

Introduction to Statistical Physics

G. M. Pastor

Institut für Theoretische Physik

Fachbereich Mathematik und Naturwissenschaften

Universität Kassel

April 25, 2022

The purpose of these lectures is to study the fundamental principles of Statistical Mechanics, also known as Statistical Physics, which govern the behavior of macroscopic systems, i.e., systems composed by a large number of atoms or molecules. Although large macroscopic systems are subject to the same fundamental laws of quantum mechanics as small systems, their behavior follows new universal laws which are specific to the presence of a very large number of degrees of freedom and to the boundary conditions imposed by experiment. Instead of becoming increasingly intricate and obscure, new regularities appear as the number of particles becomes very large which escape a purely mechanical deterministic interpretation. These statistical laws cannot be reduced to purely mechanical laws. New concept and principles are required in order to derive them, which do not apply to small systems.

One of our first goals will be to understand why the laws of quantum mechanics (or classical mechanics) are not enough for understanding macroscopic systems and to lay down the principles of statistical mechanics as clearly and explicitly as possible. Many introductory courses and a number of remarkable books on the theory of macroscopic systems (e.g., K. Huang, *Statistical Mechanics*, Wiley) follow the historical development. They start by formulating the first and second laws of the thermodynamics and develop in length their consequences by establishing relations between their mechanic magnitudes. Important concepts like entropy, heat and temperature are introduced in an ad hoc phenomenological way. The concepts of Statistical Mechanics are introduced later in order to provide a microscopic justification or rationalization of the thermodynamic principles. This is the pathway followed by R. Clausius (1822-1888), J.C. Maxwell(1831-1879), L.E. Boltzmann (1844-1906) and J.W. Gibbs (1839-1903). While such an evolution is understandable at a time when no microscopic theory of matter was available, approaching the subject in this way would be an anachronism. At present it is far more meaningful

to set first of all the basis of Statistical Mechanics taking advantage of our knowledge of classical and quantum mechanics and to derive thermodynamics as a byproduct.

This approach, which will be adopted in the following, has a number of important advantages. From a conceptual perspective it is important to state the principles of the theory of macroscopic systems in a clear transparent way, so that we understand, where they are applicable, where they are not and what kind of consequences to expect in borderline cases (e.g., nanoscale systems). In this way we set the focus on the general fundamental theory rather than on the phenomenological consequences of it. Statistical mechanics is free from a number of limitations of thermodynamics. For example, thermodynamics concerns systems in equilibrium leaving no room for thermal fluctuations of the physical observables. It provides no microscopic explanation of the observed behavior but simply general relations between different measurable properties. Since no connection to quantum mechanics can be established in thermodynamics, it is not possible to predict any observable property (e.g., the melting point of a given metal or the Curie temperature of ferromagnet). The equation of state of a material in a specific phase can never arise from thermodynamics. Statistical mechanics shares with thermodynamics its universal character, since its principles and laws apply to all systems alike. But in addition, it opens the possibility of a detailed theory of matter including the calculation material-specific properties. Subtle quantum-mechanical effects such as superconductivity, magnetism, metal insulation transitions and all other condensed-matter properties can thus be predicted. In fact, the rigorous bridge between quantum theory and macroscopic behavior has only been established thanks to Statistical Mechanics. This allows a detailed comparison between microscopic theory and macroscopic experiments, the cornerstone of scientific progress.

Any theory of the properties of macroscopic systems has to be based, or at least be consistent, with quantum mechanics (QM), which is our microscopic theory of matter. QM defines very precise methods for determining the eigenstates and eigenvalues of different observables, in particular the stationary states and eigenenergies. This is certainly not an easy task for systems containing of the order of 10^{23} atoms. In addition, QM predicts the time evolution of any given state $|\Psi\rangle$. The same holds for a classical description, whenever it is applicable. However, an atomistic deterministic mechanical theory is not able to predict in which state a system will be found under specified macroscopic constraints such as volume and total energy. Answering this question is one of the main goals of Statistical Mechanics. According to deterministic dynamics the state of any system is unambiguously defined by the initial conditions and the elapsed time. Moreover, different initial conditions yield different microscopic states (time-inversion symmetry). In contrast, from a statistical perspective an uncountable number of different microscopic initial states lead to the same macroscopic equilibrium variables

such as temperature and pressure. A simple example, which I borrowed from Erwin Müller-Hartmann, allows to clearly illustrate the different scopes of deterministic and statistical mechanics: Consider a large number of H₂O molecules in a given volume. A typical statement of QM would be that water and ice are possible states of the system. Based on this quantum mechanical result statistical mechanics allows us to conclude that in summer one finds water and in winter ice.

G. M. Pastor, WS 13/14, Kassel.

Contents

1	Basic statistical concepts	7
1.1	Random variables and probabilities	7
1.2	A single continuous random variable	8
1.2.1	Moments of a PDF	9
1.2.2	Cumulants of a PDF*	10
1.2.3	Computation of moments in terms of cumulants*	11
1.3	The Gaussian or normal distribution	13
1.4	Many random variables	14
1.4.1	Joint cumulants of independent random variables*	15
1.5	The Gaussian distribution for many variables*	16
1.6	Probability distribution of a function on N random variables	18
1.7	The central limit theorem	18
1.8	Information content of a probability distribution	22
1.9	Entropy of a probability distribution	24
1.9.1	Inferring unbiased probability distributions	25
1.9.2	Entropy of continuous probability densities	26
2	Mixed states, density matrices and distribution functions	30
2.1	Introduction	30
2.2	Mixed states	31
2.2.1	Statistically independent systems	31
2.2.2	A single quasi-closed system and its environment	32
2.3	Density matrix	34
2.4	Time dependence of the density operator of isolated systems	38
2.5	The statistical distribution function	41
2.6	Time dependence of $\rho(p, q)$: Liouville theorem	42
2.6.1	Total time derivative of classical observables	43
2.6.2	The continuity equation in n dimensions	44
2.6.3	Time dependence of $\rho(p, q, t)$	45
2.6.4	A note on ergodicity and relaxation times	46
2.7	Entropy conservation in strictly isolated classical systems	47

3	Equilibrium statistical ensembles	49
3.1	Statistical independence of macroscopic subsystems	49
3.2	The statistical independence of extensive additive properties	50
3.3	The importance of additive constants of motion	52
3.4	The density operator: General formulation	54
3.4.1	The microcanonical ensemble	54
3.4.2	The grand canonical ensemble	56
3.4.3	The canonical ensemble	57
3.5	Explicit forms of the density operator	58
3.5.1	Microcanonical ensemble	59
3.5.2	Canonical ensemble	60
3.5.3	Grand Canonical ensemble	63
3.5.4	Grand Canonical pressure ensemble	68
3.6	Implicit notation	69
4	Entropy	70
4.1	Maximum entropy theorem	72
4.2	The approach to equilibrium	74
4.2.1	The master equation	75
4.2.2	Time dependence of the entropy	76
4.2.3	A simple example of irreversible entropy growth	79
4.3	Thermodynamic equivalent descriptions	79
4.3.1	Euler theorem for homogeneous functions	86
4.4	Thermodynamic potentials: General formulation	87
4.4.1	Legendre transformations	91
5	Thermodynamic properties	94
5.1	Thermodynamic potentials: special cases	94
5.1.1	Energy E and Entropy S	94
5.1.2	Helmholtz free energy F	95
5.1.3	Enthalpy H	97
5.1.4	Free enthalpy G	97
5.1.5	Grand canonical potential Φ	98
5.1.6	Deriving properties in the grand canonical pressure ensemble	98
5.2	Derived thermodynamic properties	99
5.2.1	Heat capacities and specific heats	99
5.2.2	Compressibilities	100
5.2.3	Thermal expansion	100
5.2.4	Charge compressibility	101
5.2.5	Assessing fluctuations	101
5.3	Minimum free-energy theorem	103
6	Thermodynamic relations	106
6.1	Duhem-Gibbs relations for thermodynamic potentials	106

6.2	Intensive nature of the derivatives of extensive properties	106
6.3	Integrability of the differential of thermodynamic potentials	107
6.4	Jacobi-determinant manipulations	110
6.5	Measuring the absolute temperature scale	113
7	Thermodynamic processes	116
7.1	Carnot cycle	118
7.2	Joule Thompson process	120
7.3	Adiabatic expansion in vacuum	122
8	The Nernst theorem: The Third law of thermodynamics	124
8.1	Some experimental consequences	125
8.2	Approaching $T = 0$ in experiment	126
9	The principle of indistinguishability of identical particles	129
9.1	Many-particle wave functions	131
9.2	Fermions	133
9.3	Bosons	134
10	The classical limit	140
10.1	Boltzmann counting	142
10.2	Configurational Integrals	143
10.3	Virial and equipartition theorems	143
10.4	The ideal classical gas	145
10.5	The ideal classical gas in the pressure ensemble	147
10.6	Derivation of the classical limit	149
11	The ideal quantum gases	157
12	Fermi systems: The ideal Fermi gas	163
12.1	Energy-pressure relation	165
12.2	High temperatures and low densities	168
12.3	Low temperatures and high densities	169
12.3.1	The Sommerfeld expansion	170
12.3.2	Physics of the low temperature limit	173
13	Bose systems: Photons and black body radiation	180
13.1	Basic properties of the quantized electromagnetic field	181
13.2	The Planck distribution	183
13.3	The energy-pressure relation	184
13.4	Thermodynamic properties	185
13.5	Revisiting Planck's law	188
13.6	Phonons	190
14	Bose-Einstein condensation	192

1 Basic statistical concepts

A full mechanical description of the dynamics of a macroscopic system is both hopeless and not very meaningful, since we lack the required precise information on the initial conditions of all degrees of freedom and on the exact form of the Hamiltonian governing the system and its environment. Moreover, for describing the equilibrium properties a precise knowledge of all its constituents is not necessary at all. What we actually need to know is only the probability of finding the macroscopic system in each one of its possible microscopic states. It is the goal of statistical mechanics to provide such an inherently probabilistic description of macroscopic systems. This motivates a brief discussion of some basic concepts of probability theory.

