to that of the $TV\mu$ ensemble only in the limit of large N.

The free energy for the ideal Boltzmann gas at a given volume and number of particles is called the Helmholtz free energy[1] , denoted by a symbol A. In physics the common symbol for the Helmholtz free energy is F, whereas chemists have adopted the symbol A. To be more explicit we reserve the symbol F for a generic free energy, whereas for chemical systems such as gases etc., we use the symbols adopted by chemists.

$$\begin{aligned}
A^{BG}(T,V,N) &\overset{Eq.\ 1.61}{=} -k_BT\ln[Z^{BG}(T,V,N)]\\
&\overset{Eq.\ 2.6}{=} -k_BT\ln\left[\frac{1}{N!}\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3N}{2}}V^N\right]\\
&\overset{Stirling}{\approx} -k_BT\left(-N\ln[N]+N+N\ln\left[\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}V\right]\right)\\
&= -Nk_BT\left(-\ln[N]+1+\ln\left[\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}V\right]\right)\\
&= -Nk_BT\left(1+\ln\left[\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}\frac{V}{N}\right]\right)
\end{aligned}\tag{D.8}$$

Note that we had to make an additional approximation by applying Stirling's formula[2]. Thus this result is only accurate in the sense that the free energy per particle is is correct for large numbers of particles.

Interestingly the volume enters only in the form of the average density N/V. The box dimensions do not show up any longer in the free energy. This indicates[3] that our results are independent of the shape of the box.

---

[1] Hermann Ludwig Ferdinand von Helmholtz. German Physicist 1821-1894. Proposed the conservation of energy.

[2] James Stirling. Scottish Scientist 1692-1770.

[3] It does not necessarily prove it!

D.7.1 Energy and Entropy of the ideal Boltzmann gas

First we determine the entropy

$$\begin{aligned}
S^{BG}(T,V,N) &= -\frac{\partial A^{BG}}{\partial T} \\
&= Nk_B\left(1+\ln\left[\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}\frac{V}{N}\right]\right)+Nk_BT\frac{3}{2T} \\
&= Nk_B\left(\frac{5}{2}+\ln\left[\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}\frac{V}{N}\right]\right) \qquad \text{(D.9)}
\end{aligned}$$

Then we use Eq. 1.67 to obtain the internal energy

$$\begin{aligned}
U^{BG} &\overset{Eq.\ 1.67}{=} A^{BG}+TS^{BG} \\
&= -Nk_BT\left(1+\ln\left[\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}\frac{V}{N}\right]\right) \\
&\quad +Nk_BT\left(\frac{5}{2}+\ln\left[\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{\frac{3}{2}}\frac{V}{N}\right]\right) \\
&= \frac{3}{2}Nk_BT \qquad \text{(D.10)}
\end{aligned}$$

This is the so-called **caloric equation of state**.

Later we will need the entropy as function of the extensive variables, which we obtain by using the caloric equation of state Eq. 2.17 to replace the temperature by the internal energy in the expression for the entropy Eq. D.9. This yields the so-called **Sackur-Tetrode Equation**.

$$\begin{aligned}
S^{BG}(U,V,N) &\overset{Eq.\ D.9}{=} Nk_B\left(\frac{5}{2}+\ln\left[\left(\frac{mk_BT(U)}{2\pi\hbar^2}\right)^{\frac{3}{2}}\frac{V}{N}\right]\right) \\
&\overset{Eq.\ 2.17}{=} Nk_B\left(\frac{5}{2}+\ln\left[\left(\frac{mU}{3\pi\hbar^2N}\right)^{\frac{3}{2}}\frac{V}{N}\right]\right) \\
&= Nk_B\left(\frac{5}{2}+\ln\left[\left(\frac{m}{3\pi\hbar^2}\right)^{\frac{3}{2}}\frac{VU^{\frac{3}{2}}}{N^{\frac{5}{2}}}\right]\right) \qquad \text{(D.11)}
\end{aligned}$$

which is the well known Sackur-Tetrode equation Eq. 2.19

D.8 Thermodynamic potentials for a general N-state system with equidistant energy levels

Let us consider an N-state system with equidistant energy eigenvalues $E(x) = E_0 + cx$ for states $x = 0, 1, 2, .., N-1$ with some constant c. For $N = 2$ this model describes a two state system such as an electron spin in a magnetic field. For $N = \infty$ this example corresponds to a quantum mechanical harmonic oscillator, or to a one-dimensional particle in a box.