The sections marked with an asterisk (*) in this chapter are not needed for understanding the rest of the notes. They may be skipped without consequences.

1.1 Random variables and probabilities

Variables whose outcome cannot be predicted with certainty are usually known as *random variables*. We consider a random variable x with a *set of possible outcomes* S , which may be discrete (e.g., $S = \{x_1, x_2, \dots\}$) or continuous (e.g., $S \equiv \mathbb{R}$).

An *event* E is a subset of outcomes $E \subseteq S$, for example, $E = \{\text{even result of a dice throw}\} = \{2, 4, 6\}$. To each event $E \subseteq S$ we assign a *probability* $P(E)$ with the following three fundamental properties:

- i) Positive definiteness: $P(E) \geq 0$.
- ii) Additivity: $P(E_1) + P(E_2) = P(E_1 \cup E_2)$ if $E_1 \cap E_2 = \emptyset$ and $E_1, E_2 \subseteq S$.
 E_1 and E_2 are said to be disconnected events.
- iii) Normalization: $P(S) = 1$.

Probabilities may be assigned in two ways:

- i) Experimentally as

$$P(E) = \lim_{N \rightarrow \infty} \frac{N_E}{N},$$

where N_E is the number of actual occurrences of the event E after N “throws” or outcomes.

- ii) Theoretically by means of an estimation, which is based on the determination of the set of outcomes S , and some hypothesis about the relative probabilities for a complete set of events. For instance, knowing that $S = \{1, 2, 3, 4, 5, 6\}$ for a dice and assuming equal probabilities $P(i) = 1/6 \quad \forall i \in S$, we conclude that

$P(\text{even}) = 3 \cdot 1/6 = 1/2$. Due to the lack of knowledge of the precise mechanical properties of the *dice* (i.e., the system) and the *way of throwing it* (i.e., the initial condition and the Hamiltonian including the environment) and in the absence of any reason to believe that the dice is biased, we assume that all six possibilities of the elements in S (i.e., the states of the system) are equally probable.

All probability assignments in statistical mechanics are theoretical or “subjective” (as opposed to “objective” in i). Their validity needs to be verified by contrasting them to experiment.

1.2 A single continuous random variable

We consider a continuous random variable $x \in \mathbb{R}$. The *cumulative probability function* (CPF) is defined as the probability $P(x)$ for any outcome x' to be *smaller* than x :

$$P(x) = \text{prob}(x' \in [-\infty, x]) .$$

The basic properties of a CPF are

- i) $P(-\infty) = 0$.
- ii) $P(x)$ is monotonically increasing, i.e.,

$$P(x + \Delta) = P(x) + \text{prob}(x' \in [x, x + \Delta]) \geq P(x),$$

since any probability satisfies additivity and positiveness.

- iii) Finally, the normalization condition implies $P(+\infty) = 1$.

The *probability density function* (PDF) is defined as $p(x) = \frac{dP}{dx}$. Consequently,

$$p(x) dx = \text{prob}(x' \in [x, x + dx]) .$$

Notice that, in contrast to probabilities satisfying $P(x) \leq 1$, there is no upper bound for $p(x)$.

Very often we are interested in analyzing the statistical behavior of properties which depend on random variables. If x is a random variable, any function $F(x)$ of x is also a random variable with its own PDF which is given by

$$p_F(f) df = \text{prob}(F(x) \in [f, f + df]) .$$

Let x_i with $i = 1, \dots, \nu$ be the solutions of $F(x_i) = f$, we have

$$\begin{aligned} p_F(f) df &= \sum_{i=1}^{\nu} p(x_i) dx_i \\ \Rightarrow p_F(f) &= \sum_{i=1}^{\nu} p(x_i) \left| \frac{dx}{dF} \right|_{x=x_i=F^{-1}(f)}. \end{aligned} \tag{1.1}$$

Notice that $|dx/dF|$ is the Jacobian for the change of variables. Eq. (1.1) may be written as

$$p_F(f) = \sum_{i=1}^{\nu} p(x_i = F^{-1}(f)) \frac{1}{\left| \frac{dF}{dx} \right|_{x=x_i=F^{-1}(f)}}. \quad (1.2)$$

Such a change of variables may lead to divergencies in $p_F(f)$ which remain of course integrable, as one may easily verify by changing variables back to x .

Exercise 1.1:

A metastable nucleus decays through β emission. What is the probability density $p(\varphi, \theta)$ for the electron to be emitted with a polar angle θ (relative to the z axis) and azimuthal angle φ ? Imagine the nucleus at the coordinate origin and consider

$$p(\varphi, \theta) d\varphi d\theta = \text{prob} \{ \text{emission with angle } \theta \in [\theta, \theta + d\theta] \text{ and angle } \varphi \in [\varphi, \varphi + d\varphi] \}.$$

Exercise 1.2: Consider the Gaussian distribution

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-x^2/2\sigma^2}$$

with $x \in \mathbb{R}$ and $f = F(x) = x^2$. Show that $p_F(f) = \frac{1}{\sqrt{2\pi}\sigma^2} \frac{e^{-f/2\sigma^2}}{\sqrt{f}}$ for $f > 0$ and $p_F(f) = 0$ for $f < 0$. Verify the normalization of $p_F(f)$.

1.2.1 Moments of a PDF

The *expectation value* of a function $F(x)$ of the random variable x is given by the known expression

$$\langle F(x) \rangle = \int_{-\infty}^{+\infty} F(x) p(x) dx.$$

Particularly important expectation values are the *moments of the PDF*

$$\langle x^n \rangle = \int x^n p(x) dx$$

and the *characteristic function* $\chi(k)$

$$\chi(k) = \langle e^{-ikx} \rangle = \int e^{-ikx} p(x) dx,$$

which is just the *Fourier transform* of the PDF. Note that $\chi(0) = 1$.

The characteristic function is the *generator of the moments*:

$$\begin{aligned}\chi(k) &= \left\langle \sum_{n=0}^{+\infty} \frac{(-ik)^n}{n!} x^n \right\rangle = \sum_{n=0}^{+\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle, \\ &= 1 - ik \langle x \rangle - \frac{1}{2} k^2 \langle x^2 \rangle + \dots\end{aligned}\tag{1.3}$$

from which we obtain

$$i^n \left. \frac{d^n \chi(k)}{dk^n} \right|_{k=0} = \langle x^n \rangle.$$

The PDF can be recovered from $\chi(k)$ by the inverse transformation

$$p(x) = \frac{1}{2\pi} \int e^{ikx} \chi(k) dk.\tag{1.4}$$

One can also easily obtain the moments around any other point x_0 from

$$e^{ikx_0} \chi(k) = \langle e^{-ik(x-x_0)} \rangle = \sum_{n=0}^{+\infty} \frac{(-ik)^n}{n!} \langle (x-x_0)^n \rangle.$$

It is clear that the knowledge of *all* the moments of a PDF defines it univocally, since $\langle x^n \rangle$ defines $\chi(k)$, from which $p(x)$ can be obtained. See Eqs. (1.3) and (1.4).

1.2.2 Cumulants of a PDF*

The logarithm of the characteristic function $\chi(k)$ is known as the *cumulant generating function*

$$\begin{aligned}\ln \chi(k) &= \sum_{n=1}^{+\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle_c \\ &= -ik \langle x \rangle_c - \frac{1}{2} k^2 \langle x^2 \rangle_c + \dots\end{aligned}\tag{1.5}$$

The cumulants are defined implicitly by means of the previous series expansion of $\ln \chi(k)$. Note that $\ln \chi(k=0) = \ln 1 = 0$.