Let us evaluate first the Free energy

$$\begin{aligned}
F(T) &= -k_BT\ln[\sum_{x=0}^{N-1}\mathrm{e}^{-\beta(E_0+cx)}] = E_0-k_BT\ln[\sum_{x=0}^{N-1}(\mathrm{e}^{-\beta c})^x] \\
&= E_0-k_BT\ln[\frac{1-\mathrm{e}^{-\beta cN}}{1-\mathrm{e}^{-\beta c}}] \\
&= E_0-k_BT\ln[1-\mathrm{e}^{-\beta cN}]+k_BT\ln[1-\mathrm{e}^{-\beta c}]
\end{aligned}$$

We have used that the partition sum contains a geometric series, that can be summed analytically, as shown in the appendix.

The entropy is given by

$$\begin{aligned} S &= -\frac{\partial F}{\partial T} \\ &= k_B\Big[\ln[1-\mathrm{e}^{-\beta Nc}] - \frac{\beta Nc}{1-\mathrm{e}^{\beta Nc}}\Big] - k_B\Big[\ln[1-\mathrm{e}^{-\beta c}] - \frac{\beta c}{1-\mathrm{e}^{\beta c}}\Big] \end{aligned}$$

The internal energy

$$\begin{aligned} U = F + TS &= E_0 - \frac{Nc}{1-\mathrm{e}^{\beta Nc}} + \frac{c}{1-\mathrm{e}^{\beta c}} \\ &= E_0 - c\frac{\mathrm{e}^{-\beta c} - \mathrm{e}^{-\beta c(N+1)} - N\mathrm{e}^{-\beta cN} + N\mathrm{e}^{-\beta c(N+1)}}{1 - \mathrm{e}^{-\beta c} - \mathrm{e}^{-\beta cN} + \mathrm{e}^{-\beta c(N+1)}} \end{aligned}$$

Now we would like to investigate the internal energy at high temperatures, that is $\beta \to 0$. Thus we expand the denominator and the ... in terms of βc, and keep the two lowest order terms.

$$= E_0 - c\frac{\sum_i \frac{1}{i!}(-\beta c)^i[1-(N+1)^i - N^{i+1} + N(N+1)^i]}{\sum_i \frac{1}{i!}(-\beta c)^i[\delta_{0,i} - 1 - N^i + (N+1)^i]}$$

i	$[1-(N+1)^i - N^{i+1} + N(N+1)^i]$	$[\delta_{0,i} - 1 - N^i + (N+1)^i]$
0	0	0
1	0	0
2	$N(N-1)$	$2N$
3	$2N(N^2-1)$	$3N(N+1)$

$$\begin{aligned} U(T\to\infty) &= E_0 + c\frac{\frac{1}{2}N(N-1) - \beta c\frac{1}{3}N(N-1)(N+1)}{N - \beta c\frac{1}{2}N(N+1)} \\ &= E_0 + c\frac{N-1}{2}\frac{1-\beta c\frac{2}{3}(N+1)}{1-\beta c\frac{1}{2}(N+1)} \\ &= E_0 + c\frac{N-1}{2} - \beta c^2\frac{1}{12}(N^2-1) + O(\beta^2)] \end{aligned}$$

As anticipated the internal energy converges to the average value of all energy levels. This reflects that all states are occupied at high temperature with equal probability.

The result given here cannot generalized to $N=\infty$. Here we need to start out from the free energy of the harmonic oscillator

$$F_{N=\infty}(T) = E_0 + k_BT\ln[1-\mathrm{e}^{-\beta c}]$$

D.9 Entropy of electrons and Mermin functional

The entropy of a system is

$$S = k_B \ln X \tag{D.12}$$

where X is the number of accessible states.

Let us now consider a system of n Fermion-s, where each of the N one-particle states can either be occupied with one electron or remain unoccupied. Furthermore, Fermions are indistinguishable.

The ultimate goal would be to answer questions such as the occupation numbers as function of temperature, or to obtain the Free Energy of such a electronic system. The Free Energy of electron states in the band gap related to defects is for example important to predict the free energy of formation of the defect and hence the defect concentration.

We can now count the number of possible states

$$
\begin{aligned}
X(n=1) &= N \\
X(n=2) &= \frac{N(N-1)}{2} \\
X(n=3) &= \frac{N(N-1)(N-2)}{3!} \\
X(n) &= \frac{N!}{n!(N-n)!}
\end{aligned}
\tag{D.13}
$$

$N!/(N-n)!$ is the number of ways distinguishable particles can be arranged into N one-particle states and therefore the number of many-particle states for distinguishable particles. $n!$ is the number of ways the particles can be rearranged among the occupied states. We have to divide by this number because all those many-particle states are equivalent.

We can now use Stirling's Formula

$$
\begin{aligned}
N! &= N^N e^{-N}\sqrt{2\pi N} \approx N^N \mathrm{e}^{-N} \\
\ln N! &= N\ln(N) - N
\end{aligned}
\tag{D.14}
$$

An approximate derivation is as follows:

$$
\begin{aligned}
\ln(N!) &= \ln(1) + \ln(2) + \ldots + \ln(N) \\
&\approx \int_1^N dx \ln(x) + \frac{1}{2}\ln(N) \\
&\approx [N\ln(N) - N] - [1\ln(1) - 1] + \frac{1}{2}\ln(N) \\
&\approx [N\ln(N) - N] + 1 + \frac{1}{2}\ln(N)
\end{aligned}
\tag{D.15}
$$

We come back to counting the number of states in order to evaluate the entropy.

$$
\begin{aligned}
S = k_B \ln X &= k_B\Big([N\ln(N) - N] - [n\ln(n) - n] - [(N-n)\ln(N-n) - (N-n)]\Big) \\
&= k_B\Big(N\ln(N) - n\ln(n) - (N-n)\ln(N-n) \\
&= k_B\Big(N\ln(N) - fN\ln(fN) - (1-f)N\ln((1-f)N)\Big) \\
&= k_B N\Big(\ln(N) - f\ln(f) - f\ln(N) - (1-f)\ln((1-f)) - (1-f)\ln(N)\Big) \\
&= k_B N\Big(-f\ln(f) - (1-f)\ln((1-f))\Big) \\
&= -k_B N\Big(f\ln(f) + (1-f)\ln((1-f))\Big)
\end{aligned}
\tag{D.16}
$$

where $f = n/N$.

The Free energy is

$$
\begin{aligned}
F &= E - TS \\
&= \sum_n f_n\epsilon_n - [\sum_n f_n - N]\mu + k_BT\sum_n\Big(f_n\ln(f_n) + (1-f_n)\ln(1-f_n)\Big)
\end{aligned}
\tag{D.17}
$$

The occupation number f_n are obtained by minimizing the free energy with respect to the occupations.

$$
\begin{aligned}
\frac{dF}{df_n} &= \epsilon_n - \mu + k_BT\Big[\ln(f_n) + 1 - \ln(1-f_n) - 1\Big] \\
&= \epsilon_n - \mu + k_BT\ln(\frac{f_n}{1-f_n}) = 0 \\
\exp(-\frac{\epsilon_n - \mu}{k_BT}) &= \frac{f_n}{1-f_n} \\
\exp(-\frac{\epsilon_n - \mu}{k_BT}) &= \Big[1 + \exp(-\frac{\epsilon_n - \mu}{k_BT})\Big] f_n \\
f_n &= \Big[1 + \exp(\frac{\epsilon_n - \mu}{k_BT})\Big]^{-1}
\end{aligned}
\tag{D.18}
$$

Thus we derived the Fermi distribution function

We can now evaluate the entropy contribution of a state near the Fermi level

$$
\begin{aligned}
S &= \sum_n f_(\epsilon_n) s(\epsilon_n) \\
s(\epsilon) &= -k_B\Big(f\ln(f) + (1-f)\ln((1-f))\Big) \\
&= -k_B\Big(\ln(1-f) + f\ln(\frac{f}{1-f})\Big) \\
&= -k_B\Big(\ln(\frac{\exp(\beta(\epsilon-\mu))}{1+\exp(\beta(\epsilon-\mu))}) - f\frac{\epsilon-\mu}{k_BT}\Big) \\
&= -k_B\Big(\beta(\epsilon-\mu)) - \ln(1+\exp(\beta(\epsilon-\mu))) - f\beta(\epsilon-\mu)\Big) \\
&= -k_B\Big((1-f)\beta(\epsilon-\mu)) - \ln(1+\exp(\beta(\epsilon-\mu)))\Big)
\end{aligned}
\tag{D.19}
$$

Appendix E

Ideal gas

Let us consider indistinguishable particles. We use periodic boundary conditions.