The cumulants and the moments are of course related, since they derive from the same $p(x)$ and $\chi(k)$. One can obtain the cumulants from the following comparison. On the one hand we have

$$\chi(k) = 1 + \underbrace{\sum_{n=1}^{+\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle}_{\varepsilon},\tag{1.6}$$

where $\varepsilon = -i k \langle x \rangle - \frac{1}{2} k^2 \langle x^2 \rangle + \dots$. And on the other we have

$$\begin{aligned} \ln \chi(k) &= \ln(1 + \varepsilon) = \sum_{l=1}^{+\infty} (-1)^{l+1} \frac{\varepsilon^l}{l} \\ &= \varepsilon - \frac{\varepsilon^2}{2} + \dots \end{aligned} \quad (1.7)$$

Using the implicit definition (1.5) of the cumulants and replacing ε from (1.6) in (1.7) we have

$$-i k \langle x \rangle_c - \frac{1}{2} k^2 \langle x^2 \rangle_c + \dots = -i k \langle x \rangle - \frac{1}{2} k^2 \underbrace{(\langle x^2 \rangle - \langle x \rangle^2)}_{=\sigma^2} + \dots, \quad (1.8)$$

or in a general form valid for all n ,

$$\sum_{n=1}^{+\infty} \frac{(-i k)^n}{n!} \langle x^n \rangle_c = \sum_{l=1}^{+\infty} \frac{(-1)^{l+1}}{l} \left(\sum_{m=1}^{+\infty} \frac{(-i k)^m}{m!} \langle x^m \rangle \right)^l.$$

This leads to

$$\langle x \rangle_c = \langle x \rangle \quad \text{mean}$$

$$\langle x^2 \rangle_c = \langle x^2 \rangle - \langle x \rangle^2 \quad \text{variance}$$

$$\langle x^3 \rangle_c = \langle x^3 \rangle - 3 \langle x^2 \rangle \langle x \rangle + 2 \langle x \rangle^3 \quad \text{skewness}$$

$$\langle x^4 \rangle_c = \langle x^4 \rangle - 4 \langle x^3 \rangle \langle x \rangle - 3 \langle x^2 \rangle^2 + 12 \langle x^2 \rangle \langle x \rangle^2 - 6 \langle x \rangle^4 \quad \text{kurtosis (or kurtosis)}.$$

A PDF can be described indistinctively in terms of its cumulants or of its moments.

1.2.3 Computation of moments in terms of cumulants*

Theorem: The m -th moment $\langle x^m \rangle$ is obtained by considering all possible subdivisions of m points in p_n groups or connected clusters of n points each. Of course $\sum_n p_n n = m$. Each possible subdivision contributes with the product of the cumulants $\langle x^n \rangle_c$ associated with the connected cluster having n points.

Examples:

$$\begin{aligned} \langle x \rangle &= (\bullet) & &= \langle x \rangle_c \\ \langle x^2 \rangle &= (\bullet)(\bullet) + (\bullet\bullet) & &= \langle x \rangle_c^2 + \langle x^2 \rangle_c \\ \langle x^3 \rangle &= (\bullet)(\bullet)(\bullet) + 3(\bullet)(\bullet\bullet) + (\bullet\bullet\bullet) & &= \langle x \rangle_c^3 + 3 \langle x \rangle_c \langle x^2 \rangle_c + \langle x^3 \rangle_c \end{aligned}$$

Exercise 1.3:

Obtain the expression for $\langle x^4 \rangle$ in terms of $\langle x^l \rangle_c$ and $l = 1-4$. Deduce $\langle x^l \rangle_c$ in terms of $\langle x^l \rangle$ for $l \leq 4$.

The theorem can be demonstrated by noting that

$$\begin{aligned}\chi(k) &= \sum_{m=0}^{+\infty} \frac{(-ik)^m}{m!} \langle x^m \rangle = \exp \{ \ln \chi(k) \} \\ &= \exp \left\{ \sum_{n=1}^{+\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle_c \right\} \\ &= \prod_{n=1}^{+\infty} \exp \left\{ \frac{(-ik)^n}{n!} \langle x^n \rangle_c \right\} \\ &= \prod_{n=1}^{+\infty} \left[\sum_{p_n=0}^{+\infty} \frac{(-ik)^n p_n}{(n!)^{p_n}} \frac{\langle x^n \rangle_c^{p_n}}{p_n!} \right].\end{aligned}$$

Matching the coefficients of the powers of $(ik)^m$ with all the possibilities yielding $\sum_n p_n = m$ we have

$$\frac{\langle x^m \rangle}{m!} = \sum_{\{p_n\}} \prod_n \frac{\langle x^n \rangle_c^{p_n}}{p_n! (n!)^{p_n}} \quad (1.9)$$

where the sum runs over all the possibilities of forming subgroups with $\sum_{p_n} n = m$. After rewriting Eq. (1.9) as

$$\langle x^m \rangle = \sum_{\{p_n\}} \prod_n \left(\frac{m!}{p_n! (n!)^{p_n}} \langle x^n \rangle_c^{p_n} \right)$$

we can identify the different variables and factors as follows:

n : number of points in one cluster.

p_n : number of clusters with the same number n of points inside.

$m!$: number of permutations of all the m points.

$(n!)^{p_n}$: permutations of the points within each cluster.

$p_n!$: number of permutations of the clusters with n points among them.

$\frac{m!}{p_n! (n!)^{p_n}}$: number of ways of splitting m points in $\{p_n\}$ subgroups with n points each.

1.3 The Gaussian or normal distribution

The Gaussian distribution is given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\lambda)^2}{2\sigma^2}},$$

where $\lambda = \langle x \rangle$ is the mean value and $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$ is the variance. The characteristic function

$$\chi(k) = \frac{1}{\sqrt{2\pi\sigma^2}} \int dx e^{-\frac{(x-\lambda)^2}{2\sigma^2} - ikx}$$

has also a Gaussian form. In order to prove it we set $\xi = x - \lambda$ and rewrite the exponent as

$$\begin{aligned} \frac{(x-\lambda)^2}{2\sigma^2} + ikx &= \frac{\xi^2}{2\sigma^2} + ik\xi + ik\lambda \\ &= \frac{(\xi + i\sigma^2 k)^2}{2\sigma^2} + \frac{\sigma^2 k^2}{2} + ik\lambda. \end{aligned}$$

One obtains

$$\begin{aligned} \chi(k) &= e^{-ik\lambda - \frac{k^2\sigma^2}{2}} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} d\xi e^{-\frac{(\xi + i\sigma^2 k)^2}{2\sigma^2}}}_{=1} \\ &= e^{-ik\lambda - \frac{k^2\sigma^2}{2}}. \end{aligned}$$

In the last step we have applied Cauchy's residue theorem and the fact that the exponential function is analytic in all \mathbb{C} . The cumulant generating function is thus simply given by

$$\ln \chi(k) = -ik\lambda - \frac{k^2\sigma^2}{2}, \quad (1.10)$$

which implies

$$\begin{aligned} \langle x \rangle_c &= \langle x \rangle = \lambda, \\ \langle x^2 \rangle_c &= \langle x^2 \rangle - \langle x \rangle^2 = \sigma^2, \end{aligned}$$

and

$$\langle x^n \rangle_c = 0 \quad \text{for } n \geq 3.$$

This makes the calculations using the cluster expansion particularly simple, since the graphical expansions involve only one- and two-point clusters (see Sec. 1.2.3). Conversely, if all the higher cumulants $\langle x^n \rangle_c$ with $n \geq 3$ vanish (or can be asymptotically neglected) the underlying PDF is a Gaussian (or tends asymptotically to it). This property will be exploited in the derivation of the central limit theorem.

1.4 Many random variables

For more than one variable $\vec{x} = (x_1, \dots, x_N) \in \mathbb{R}^n$ the set of outcomes $S \subseteq \mathbb{R}^n$. The *joint probability distribution function (PDF)* is defined by

$$p(\vec{x}) \prod_{i=1}^N dx_i = \text{prob} \{ \text{event } \vec{x}' \text{ in } x_i < x'_i < x_i + dx_i \quad \forall i \} .$$

It represents the probability per unit volume in \mathbb{R}^n that the random variable \vec{x}' takes a value \vec{x} in a volume element $d\vec{x} = \prod_i dx_i$ centered at \vec{x} and satisfies the normalization condition

$$\int p(\vec{x}) d^N x = 1.$$

If, and only if, the variables are *statistically independent* we have $p(\vec{x}) = \prod_{i=1}^N p_i(x_i)$, where $p_i(x)$ is the PDF of the random variable x_i . In this case one also says that the variables x_i are uncorrelated.

The *unconditional probability density* for a subset of random variables x_1, \dots, x_m is given by

$$p(x_1 \dots x_m) = \int p(x_1, \dots, x_m, x_{m+1}, \dots, x_N) \prod_{i=m+1}^N dx_i.$$

It describes the behavior of the variables x_1, \dots, x_m irrespective of all the others. For instance $p(\vec{x}) = \int d^3 v p(\vec{x}, \vec{v})$ gives the *particle density* (i.e., probability distribution for the position) irrespective of the velocity \vec{v} .

The *conditional PDF* $p(x_1 \dots x_m | x_{m+1} \dots x_N)$ describes the behavior of some variables $x_1 \dots x_m$, subject to the constraint that the other variables $x_{m+1} \dots x_N$ have specified values. For example, one may search for the velocity distribution at a given point \vec{x} , which we denote by $p(\vec{v} | \vec{x})$.