We can obtain the partition function from the density of states.

$$Z(T,V)=\sum_{states}\mathrm{e}^{-\frac{E(i,j,k)}{k_BT}}$$

$$\int d\epsilon D(\epsilon)\mathrm{e}^{-\frac{\epsilon}{k_BT}}$$

The partition function is

$$\begin{aligned}Z(T,V)&=\sum_{i,j,k=-\infty}^{\infty}\exp\Big[-\frac{1}{2mk_BT}(\frac{2\pi\hbar}{L})^2(i^2+j^2+k^2)\Big]\\&=\Big[\sum_{i=-\infty}^{\infty}\exp\Big[-\frac{1}{2mk_BT}(\frac{2\pi\hbar}{L})^2i^2\Big]^3\\&\approx\Big[\sum_{i=-\infty}^{\infty}\exp\Big(-\pi(\frac{\lambda_Ti}{L})^2\Big)\Big]^3\end{aligned}$$

Here we introduced the **thermal de-Broglie wavelength**[1]

$$\lambda_T=\sqrt{\frac{2\pi\hbar^2}{mk_BT}}\qquad\text{(E.1)}$$

The de-Broglie wavelength is a measure of the length scale over which we can detect quantum mechanical discretization of the states...???

If the de-Broglie wavelength is much smaller than the dimensions of the box, namely L, the sum can be approximated by an integral

$$\begin{aligned}Z(T,V)&=\Big[\int_{-\infty}^{\infty}dx\exp\Big(-\pi(\frac{\lambda_Tx}{L})^2\Big)\Big]^3\\&=\Big[\frac{L}{\lambda_T}\int dye^{-\pi y^2}\Big]^3\\&=\frac{V}{\lambda_T^3}\end{aligned}$$

---

[1] Louis Victor Pierre Raymond duc de Broglie. French Physicist 1892-1987. Professor for theoretical Physics at Sorbonne in Paris. Founded the particle-wave duality of matter in his doctoral thesis 1924. Nobel price 1929.

Here we have used that $\int_{-\infty}^{\infty} dx e^{\pi x^2} = 1$, and that the volume is $V = L^3$.

Now we generalize this result from one particle to N indistinguishable particles. To obtain the sum over all N-particle states we have to multiply the partition function N times. However in that case we included every set of of a number of times. For example if we have two one-particle states A and B each occupied with one electron, we included the states A(1)B(2) and the state A(2)B(1), where 1 and 2 denote the two particles. To avoid the "double counting" we have to divide the result by the number of permutations of particles $N!$. We use the relation $N! \approx (N/e)^N$

$$\begin{aligned} Z(T, V, N) &= \frac{1}{N!} Z(T, V, 1)^N \\ &\approx \Big[\frac{\mathrm{e}V}{N\lambda_T^3}\Big]^N \end{aligned}$$

Next we evaluate the Free energy

$$\begin{aligned} F(T, V, N) &= -k_B T \ln Z(T, V, N) \\ &= -N k_B T \ln\Big[\frac{\mathrm{e}V}{N\lambda_T^3}\Big] \\ &= N k_B T \Big[1 - \ln \frac{V}{N\lambda_T^3}\Big] \end{aligned}$$

Given that the de-Broglie wavelength is small compared to the length dimensions, we can ignore the term unity and obtain

$$F(T, V, N) = -N k_B T \ln \frac{V}{N\lambda_T^3} \tag{E.2}$$

From the free energy we can derive the pressure at a given temperature and density

$$p = -\frac{\partial F}{\partial V} = k_B T \frac{N}{V} \tag{E.3}$$

This is nothing but the ideal gas equation

$$pV = N k_B T = NRT \tag{E.4}$$

The gas constant is k_B=R=8.314510 J/(mol K). (Pascal:Pa=Jm^{-3};) The gas constant is simply the Boltzmann constant in different units. Typically one represents $k_B = R/N_A$ where N_A is the Avogadro constant. However if we use the unit $mol = N_A$, this relation is absorbed in the units.

The pressure rises linearly with temperature and with the gas density. However the pressure is independent of the type of particles. This implies that any ideal gas with a fixed number of molecules fills the same volume at a given temperature and pressure. Hence the relative masses of the molecules can be measured, by weighing gases in a container of a given volume. It is the reason why balloons filled with helium to make them fly. Helium atoms is lighter than nitrogen molecules and so is the helium gas lighter than nitrogen gas. Therefore the balloon filled with helium will be lighter than the surrounding air and will fly.

Most gases will actually sink to the bottom.

Air consists in fraction of volume of 78.9 % N_2, 20.95 % O_2, 0.93 % Ar and 0.002 % other Nobel gases. (Furthermore it contains water, Dust, CO_2 and carbohydrides such as CH_4, C_2H_4.)[20] It has therefore a weight density at atmospheric pressure of ..... Only a small number of gases are lighter than air, namely H_2, He, Ne, CH_4, C_2H_2. This is important to know to prevent explosions, because one has to allow the gases to leave a closed space at the bottom. Often it does not help to open the window, unless there is sufficient circulation. Another important example is the accumulation of carbon dioxide CO_2 in wells or corn silos. As there is no exit for the gas at the bottom, CO_2 will accumulate at the bottom of these wells. People entering will suffocate. Also Radon, a radioactive noble gas tends to accumulate in cellars of houses, where it can pose health problems.