The joint probability is given by

$$p(x_1 \dots x_m, x_{m+1} \dots x_N) = p(x_{m+1} \dots x_N) p(x_1 \dots x_m | x_{m+1} \dots x_N),$$

where $p(x_{m+1} \dots x_N)$ is the unconditional probability density for $x_{m+1} \dots x_N$, irrespective of the other variables $x_1 \dots x_m$, and $p(x_1 \dots x_m | x_{m+1} \dots x_N)$ is the probability of $x_1 \dots x_m$ given the values $x_{m+1} \dots x_N$. Thus

$$p(x_1 \dots x_m | x_{m+1} \dots x_N) = \frac{p(x_1 \dots x_N)}{p(x_{m+1} \dots x_N)},$$

where $p(x_1 \dots x_N)$ is the number of events $x_1 \dots x_m, x_{m+1} \dots x_N$ (divided by the number N of trials) and $p(x_{m+1} \dots x_N)$ is the number of events $x_{m+1} \dots x_N$ (divided by the

number N of trials). An example would be the velocity distribution at point \vec{x} , namely, $p(\vec{v}|\vec{x}) = p(\vec{x}, \vec{v})/p(\vec{x})$.

The *expectation values* of some function $F(\vec{x})$ is calculated as usual from

$$\langle F(\vec{x}) \rangle = \int p(\vec{x}) F(\vec{x}) d^N x.$$

Thus, the *joint characteristic function* is given by the Fourier transform

$$\begin{aligned} \chi(\vec{k}) &= \int d\vec{x} e^{-i\vec{k}\cdot\vec{x}} p(\vec{x}) \\ &= \langle e^{-i\sum_{j=1}^N k_j x_j} \rangle \end{aligned}$$

and the *joint cumulant generating function* is $\ln \chi(\vec{k})$. In the particular case of independent variables we have $p(\vec{x}) = \prod_{i=1}^N p_i(x_i)$, which implies $\chi(\vec{k}) = \prod_{i=1}^N \chi_i(k_i)$ and $\ln \chi(\vec{k}) = \sum_{i=1}^N \ln \chi_i(k_i)$, where $\chi_i(k)$ is the characteristic function of the probability distribution of the variable i .

The *joint moments* and *joint cumulants* are then obtained from

$$\langle x_1^{n_1} \dots x_N^{n_N} \rangle = \frac{\partial^{n_1}}{\partial(-i k_1)} \dots \frac{\partial^{n_N}}{\partial(-i k_N)} \chi(\vec{k}) \Big|_{\vec{k}=0}$$

and

$$\langle x_1^{n_1} * \dots * x_N^{n_N} \rangle_c = \frac{\partial^{n_1}}{\partial(-i k_1)} \dots \frac{\partial^{n_N}}{\partial(-i k_N)} \ln \chi(\vec{k}) \Big|_{\vec{k}=0}.$$

The *graphical relation* between moments and cumulants, that was demonstrated for one variable, also applies to N variables. For instance,

$$\langle x_1 x_2 \rangle = \binom{\bullet}{1} \binom{\bullet}{2} + \binom{\bullet \bullet}{1 \ 2} = \langle x_1 \rangle_c \langle x_2 \rangle_c + \langle x_1 * x_2 \rangle_c$$

or

$$\begin{aligned} \langle x_1^2 x_2 \rangle &= \binom{\bullet}{1} \binom{\bullet}{2} \binom{\bullet}{1} + \binom{\bullet \bullet}{1 \ 1} \binom{\bullet}{2} + 2 \binom{\bullet \bullet}{1 \ 2} \binom{\bullet}{1} + \binom{\bullet \bullet \bullet}{1 \ 2 \ 1} \\ &= \langle x_1 \rangle_c^2 \langle x_2 \rangle_c + \langle x_1^2 \rangle_c \langle x_2 \rangle_c + 2 \langle x_1 * x_2 \rangle_c \langle x_1 \rangle_c + \langle x_1^2 * x_2 \rangle_c. \end{aligned}$$

1.4.1 Joint cumulants of independent random variables*

It is easy to see that $\langle x_\alpha x_\beta \rangle = \langle x_\alpha \rangle \langle x_\beta \rangle$ and $\langle x_\alpha * x_\beta \rangle_c = 0$, if x_α and x_β are *independent random variables*. Let the PDF be of the form

$$p(\vec{x}) = p_1(x_1 \dots x_m) p_2(x_{m+1} \dots x_N),$$

where $1 \leq \alpha \leq m$ and $m + 1 \leq \beta \leq N$. Then

$$\begin{aligned}\chi(\vec{k}) &= \int d\vec{x} e^{-i\vec{k}\cdot\vec{x}} p(\vec{x}) \\ &= \langle e^{-i \sum_{j=1}^m k_j x_j} \rangle_1 \langle e^{-i \sum_{j=m+1}^N k_j x_j} \rangle_2 \\ &= \chi_1(\vec{k}_1) \chi_2(\vec{k}_2).\end{aligned}$$

The joint moment is given by

$$\langle x_\alpha x_\beta \rangle = \frac{\partial \chi_1}{\partial(-i k_\alpha)} \frac{\partial \chi_2}{\partial(-i k_\beta)} = \langle x_\alpha \rangle_1 \langle x_\beta \rangle_2$$

for $1 \leq \alpha \leq m$ and $m + 1 \leq \beta \leq N$. It follows that

$$\ln \chi(\vec{k}) = \ln \chi_1(\vec{k}_1) + \ln \chi_2(\vec{k}_2)$$

and consequently

$$\langle x_\alpha * x_\beta \rangle_c = \frac{\partial}{\partial k_\alpha} \frac{\partial}{\partial k_\beta} \ln \chi(\vec{k}) = 0$$

if $1 \leq \alpha \leq m$ and $m + 1 \leq \beta \leq N$. The joint cumulant $\langle x_\alpha * x_\beta \rangle_c$ is also known as the *connected correlation*. It can be regarded as a measure of the correlation or mutual dependence between the variables x_α and x_β .

1.5 The Gaussian distribution for many variables*

The generalization to N variables of the normal distribution has the form

$$p(\vec{x}) = \frac{1}{\sqrt{(2\pi)^N \det(\Sigma)}} \exp \left\{ -\frac{1}{2} \sum_{mn} (\Sigma^{-1})_{mn} (x_m - \lambda_m) (x_n - \lambda_n) \right\},$$

where Σ is a positive-definite symmetric matrix and Σ^{-1} refers to its inverse. Note that Σ^{-1} is also positive definite. In other words, the argument of the exponential is an arbitrary positive-definite quadratic form.

The characteristic function is given by

$$\chi(\vec{k}) = \int d\vec{x} e^{-i\vec{k}\cdot\vec{x}} e^{-\frac{1}{2}(\vec{x}-\vec{\lambda})\cdot\Sigma^{-1}(\vec{x}-\vec{\lambda})} \frac{1}{\sqrt{(2\pi)^N \det(\Sigma)}},$$

where we have introduced $\vec{k} = (k_1, \dots, k_N)$ and $\vec{\lambda} = (\lambda_1, \dots, \lambda_N)$.

One may easily verify the normalization of $p(\vec{x})$ and compute $\chi(\vec{k})$ by changing variables to $\vec{y} = \vec{x} - \vec{\lambda}$, so that the Gaussian distribution is centered at the origin of

the coordinate system, and by performing an orthogonal transformation U such that $U^t \Sigma^{-1} U = \sigma_m^{-2} \delta_{mn}$ is diagonal (Σ^{-1} is symmetric). The Jacobian of the orthogonal transformation being equal to 1 ($\det U = 1$) and denoting the eigenvalues of Σ^{-1} by $1/\sigma_m^2 > 0$ we have

$$(U^t \Sigma^{-1} U)_{mn} = \delta_{mn} \frac{1}{\sigma_m^2} \quad \text{with} \quad U^t U = \mathbb{1}$$

and

$$(U^t \Sigma U)_{mn} = \delta_{mn} \sigma_m^2.$$

Setting $U \vec{\xi} = \vec{y} = (\vec{x} - \vec{\lambda})$ we have

$$\chi(\vec{k}) = \frac{e^{-i \vec{k} \cdot \vec{\lambda}}}{\sqrt{(2\pi)^N \det(\Sigma)}} \int d\vec{\xi} e^{-i \vec{k} \cdot U \vec{\xi}} \exp \left\{ -\frac{1}{2} \underbrace{(U \vec{\xi}) \cdot \Sigma^{-1} U \vec{\xi}}_{\vec{\xi} \cdot U^t \Sigma^{-1} U \vec{\xi}} \right\}.$$

If we set for a moment $\vec{k} = 0$ to verify the normalization, we see that the integral splits in a product of N one-dimensional Gaussians each yielding an integral $\sqrt{2\pi\sigma_m^2}$, so that $\prod_m \left(\sqrt{2\pi\sigma_m^2} \right) = \sqrt{(2\pi)^N \det(\Sigma)}$. The joint PDF is therefore properly normalized.