(below 16% oxygen is the strength strongly reduced, below 10% unconsciousness can suddenly occur; CO_2 ist narkotisierend und f§hrt bei §ber 7 % zu Bewustlosigkeit. 1L Trockeneis entspricht 500 l CO_2 gas (Erstickungsgefahr) ;N2O ist lachgas)

We can also derive the chemical potential.

$$\begin{aligned}\mu &= -\frac{\partial F}{\partial N} = k_B T \ln \frac{V}{N\lambda_T^3} - k_B T \\ &= -k_B T (\ln[\frac{N}{V}\lambda_T^3] + 1)\end{aligned}$$

The requirement is that the volume per particle must be large compared to the de-Broglie wavelength, allows us to drop the constant term relative to the logarithm.

$$\mu = -k_B T \ln[\frac{N}{V}] + 3 k_B T \ln[\lambda_T] \tag{E.5}$$

Interestingly a small term depending on the mass of the gas molecules occurs. The thermal de-Broglie wavelength of air at room temperature is 0.2 Å, which is much smaller than the average distance of gas molecules at atmospheric pressure is about 35 AA. At atmospheric pressure the de-Broglie wavelength becomes comparable to the de-Broglie wavelength at about 1K. All materials are liquid or solid at this temperature. He^4, a viscous liquid at low temperatures, becomes superfluid at 2 K, that is it forms a Bose condensate.

Let us perform a Lagrange transform to $U(S, V, N)$

$$\begin{aligned}S &= -\frac{\partial F}{\partial T} \\ &= N k_B \ln \frac{V}{N\lambda_T^3} + N k_B T \frac{3}{\lambda_T}\frac{\partial \lambda_T}{\partial T} \\ &= +N k_B \ln \frac{V}{N\lambda_T^3} + N k_B T \frac{3}{\lambda_T}(-\frac{1}{2T}) \\ &= +N k_B \ln \frac{V}{N\lambda_T^3} - \frac{3}{2} N k_B \frac{1}{\lambda_T} \\ &= -N k_B \Big[\frac{3}{2\lambda_T} - \ln \frac{V}{N\lambda_T^3}\Big] \\ -TS &= F(T) + \frac{3}{2} N k_B T \frac{1}{\lambda_T} \\ U &= F + TS = -\frac{3 N k_B T}{2\lambda_T} \\ &= -\frac{3N\sqrt{\pi}}{\hbar\sqrt{m}}(\frac{k_B T}{2})^{\frac{3}{2}}\end{aligned}$$

E.1 Quantities

1 atm	1.01325×10^6 J/m^3
	=6.10193×10^{-4} kJ/mol/Å^3
thermal de-Broglie wavelength of air at 300K	0.2 Å
Mean distance between gas molecules at 300 K and 1atm	35 Å
Mean distance between gas molecules at 300 K and 10^{-8} Torr	15 μm
k$_B T$ at 300 K	2.49435 kJ/mol
	=25.85212 meV
chem. pot. of air at 300K and 1 atm	-0.36 eV

E.2 Aufgaben

E.2.1 Configurational Entropy

Vacant lattice positions, so-called vacancies, in a semiconductor can have an impact on the electronic properties and dopant diffusion. Therefore it is important to estimate the number of lattice vacancies in a semiconductor. The number of lattice positions in Si is estimated from the crystal structure. The cubic unit cell of silicon has a lattice constant of 5.43 Åand contains 8 atoms. Let us assume that the energy to form a vacancy in a crystal is $E_F = 3\ eV$.

What is the fraction of empty lattice sites at 200^{o}C? Determine the free energy as function of the number of vacancies in a given volume and temperature. Determine the number of lattice sites from this free energy.

Help: Consider a lattice with M lattice sites, of which N sites are vacant. Determine the entropy as function of N from the number of configurations with N vacancies. Use Stirling's formula $\ln(n!) \approx n\ln[n] + n$, which is valid for large n.

Answer:

First we determine the partition function as function of number of vacancies

$$Z(T,N) = \sum_i \mathrm{e}^{-\beta E_i}) = \sum_N \mathrm{e}^{-\beta(E(N)-TS(N))}$$

Let us consider a crystal with M sites. A state with N vacancies has the energy NE_f. There are $\frac{M!}{N!(M-N)!}$ different configurations.[2] The number of configurations is related to the entropy by $\mathrm{e}^{S/k_B} = \frac{M!}{N!(M-N)!}$.

$$\begin{aligned}
S &= k_B \ln[\frac{M!}{N!(M-N)!}] = k_B\Big(\ln[M!] - \ln[N!] - \ln[(M-N)!]\Big) \\
&\approx k_B\Big(M\ln[M] - M - N\ln[N] + N - (M-N)\ln[(M-N)] + (M-N)\Big) \\
&= k_B\Big(M\ln[M] - N\ln[N] - (M-N)\ln[M-N]\Big) \\
&= k_B\Big(M\ln[M] - N\ln[N] - (M-N)\ln[M-N]\Big) \\
&= k_B\Big(M\ln[\frac{M}{M-N}] - N\ln[\frac{N}{M-N}]\Big) \\
&= k_B\Big(M\ln[\frac{M}{M-N}] - N\ln[\frac{N}{M-N}]\Big)
\end{aligned}$$

Let us now introduce a new symbol $c \overset{\text{def}}{=} N/M$ for the vacancy concentration.

$$\begin{aligned}
S &= Mk_B\Big(\ln[\frac{1}{1-c}] - c\ln[\frac{c}{1-c}]\Big) \\
&= Mk_B\Big(-\ln[1-c] - c\ln[c] + c\ln[1-c]\Big) \\
&= -Mk_B\Big((1-c)\ln[1-c] + c\ln[c]\Big)
\end{aligned}$$

---

[2] The degeneracy is obtained as follows. I place the first vacancy on one of M positions, the second on one of $M-1$ positions and so on until I have created $M(M-1)\dots(M-N+1) = M!/(M-N)!$ configurations. Each configuration however has been counted $N!$ times. Therefore I divide by $N!$, the number of permutations of the vacancies among each other.

Thus we obtain the entropy per vacancy[3]

$$s(c)\overset{\text{def}}{=}\frac{S}{M}=-k_B\Big((1-c)\ln[1-c]+c\ln[c]\Big)$$

Now we can obtain the free energy

$$\begin{aligned}F_M(T,c)&=-k_BT\ln\Big[\sum_{N=1}^{M}\mathrm{e}^{S(N)/k_B}\mathrm{e}^{-\beta NE_0}\Big]\\&=-k_BT\ln\Big[\sum_{N=1}^{M}\mathrm{e}^{-\beta(NE_0-TS(N))}\Big]\\&=-k_BT\ln\Big[M\int_0^1 dc\mathrm{e}^{-\beta M(cE_0-Ts(c))}\Big]\\&=-k_BT\ln\Big[M\int_0^1 dc\Big(\mathrm{e}^{-\beta(cE_0-Ts(c))}\Big)^M\Big]\end{aligned}$$

For large M only the maximum value of the integrand will contribute. Therefore we will perform the harmonic approximation of the integrand by expanding $Y=cE_0-Ts(c)$ to second order in c about the maximum of Y. First we determine the maximum of Y [4] :

$$\begin{aligned}Y&=cE_0+k_BT\Big((1-c)\ln[1-c]+c\ln[c]\Big)\\\frac{dY}{dc}&=0=E_0+k_BT\Big(-\ln[1-c]-1+\ln[c]+1\Big)=E_0+k_BT\ln[\frac{c}{1-c}]\\\Rightarrow\frac{c}{1-c}&=\mathrm{e}^{-\beta E_0}\qquad\Rightarrow\qquad c=\mathrm{e}^{-\beta E_0}-c\mathrm{e}^{-\beta E_0}\\\Rightarrow c&=\frac{1}{\mathrm{e}^{\beta E_0}+1}\end{aligned}$$

Let us evaluate the value of Y at its maximum

$$\begin{aligned}Y&=cE_0-Ts(c)=cE_0+k_BT\Big((1-c)\ln[1-c]-c\ln[c]\Big)\\&=\frac{E_0}{1+\mathrm{e}^{\beta E_0}}+k_BT\Big(-\frac{\ln[1+\mathrm{e}^{-\beta E_0}]}{1+\mathrm{e}^{-\beta E_0}}\frac{\ln[1+\mathrm{e}^{\beta E_0}]}{1+\mathrm{e}^{\beta E_0}}\Big)\end{aligned}$$

Now we need the second derivative of Y at its maximum.

Thus we obtain

$$\begin{aligned}F_M&=-k_BT\ln\Big[M\int_0^\infty dc\Big(\mathrm{e}^{-\beta(Y(c_0)+\frac{1}{2}Y''(c-c_0)^2)}\Big)^M\Big]\\&=-k_BT\ln\Big[M\int_0^\infty dc\Big(\mathrm{e}^{-\beta(Y(c_0)+\frac{1}{2}Y''(c-c_0)^2)}\Big)^M\Big]\end{aligned}$$

---

[3] Note here that this expression is identical to the entropy of a one-particle state due to filling with electrons. In this entropy contribution the concentration c is replaced by the fractional occupation.