In order to compute $\chi(\vec{k})$ one can use the result for one-dimensional Gaussians for $\vec{k}' = U^t \vec{k}$ noting that $(U^t \Sigma^{-1} U)_{mn} = \frac{\delta_{mn}}{\sigma_m^2}$. In this way one has

$$\begin{aligned} \chi(\vec{k}) &= e^{-i \vec{k} \cdot \vec{\lambda}} \prod_m e^{-\frac{k_m'^2 \sigma_m^2}{2}} \\ &= e^{-i \vec{k} \cdot \vec{\lambda}} \exp \left\{ -\frac{1}{2} \sum_m k_m'^2 \sigma_m^2 \right\}. \end{aligned}$$

Using that

$$\begin{aligned} \sum_m k_m'^2 \sigma_m^2 &= \sum_m k_m' \sigma_m^2 k_m' = \vec{k}' \cdot U^t \Sigma U \vec{k}' = \\ &= U^t \vec{k} \cdot U^t \Sigma U \vec{k} = \vec{k} \cdot \Sigma \vec{k} = \sum_{mn} \Sigma_{mn} k_m k_n, \end{aligned}$$

we finally obtain

$$\begin{aligned} \chi(\vec{k}) &= e^{-i \vec{k} \cdot \vec{\lambda} - \frac{1}{2} \vec{k} \cdot \Sigma \vec{k}} \\ &= \exp \left\{ -i \sum_m k_m \lambda_m - \frac{1}{2} \sum_{mn} \Sigma_{mn} k_m k_n \right\}. \end{aligned}$$

Consequently,

$$\ln \chi(\vec{k}) = -i \sum_m k_m \lambda_m - \frac{1}{2} \sum_{mn} \Sigma_{mn} k_m k_n,$$

which implies

$$\begin{aligned} \langle x_m \rangle_c &= \lambda_m, \\ \langle x_m * x_n \rangle_c &= \Sigma_{mn}, \end{aligned}$$

and all higher cumulants vanish.

In the special case of vanishing mean values, i.e., $\lambda_m = 0 \forall m$, we have that all odd cumulants vanish. Thus, all odd moments vanish and any even moment is given by the sum of the products of cumulants obtained from all possible ways of forming pairs of variables. For instance,

$$\begin{aligned} \langle x_\alpha x_\beta x_\gamma x_\delta \rangle &= \langle x_\alpha * x_\beta \rangle_c \langle x_\gamma * x_\delta \rangle_c + \\ &+ \langle x_\alpha * x_\gamma \rangle_c \langle x_\beta * x_\delta \rangle_c \\ &+ \langle x_\alpha * x_\delta \rangle_c \langle x_\beta * x_\gamma \rangle_c. \end{aligned}$$

This is analogous to Wick's theorem in many-body Green's function theory.

1.6 Probability distribution of a function on N random variables

We consider an arbitrary function $f : \mathbb{R}^N \rightarrow \mathbb{R}$ of N random variables $\vec{x} = (x_1, \dots, x_N)$. The PDF $p(f)$ for the variable $f = f(x_1, \dots, x_N)$ is given by

$$p(f') = \int dx_1 \dots dx_N p(x_1, \dots, x_N) \delta[f' - f(x_1, \dots, x_N)], \quad (1.11)$$

where the delta function ensures that the integral sums up only the probability density at the points \vec{x} for which the function $f(\vec{x})$ yields the value f' . In other words, the right hand side can be interpreted as the sum over all possible values of the random variable \vec{x} of the product of probability density of the variable \vec{x} times the conditional probability $p(f'|\vec{x})$ that the random variable $f(\vec{x})$ takes the value f' for the given \vec{x} . The latter is given by the delta function, since it vanishes for all $f' \neq f(\vec{x})$ and its integral over all f' is equal to 1. The reader may wish to apply Eq. (1.11) to a one-variable function $f(x)$ and thus recover Eq. (1.2).

1.7 The central limit theorem

We consider the average $\bar{x} = \frac{1}{N} \sum_{\nu=1}^N x_\nu$ of N random variables x_1, \dots, x_N having the joint PDF $p(x_1, \dots, x_N)$. These variables can be, for example, the kinetic energy $p_\nu^2/2m$

of the atoms in a classical gas or the pair interactions $w(\vec{r}_i, \vec{r}_j)$ between pairs of molecules in liquid ($\nu \equiv ij$). In these cases the average \bar{x} represents the average kinetic energy, respectively interaction energy, of the system. The PDF for the random variable \bar{x} is obtained by integrating over all possible values of x_1, \dots, x_N as

$$p_{\bar{x}}(x) = \int dx_1 \dots dx_N p(x_1, \dots, x_N) \delta\left(x - \frac{1}{N} \sum_{\nu=1}^N x_\nu\right). \quad (1.12)$$

The corresponding characteristic function is

$$\begin{aligned} \chi_{\bar{x}}(k) &= \int dx p_{\bar{x}}(x) e^{-ikx} \\ &= \int dx e^{-ikx} \int dx_1 \dots dx_N p(x_1, \dots, x_N) \delta\left(x - \frac{1}{N} \sum_{\nu} x_\nu\right) \\ &= \int dx_1 \dots dx_N p(x_1, \dots, x_N) e^{-i \frac{k}{N} \sum_{\nu=1}^N x_\nu} \\ &= \chi_p\left(k_1 = \frac{k}{N}, \dots, k_N = \frac{k}{N}\right), \end{aligned} \quad (1.13)$$

where $\chi_p(\vec{k})$ is the characteristic function of the joint probability distribution $p(x_1 \dots x_N)$. Let us recall that $\chi_p: \mathbb{R}^N \rightarrow \mathbb{R}$. Based on Eq. (1.13) one may already observe that for a large number of variables N the properties of $\chi_{\bar{x}}(k)$ and thus of $p_{\bar{x}}(x)$ are given by the behavior of $\chi_p(\vec{k})$ for \vec{k} near the origin. They could thus be assessed by an appropriate Taylor expansion.

Let us now assume that the x_ν are *independent variables* having all the same PDF $p_1(x)$. A memorable example of random variables, which can be considered to be independent, is the kinetic energy the particles in a classical gas at low densities, so that the interactions among the particles can be neglected; the problem considered by Maxwell and Boltzmann in their kinetic theory of gases. Other commonplace examples are the sex of the next N children to be born, the outcomes of throwing a dice, the party voters will chose in the next general election, the velocity of cars in a highway when traffic density is very low, etc. As you can see, no assumption is made neither on the nature of the random variable nor on the form discrete or continuous of the probability distribution $p_1(x)$. Once statistical independence is granted the following considerations are universal. Our purpose is to infer how the probability distribution $p_{\bar{x}}(x)$ of the average \bar{x} of N such uncorrelated variables looks like.

Concerning the joint PDF, statistical independence implies that it splits in the simple product

$$p(x_1, \dots, x_N) = \prod_{\nu=1}^N p_1(x_\nu), \quad (1.14)$$

where the index 1 in $p_1(x_\nu)$ indicates that only one variable is involved. The characteristic function associated with $p(\vec{x})$ is therefore

$$\chi_p(\vec{k}) = \prod_{i=1}^N \chi_1(k_i), \quad (1.15)$$

where $\chi_1(k)$ is the characteristic function of $p_1(x)$. The characteristic function $\chi_{\bar{x}}(k)$ for the PDF $p_{\bar{x}}(x)$ of \bar{x} takes then the form

$$\chi_{\bar{x}}(k) = \prod_{\nu=1}^N \left(\int dx_\nu p_1(x_\nu) e^{-i \frac{k}{N} x_\nu} \right) = \left[\chi_1 \left(\frac{k}{N} \right) \right]^N,$$

where χ_1 is the characteristic function of $p_1(x)$. This result can be obtained either directly from the definition of $\chi_{\bar{x}}(k)$ or by applying Eq. (1.13) to the particular case (1.15). The corresponding cumulant generating function then reads

$$\ln \chi_{\bar{x}}(k) = N \ln \chi_1 \left(\frac{k}{N} \right). \quad (1.16)$$

We can now expand χ_1 or $\ln \chi_1$ for small k/N , i.e., for large N , in terms of the cumulants of the probability distribution for one variable as

$$\ln \chi_1 \left(\frac{k}{N} \right) = -i \frac{k}{N} \langle x \rangle_c - \frac{1}{2} \left(\frac{k}{N} \right)^2 \langle x^2 \rangle_c + O(N^{-3}). \quad (1.17)$$

Combining Eqs. (1.16) and (1.17) we finally have

$$\begin{aligned} \ln \chi_{\bar{x}}(k) &= -i k \langle x \rangle_c - \frac{1}{2} k^2 \frac{\langle x^2 \rangle_c}{N} + O(N^{-2}) \\ &= -i k \langle x \rangle - \frac{1}{2} k^2 \frac{\sigma^2}{N} + O(N^{-2}). \end{aligned}$$

Comparing this equation with the general expression 1.5 for $\ln \chi$ in terms of the cumulants one concludes that the average of \bar{x} is given by $\langle \bar{x} \rangle = \langle x \rangle$ and that its variance is given by $\sigma_{\bar{x}}^2 = \langle \bar{x}^2 \rangle_c = \langle x^2 \rangle_c / N = \sigma^2 / N$. That the averages coincide was certainly expected. Far more remarkable is the fact that the variance of \bar{x} is reduced by a factor $1/N$ with respect to the variance σ^2 of the probability distribution $p_1(x)$ of a single event. In particular $\sigma_{\bar{x}}^2$ vanishes for $N \rightarrow \infty$!

Fourier-transforming back in order to obtain $p_{\bar{x}}(x)$ we have asymptotically for large N that

$$p_{\bar{x}}(x) \cong \frac{1}{\sqrt{2\pi\sigma_{\bar{x}}^2}} \exp \left\{ -\frac{1}{2\sigma_{\bar{x}}^2} (x - \langle x \rangle)^2 \right\}$$

with a variance $\sigma_{\bar{x}}^2 = \sigma^2 / N$. Thus, in the limit of $N \rightarrow \infty$, the probability density of the average always follows a Gaussian distribution whose variance tends to zero. This

important result is known as the *central limit theorem*. Notice that the theorem holds independently of the form of the PDF of the random variable x , i.e., for any $p_1(x)$. The details of the physics behind $p_1(x)$ are therefore irrelevant. In particular, it holds for both continuous and discrete $p_1(x)$ alike. On the one hand, this stresses the universality of the theorem and of the statistical approach for large number of variables. On the other, however, it also implies that the distribution $p_{\bar{x}}(x)$ for large N gives no information on the behavior of the single-event probability $p_1(x)$ from which \bar{x} is issued. Only $\langle x \rangle$ and σ^2 can be inferred from $p_{\bar{x}}(x)$, provided that N is finite and known.

The reasoning used to derive $p_{\bar{x}}(x)$ also applies to the *sum of N independent random variables* $X = \sum_{\nu} x_{\nu}$. In this case the average is $\langle X \rangle = N\langle x \rangle$ and the variance is $\sigma_X^2 = \langle X^2 \rangle_c = N\langle x^2 \rangle_c = N\sigma^2$, so that the amplitude of the fluctuations around the mean, as measured by the standard deviation $\sigma_X = \sqrt{\langle X^2 \rangle_c} = \sqrt{N}\sigma$, scales with \sqrt{N} . The fact that the fluctuations of X increase monotonously with N is certainly reasonable. However, notice that the increase of σ_X is much slower than the average itself ($\langle X \rangle = N\langle x \rangle$). As a result, the relative fluctuation vanishes for large N as $\frac{\sigma_X}{\langle X \rangle} \sim \frac{1}{\sqrt{N}} \rightarrow 0$ for $N \rightarrow \infty$, provided that $\langle X \rangle \neq 0$. This behavior is sometimes referred to as self-averaging.

The consequences of the central limit theorem are far reaching, not only in physical applications but in statistics in general. Consider an arbitrary macroscopic system, for example, a gas, liquid or solid which is not isolated and can thus exchange energy with its environment. As a result of the interactions with the environment the total energy E of the system will fluctuate to some extent, i.e., it will not have a well defined sharp value. The central limit theorem tells us how the probability distribution for the average energy per atom looks like *independently of the details of the interactions between the system and its environment*. Indeed, if the system is macroscopic, it contains of the order of 10^{20} atoms and molecules. One may then divide it in a large number N of equal subvolumes, which still contain a very large number of atoms and can thus be regarded as essentially uncorrelated (e.g., $N = 10^{10}$ subsystems each containing 10^{10} atoms). The probability of finding the subsystems $i = 1, \dots, N$ in some particular quantum states q_1, q_2, \dots, q_N is then given, to a very good approximation, by the product of probabilities $p(q_1, q_2, \dots, q_N) = \prod_{i=1}^N p_1(q_i)$, where $p_1(q)$ is the probability that a subsystem is in the state q . In equilibrium, $p_1(q)$ is the same for all subsystems since homogeneity can be assumed. This reasoning applies to the physics of macroscopic systems in general. The hypothesis of the central limit theorem are then fulfilled.

Let us turn our attention to the probability distribution of the average energy per atom $\bar{\varepsilon}$ and total energy E in such an open macroscopic system. The energies per atom ε_i of the different subsystems $i = 1, \dots, N$ are independent random variables following some probability distribution $p_1(\varepsilon)$. The energy per particle of the whole system is given by the average $\bar{\varepsilon} = \sum_i \varepsilon_i / N$ of the random variables ε_i . In the macroscopic limit both the number N of subsystems and the size of them tend to infinity. The central limit theorem implies —without any knowledge of the physics behind ε_i or $p_1(\varepsilon)$ — that the energy per particle of any macroscopic system in equilibrium is normal distributed with

a mean-square deviation which tends to zero as $\sigma_\varepsilon \propto 1/\sqrt{N}$ in the macroscopic limit, i.e., when the number of subsystems N or system size tends to infinity. Concerning the total energy E one concludes that it also follows a Gaussian distribution with a standard deviation $\sigma_E \propto \sqrt{N}$. Note that the fluctuations of the total energy grow much more slowly with system size than E , which is proportional to N , rendering the fluctuations of the energy per atom negligible. The same reasoning applies to any other extensive property such as the number of particles, volume, or magnetization. The fact that these general conclusions are drawn independently of the nature of the system and of any details about its internal interactions (gas, liquid, solid, metallic, insulating, etc.) demonstrates the amazing universality of the statistical approach.

Exercise 1.4:

Calculate $\langle \bar{x} \rangle$ and $\sigma_{\bar{x}}^2 = \langle \bar{x}^2 \rangle - \langle \bar{x} \rangle^2$ in terms of $\langle x \rangle$ and σ^2 starting from Eqs. (1.12) and (1.14) for arbitrary, not necessarily small N .

1.8 Information content of a probability distribution

Consider a random variable with a discrete set of outcomes $S = \{x_i, i = 1, \dots, M\}$ having the probabilities $\{p_i\}$. Suppose we construct a message $x_1 \dots x_N$ with N independent outcomes of the random variable x_i . We intend to quantify the possible information content of such a message as a function of the probability distribution $\{p_i, i = 1, \dots, M\}$. Analyzing the number of different possible messages will allow us to infer how much of the apparent information content of the message is already contained in the probability distribution. For instance, if the probability distribution is $\{p_1 = 1 \text{ and } p_i = 0 \ \forall i > 1\}$ there is just one possible message ($x_1 \dots x_1$) and actually no information can be conveyed. All the information is in the probability distribution (PD). In the other extreme case, where x_i is uniformly distributed, the PD carries no information at all.

Let us first consider the case where the values of x_i in the message $x_1 \dots x_N$ can be chosen at will. Since there are M possibilities for each x_i , the number of different messages is $g = M^N$. The number of bits K necessary to transmit or store such messages or, in other words, the number of bits required in order to distinguish one message from the others, is

$$K = \ln_2 g = N \ln_2 M. \quad (1.18)$$

On the other hand, if the x_i are taken from the probability distribution p_i , the possible choices of x_i are limited. For instance, if $p_1 \gg p_2$ it is unlikely to construct a message with more occurrences of x_2 than x_1 . In the limit of large number of message elements N , the number of occurrences of x_i in the message approaches asymptotically $N_i = p_i N$. In fact the probability of finding $|N_i - N p_i| > \sqrt{N_i}$ becomes exponentially small as $N \rightarrow \infty$. Taking into account that the message contains nearly N_i occurrences of x_i , the

number of possible messages is reduced to

$$g = \frac{N!}{\prod_{i=1}^M N_i!}.$$

This corresponds to the number of different ways of arranging the N_1, \dots, N_M occurrences of the outcomes $x_1 \dots x_M$, where $\sum_i N_i = N$. To specify the message we therefore need

$$\begin{aligned} K &= \ln_2 g = \ln_2 N! - \sum_{i=1}^M \ln_2 N_i! \\ &\cong N \ln_2 N - N \ln_2 e - \sum_{i=1}^M (N_i \ln_2 N_i - N_i \ln_2 e) \\ &= -N \sum_{i=1}^M \frac{N_i}{N} \ln_2 \frac{N_i}{N} \\ &= -N \sum_{i=1}^M p_i \ln_2 p_i = -N \langle \ln_2 p \rangle \end{aligned}$$

bits of information.¹ K gives a measure of the diversity of the probability distribution $\{p_i\}$. As expected, we recover here the two limiting cases discussed above: $\ln_2 g = 0$ for $p_i = 1$ and $p_j = 0 \forall j \neq i$ and $\ln_2 g = N \ln_2 M$ for $p_i = 1/M \forall i$ (uniform distribution). For any non-uniform probability distribution the information content of the message $\ln_2 g$ is smaller than $N \ln_2 M$, which is the information content in the absence of any information on the relative probabilities p_i . One assigns this difference to the information carried by the probability distribution $\{p_i\}$. Consequently, the *information content of the probability distribution* $\{p_i\}$ is given by

$$\begin{aligned} I(\{p_i\}) &= \frac{1}{N} [K_{\max} - K(\{p_i\})] \\ &= \ln_2 M + \sum_{i=1}^M p_i \ln_2 p_i. \end{aligned}$$

The information content represents the reduction of the logarithm of the number of possible messages g that can be constructed out of $\{p_i\}$ in the limit of large N per unit of message length. It is a property of the probability distribution and as such it is independent N .