[4] The resulting expression for the concentration is formally identical with the Fermi distribution function. This finding is not unexpected since only one vacancy can occupy a site, just as for Fermions. **Editorial remark: this is the exact Pendent to the photon gas: chemical potential vanishes**

Appendix F

Philosophy of the ΦSX series

The ΦSX series tries to implement a few rules that I learned mostly from the feedback given to me by the students attending the course and that relied on this books as background material.

The course should be self-contained. There should not be any statements "as shown easily..." if, this is not true. The reader should not need to rely on the author, but he should be able to convince himself, if what is said is true. I am trying to be as complete as possible in covering all material that is required. The basis is the mathematical knowledge. With few exceptions the material is also developed in a sequence so that the material covered can be understood entirely from the knowledge covered earlier.

The derivations shall be explicit. The novice should be able to step through every single step of the derivation with reasonable effort. An advanced reader should be able to follow every step of the derivations even without paper and pencil.

All units are explicit. That is, formulas contain all fundamental variables, which can be inserted in any desirable unit system. Expressions are consistent with the SI system, even though I am quoting some final results in units, that are common in the field.

The equations that enter a specific step of a derivation are noted ontop of the equation sign. The experience is that the novice does not immediately memorize all the material covered and that he is struggling with the math, so that he spends a lot of time finding the rationale behind a certain step. This time is saved by being explicit about it. The danger that the student, gets dependent on these indications, is probably minor, as it is still some effort for the advanced reader to look up the assumptions, rather than memorizing the relevant material.

Important results and equations are included in boxes. This should facilitate the preparations for examinations.

Portraits of the key researchers and short biographical notes provide independent associations to the material. A student may not memorize a certain formula directly, but a portrait. From the portrait, he may associate the correct formula. The historical context provides furthermore an independent structure to organize the material.

The two first books are in german (That is the intended native language) in order to not add complications to the novice. After these first books, all material is in English. It is mandatory that the student masters this language. Most of the scientific literature is available only in English. English is currently the language of science, and science is absolutely dependent on international contacts.

I tried to include many graphs and figures. The student shall become used to use all his senses in particular the visual sense.

I have slightly modified the selection of the material. Some topics, which I consider of mostly historical relevance I have removed. Others such as the Noether theorem, I have added. Some, like Chaos, Stochastic processes, etc. I have not added yet.

Appendix G

About the author

Prof. Dr. rer. nat Peter E. Blöchl studied physics at Karlsruhe University of Technology in Germany. Subsequently he joined the Max Planck Institutes for Materials Research and for Solid state Research in Stuttgart, where he worked on development of electronic-structure methods related to the LMTO method and on first-principles investigations of interfaces. He received his doctoral degree in 1989 from the University of Stuttgart.

Following his graduation, he joined the renowned T.J. Watson Research Center in Yorktown Heights, NY in the US on a World Trade Fellowship. In 1990 he accepted an offer from the IBM Zurich Research Laboratory in Ruschlikon, Switzerland, which had just received two Nobel prices in Physics (For the Scanning Tunneling Microscope in 1986 and for the High-Temperature Superconductivity in 1987). He spent the summer term 1995 as visiting professor at the Vienna University of Technology in Austria, from where was awarded the habilitation in 1997. In 2000 he left the IBM Research Laboratory after a 10-year period and accepted an offer to be professor for theoretical physics at Clausthal University of Technology in Germany. Since 2003, Prof. Blöchl is member of the Braunschweigische Wissenschaftliche Gesellschaft (Academy of Sciences).

The main thrust of Prof. Blöchl's research is related to ab-initio simulations, that is the parameter-free simulation of materials processes and molecular reactions based on quantum mechanics. He developed the Projector Augmented Wave (PAW) method, one of the most widely used electronic structure methods to date. Next to the research related to simulation methodology, his research covers a wide area from biochemistry, solid state chemistry to solid state physics and materials science. Prof. Bloechl contributed to 8 Patents and published about 100 research publications, among others in well-known Journals such as "Nature". The work of Prof. Blöchl has been cited over 10,000 times, and he has an H-index of 37.