¹In the limit of large N one may approximate $\ln N! \simeq N \ln N - N + \ln(2\pi N)/2$. Thus, taking into account only the leading contributions, one may write $\ln N! \simeq N \ln N - N$ for $N \rightarrow \infty$.

1.9 Entropy of a probability distribution

In this context it is useful to introduce the *entropy* S of a probability distribution, which is defined by

$$S(\{p_i\}) = - \sum_{i=1}^M p_i \ln p_i = - \langle \ln p_i \rangle \geq 0 \quad (0 \leq p_i \leq 1).$$

In terms of S the information contents is given by

$$I(\{p_i\}) = \frac{1}{\ln(2)} [S_{\max} - S(\{p_i\})],$$

where

$$S_{\max} = - \sum_{i=1}^M \frac{1}{M} \ln \frac{1}{M} = \ln M$$

is the maximum of $S(\{p_i\})$ corresponding to $p_i = 1/M$. A probability distribution with maximum entropy carries the least possible information. Therefore, S gives a measure of diversity of the distribution. Notice that S is equal to the logarithm of the number of different messages or states that can be constructed with outcomes $x_1 \dots x_M$ satisfying $N_i = Np_i$. For the distribution $p_i = \delta_{ij}$ (for some j) there is only one possible microscopic configuration or message, namely, $(x_j x_j \dots x_j)$. In this case, and only in this case, we have $S = 0$.

The entropy does not depend on the values or physical meaning of the random variables x_i . Any one-to-one mapping of the form $x_i \rightarrow f_i$ leaves the entropy unchanged since $p(x_i) = p(f_i)$. This implies in particular that the (non-equilibrium) entropy of a system of interacting particles (e.g., an interacting electron gas) with occupation probabilities $n_{\vec{k}}$ for each quasi-particle state \vec{k} is the same as the entropy of a non-interacting system *having the same* $n_{\vec{k}}$. The actual equilibrium entropy will of course be different, since in equilibrium the entropy corresponds to the maximum value of $S(\{n_{\vec{k}}\})$ compatible with the constraint of a fixed total energy E .

In contrast, any many-to-one mapping reduces the entropy of the probability distribution, since it reduces its diversity or, in other words, it increases the definiteness or the information content. For example, given p_1 and p_2 , the mapping

$$\left. \begin{array}{l} x_1 \\ x_2 \end{array} \right\} \rightarrow f$$

gives

$$p(f) = p_1 + p_2.$$

The resulting change in the entropy reads

$$\begin{aligned}\Delta S &= S_f - S_{12} \\ &= -p(f) \ln p(f) + (p_1 \ln p_1 + p_2 \ln p_2) \\ &= p_1 \ln \frac{p_1}{p_1 + p_2} + p_2 \ln \frac{p_2}{p_1 + p_2},\end{aligned}$$

which is negative, provided that $p_1 \neq 0$ and $p_2 \neq 0$. Conversely, removing a constraint in a probability distribution systematically increases S .

1.9.1 Inferring unbiased probability distributions

The entropy S (or equivalently the information content) of a probability distribution can also be used as a criterion to infer subjective (i.e., theoretical) estimates of probability distributions. For instance, in the absence of any information or constraint on the processes behind the individual events i , the *best unbiased estimate* of $\{p_i\}$ is that all possible outcomes are equally probable, i.e., $p_i = 1/M$ for all i . The distribution maximizing S has the largest diversity and therefore yields the largest number of messages or ‘states’ which are likely to be obtained as the outcomes of the probability distribution $\{p_i\}$ for any given M . One may also say that this choice minimizes the information content of the probability distribution $\{p_i\}$.

If additional information is available the unbiased estimate for p_i is obtained by maximizing S subject to the constraints imposed by the available information. As an example let us assume that we know the average value $\langle F(x) \rangle = f$ of some function $F(x)$ of the random variable x . In this case we obtain the unbiased p_i from the extremum of

$$L(\{p_i\}) = \underbrace{S(\{p_i\})}_{-\sum_{i=1}^M p_i \ln p_i} - \alpha \left(\sum_{i=1}^M p_i - 1 \right) - \beta \left(\sum_{i=1}^M p_i F(x_i) - f \right),$$

where α and β are Lagrange multipliers. Straightforward derivation yields

$$\begin{aligned}\frac{\partial L}{\partial p_i} &= -\ln p_i - 1 - \alpha - \beta F(x_i) = 0 \\ \Rightarrow p_i &= e^{-(1+\alpha)} e^{-\beta F(x_i)}.\end{aligned}$$

Imposing the normalization condition $\sum_i p_i = 1$ we obtain $e^{1+\alpha} = \sum_i e^{-\beta F(x_i)}$ and thus

$$p_i = \frac{e^{-\beta F(x_i)}}{\sum_{i=1}^M e^{-\beta F(x_i)}}, \quad (1.19)$$

where β is such that

$$f = \langle F(x) \rangle = \frac{\sum_{i=1}^M e^{-\beta F(x_i)} F(x_i)}{\sum_i e^{-\beta F(x_i)}}. \quad (1.20)$$

Exercise 1.5:

Find the unbiased probability $p(x_i)$ for a random variable x_i ($i = 1, \dots, M$) knowing the first n moments of $p(x_i)$ (i.e., $\langle x^\nu \rangle = M_\nu$ for $\nu \leq n$).

- i) Show that $p(x_i) \propto \exp\left(\sum_{\nu=0}^n a_\nu x_i^\nu\right)$ with certain coefficients a_ν .
- ii) Consider the partition function $Z = \sum_{i=1}^M \exp\left\{\sum_{\nu=0}^n a_\nu x_i^\nu\right\}$ and show that the coefficients are given by the equations $\frac{\partial \ln Z}{\partial a_\nu} = M_\nu$.
-

1.9.2 Entropy of continuous probability densities

In analogy with the discrete case we can define the entropy of a continuous probability density distribution $p(\vec{x})$ as

$$S = -\langle \ln p(\vec{x}) \rangle = -\int p(\vec{x}) \ln p(\vec{x}) d\vec{x}.$$

However, notice that this definition does not have some of the nice properties of $S = -\sum_i p_i \ln p_i$ for discrete random variables. For instance, for a uniform distribution in the interval $[a, b]$, i.e.,

$$p(x) = \begin{cases} 1/(b-a) & \text{for } a \leq x \leq b \\ 0 & \text{elsewhere,} \end{cases}$$

we have

$$S = -\int_a^b \frac{1}{b-a} \ln\left(\frac{1}{b-a}\right) = \ln(b-a).$$

For large intervals this is positive and diverges logarithmically [$S \rightarrow +\infty$ for $(b-a) \rightarrow +\infty$]. In the opposite limit of a very narrow PDF around some point x_0 , we have

$$p(x) = \begin{cases} 1/\varepsilon & \text{for } x_0 - \varepsilon/2 < x < x_0 + \varepsilon/2 \\ 0 & \text{elsewhere.} \end{cases}$$

Therefore, $S = -\langle \ln p \rangle = \ln(\varepsilon)$, which tends to $-\infty$ for ε tending to 0. Notice that S can take negative values for very sharp $p(x) \simeq \delta(x)$ since the probability density $p(x)$ is not bounded, in contrast to the probabilities p_i of discrete variables. Such situations,

however, never appear in the description of macroscopic systems in equilibrium. Despite these diverging limits, it is interesting to observe that S always increases (decreases) as the diversity of the distribution increases (decreases). Indeed, $S \rightarrow +\infty$ when the width $|b - a|$ of the accessible interval in a uniform distribution $p(x) = 1/|b - a|$ diverges, and $S \rightarrow -\infty$ when $p(x)$ tends to a delta function.

In order to avoid the problem of the negative divergence of S for very sharp $p(x)$, or rather to understand its origin, it is useful to derive the expression for S for continuous PDF by starting from the expression

$$S = - \sum_i p_i \ln p_i$$

for discrete variables. To this aim, a finite lower bound or threshold Δ is introduced on the continuous variable x , below which two outcomes (differing by less than Δ) are considered to be equivalent. With this coarse graining the logarithm of the number of possible messages (i.e., the diversity of the probability distribution) is given by

$$S = - \sum_i \tilde{P}(x_i < x < x_i + \Delta) \ln \left[\tilde{P}(x_i < x < x_i + \Delta) \right]$$

where $x_i = i \Delta$ and $p_i = \tilde{P}(x_i < x < x_i + \Delta)$ is the probability for x to lie in the interval $[x_i, x_i + \Delta]$. Using that

$$\tilde{P}(x_i < x < x_i + \Delta) = P(x_i + \Delta) - P(x_i) = \int_{x_i}^{x_i + \Delta} p(x) dx,$$

where $P(x)$ refers to the cumulative probability function and $p(x) = \frac{dP}{dx}$ to the probability-density function, we have

$$S = - \sum_i \left(\int_{x_i}^{x_i + \Delta} p(x) dx \right) \ln \left(\int_{x_i}^{x_i + \Delta} p(x) dx \right) \geq 0.$$

If the spectrum of outcomes is discrete or shows very narrow peaks (narrower than Δ) we can still compute S and recover the limit of discrete random variables. In contrast, if $p(x)$ is smooth we can write

$$\begin{aligned} S &\cong - \sum_i \left(\int_{x_i}^{x_i + \Delta} p(x) dx \right) \ln [p(x_i) \Delta] \\ &\cong - \sum_i \left(\int_{x_i}^{x_i + \Delta} p(x) dx \right) [\ln p(x_i) + \ln \Delta] \\ &\cong - \int p(x) \ln [p(x)] dx - \ln \Delta. \end{aligned}$$

The term $\ln \Delta$ cancels the divergence of $S = -\langle \ln p(x) \rangle$ for $p(x) \rightarrow \delta(x)$.