Appendix H

Greek alphabet

A	α	alpha	N	ν	nu
B	β	beta	Ξ	ξ	ksi
$\mathit{\Gamma}$	γ	gamma	O	$o,$	omicron
Δ	δ	delta	Π	π, ϖ	pi
E	ϵ, ε	epsilon	P	ρ, ϱ	rho
Z	ζ	zeta	Σ	σ, ς	sigma
H	η	eta	T	τ	tau
Θ	θ, ϑ	theta	Υ	υ	upsilon
I	ι	iota	Φ	ϕ, φ	phi
K	κ	kappa	X	χ	chi
Λ	λ	lambda	Ψ	ϕ	phi
M	μ	mu	Ω	ω	omega

Appendix I

A small Dictionary

acceptor	Akzeptor
attempt	Versuch
cam shaft	Nockenwelle
chamber	Kammer
combinatorics	Kombinatorik
combustion	das Brennen
combustion chamber	Brennkammer
compartment	Abteil
composite	zusammengesetzt
constraint	Zwangsbedingung
crank shaft	Kurbelwelle
denominator	Nenner
die; pl. dice	Würfel
donor	Donator
dopant atom	Dotieratom
ensemble	Gesamtheit
exert	ausüben
exhaust valve	Auslassventil
extrinsic	extrinsisch
extensive	extensiv
factorial	Fakultät ($n!$)
forecast	Vorhersage
frequency	Häufigkeit; Frequenz

.

cont'd

hallmark	Markenzeichen
handwaving	grob, vereinfacht
inlet valve	Einlassventil
intensive	intensiv
intersection (of sets)	Vereinigung (von Mengen)
intersection (of streets)	Kreuzung (von Straßen)
intrinsic	intrinsisch, eigenleitend
law of mass action	Massenwirkungsgesetz
heat	Wärme
intrinsic	intrinsisch
moment of inertia	Trägheitstensor
multiplier	Multiplikator
numerator	Zähler
number representation	Besetzungszahldarstellung
partition function	Zustandssumme
piston	Kolben
Planck's law	Planck's Strahlungsgesetz
reaction parameter	Reaktionslaufzahl
set	Menge (math.)
shallow	oberflächlich, seicht
toy	Spielzeug
transition state	Übergangszustand
trace	Spur
ubiquitous	allgegenwärtig
union (of sets)	Vereinigung (von Mengen)
valve	Ventil
Wien's displacement law	Wiensches Verschiebungsgesetz
work	Arbeit

Bibliography

[1] Harald Cramer. *The Elements of Probability Theory and some of its applications*. John Wiley and Sons New York, 1954.

[2] C.E. Shannon. A mathematical theory of computation. *Bell. Syst. Techn. J.*, 22:379, 1948.

[3] E.T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620, 1957.

[4] E.T. Jaynes. Information theory and statistical mechanics ii. *Phys. Rev.*, 108:171, 1957.

[5] D. Chowdhury and D. Stauffer. *Principles of Equilibrium Statistical Mechanics*. Wiley-VCH, 2000.

[6] J. Schnakenberg. *Thermodynamik und Statistische Physik*. Wiley-VCH, 2002.

[7] J. Clerk Maxwell. *Theory of Heat*. Longmanns Green and Co, 1871.

[8] J. von Neumann. *Mathematische Grundlagen der Quantenmechanik*, volume 38 of *Grundlehren der mathematischen Wissenschaften*. Springer, 1932.

[9] S. Nose. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81:511, 1984.

[10] S. Nose. A molecular dynamics method for the simulation in the canonical ensemble. *Mol. Phys.*, 52:255, 1984.

[11] S. Chapman. Boltzmann's h-theorem. *Nature*, 139:931, 1937.

[12] A. Wehrl. General properties of entropy. *Rev. Mod. Phys.*, 50:221, 1978.

[13] N.W. Ashcroft and N.D. Mermin. *Solid State Physics*. Thomson Learning, 1987.

[14] G.J. Ehnholm, J.P. Eckström, J.F. Jacquinot, M.T. Loponen, O.V. Lounasmaa, and J.K. Soini. Evidence for nuclear antiferromagnetism in copper. *Phys. Rev. Lett*, 42:1702, 1979.

[15] H. Haken. *Quantenfeldtheorie des Festkörpers*. B.G.Teubner, Stuttgart, 1973.

[16] B.K. Agarwal. *Statistical mechanics*. New Age international Publishers, New Dehli, 1988,1998.

[17] Max Planck. Ueber das gesetz der energieverteilung im normalspectrum. *Ann. Phys.*, 309:553, 1901.

[18] Global warming art project. `http://www.globalwarmingart.com/`.

[19] P.J. Mohr and B.N. Taylor. Codata recommended values of the fundamental physical constants: 1989. *Rev. Mod. Phys.*, 72:351, 2000.

[20] Communication by: Linde technische gase.

Index