Another potential problem of the definition

$$S = - \int p(x) \ln[p(x)] dx$$

is that it is not necessarily invariant under a bijective mapping $f = F(x)$. In fact, one has

$$p(f) df = p(x) dx \quad \Rightarrow \quad p(f) = p(x) \frac{1}{\left| \frac{dF}{dx} \right|}$$

and therefore

$$\begin{aligned} S[p(f)] &= - \int p(f) \ln p(f) df = - \int p(x) \left[\ln p(x) - \ln \left| \frac{dF}{dx} \right| \right] dx \\ &= - \int p(x) \ln p(x) dx + \int p(x) \ln \left| \frac{dF}{dx} \right| dx \\ &= S[p(x)] + \left\langle \ln \left| \frac{dF}{dx} \right| \right\rangle. \end{aligned}$$

A simple change of scale, i.e., $f = \alpha x$, would thus change the value of S. An expansion ($\alpha > 1$) increases S, since it increases the diversity, while a compression ($\alpha < 1$) reduces S since it flattens the variations of the random variable.

In the case of many random variables we define S analogously as

$$S = -\langle \ln p(\vec{x}) \rangle = - \int p(\vec{x}) \ln p(\vec{x}) d\vec{x}.$$

Following a change of variables $\vec{x} \rightarrow \vec{f}$ we have

$$S[p(\vec{f})] = S[p(\vec{x})] + \langle \ln |J(\vec{f})| \rangle,$$

where $J(\vec{f}) = \left| \frac{\partial f_i}{\partial x_j} \right|$ is the Jacobian of the variable transformation $f_i = f_i(x_1, \dots, x_N)$.

The entropy is thus invariant under canonical transformations in classical mechanics and unitary transformation in quantum mechanics, since in these cases the Jacobian equal to 1.

Exercise 1.6: Loaded dice

A dice is loaded such that $p_6 = n p_1$ (e.g., $n = 2$). In other words, six occurs n times more often than 1.

- i) Find the unbiased probabilities for the six faces of the dice. How are p_i for $i = 2-4$ related? Is this reasonable according to the notion of unbiased probability distribution?
 - ii) Find $\{p_i, i = 1-6\}$ in the limit $n \rightarrow \infty$ and analyze the results for the different i . Compare p_i for $i = 2-4$, p_6 and $p_1 + p_6$ in this limit with the unloaded $n = 1$ case. Are the results reasonable from the physical or mechanical perspective of a real loaded dice?
 - iii) Analyze the information content of the probability distribution function $\{p_i, i = 1-6\}$ as a function of n ?
-

2 Mixed states, density matrices and distribution functions

2.1 Introduction

The usual point of view in mechanics, either quantum or classical, is to predict the state of a system at a given time t provided that one knows the state of the system at a previous time $t_0 < t$ and that one know the exact form of the Hamiltonian. This is actually the scope of any deterministic mechanical theory. The information required to define a state, the very nature of it, and its connection to the observable properties is very different in the quantum and classical cases. Still, the time evolution is in both cases deterministic. While we all agree that QM should be the basis of any theory of matter, it is easy to see that a deterministic approach to macroscopic systems is neither practical nor meaningful. From a practical perspective first, it is clear that we can never succeed in recording all the information needed to determine the initial state of a system composed of 10^{20} atoms. Even the exact form of the Hamiltonian governing the system and its environment is, strictly speaking, inaccessible. Needless to say that solving the time evolution of such a system is impossible both analytically or numerically. The other important reason not to pursue the deterministic perspective is understanding. Suppose we would be able to propagate during $1 \mu\text{s}$ a wave function, also known as pure state, depending on 10^{20} coordinates. What would we do with all these numbers?² From an experimental perspective we are only interested in mean values, either within an interval of time, or in space, or both. The information consisting of the complete wave function (or of the coordinates and momenta of all particles in the classical case) as a function of time does not provide any understanding.

In order to successfully approach this problem a change of perspective is needed. Instead of following the dynamics of a given system having a given initial state as a function of time, it is far more meaningful to consider a large number of equivalent systems, which have a wide variety of initial states, ideally all possible states compatible with the global boundary conditions and constants of motion. This ensemble of systems should be distributed across the accessible states with some initial probability distribution. Actually, only the probability of finding a system in a given state is needed in order to compute any average value. Notice that with average values we not only mean global properties of the system (e.g., total energy, kinetic energy, magnetization, equilibrium volume, etc.) but also the most detailed microscopic information, such as, spin-polarized density distribution, density and spin correlation functions including their dependence on experimentally tunable external fields. The challenge is then to determine the time evolution of the probability distribution when the initial distribution has been given. This is the perspective of statistical mechanics as proposed by Maxwell, Boltzmann and Gibbs. The properties

²The reader may wish to estimate the number of bits needed to store the wave function of a system containing $N = 10^{20}$ particles (just at a single time t) and compare it with the total number of particles in the universe ($\sim 10^{80}$) or estimate the time needed to go through all of them once at the clock rate of the fastest CPUs ($\sim 10\text{GHz}$) and compare it with the current estimation of the age of the universe ($\sim 10^{10}$ years).

of such statistical ensembles of systems is characterized by density operators $\hat{\rho}$ in quantum mechanics and by statistical distribution functions $\rho(p, q)$ in classical mechanics. In the following we shall review the properties of $\hat{\rho}$ and $\rho(p, q)$. A rigorous justification of the ensemble or mixed state approach to macroscopic systems will be given later on. As we shall see, mixed states and ensembles appear quite naturally once the crucial, even if very weak, coupling between the system and its environment is taken into account.

2.2 Mixed states

The fundamental principles of quantum mechanics establish that the state of any system is characterized by an element $|\Psi\rangle$ of a vector space or Hilbert space \mathbb{V} . The superposition principle implies that the set of all possible states form a vector space, since $\forall |\Psi_1\rangle$ and $|\Psi_2\rangle \in \mathbb{V}$, we have $a|\Psi_1\rangle + b|\Psi_2\rangle \in \mathbb{V}$ for all $a, b \in \mathbb{C}$. These states are called pure states or microstates in the broader context of statistical mechanics. In order to characterize $|\Psi\rangle$ one usually expands it in a basis of states having well-defined values of a complete set of compatible observables f :

$$|\Psi\rangle = \sum_n a_n |f_n\rangle + \int a_f |f\rangle df.$$

It is clear that knowing all these expansion coefficient is impossible in practice, except for very small systems (atoms and small molecules). Strictly speaking, all possible states of the system at any time t are pure states, i.e., an element of \mathbb{V} . Nevertheless, this is not the only physical situation that one may encounter. In many cases of interest, and in particular in the study of macroscopic systems, one has to deal with an incoherent superposition of microstates $|\alpha_i\rangle$ each having a probability w_i but bearing no correlation whatsoever. Such ensembles of microstates are called *mixed states* or mixed ensembles.

2.2.1 Statistically independent systems

Mixed states are necessary in order to describe the properties of ensembles of physically equivalent, statistically independent systems, each being in a state $|\alpha_i\rangle$ with probability w_i . The reader is probably familiar with calculating the properties of a beam of atoms coming out of an oven in the context of the Stern-Gerlach experiment. In the case of a macroscopic system it is reasonable to regard a large system (containing for example 10^{20} particles) as an ensemble of N subsystems each containing a large number of particles as well (for example, 10^{10} subsystems with 10^{10} particles each). In the macroscopic limit the subsystems are also macroscopic, and the interaction between them are weak, so that their states can be regarded as statistically independent. The surface to volume ratio of the subsystems tends to zero as the size tends to infinity. It is important to note that full statistical independence requires the system to be open or that it is part of a much larger closed system, since in the case of closed system the subsystems must satisfy the usual conservation laws. For example, the sum of all the energies of the subsystems of a

closed system must be constant. In any case it is clear that the predictions of a theory of macroscopic systems must be the same, whether we consider fewer larger systems or a larger number of smaller subsystems, provided that they remain macroscopic.

We consider an ensemble of systems distributed over the pure states $|\alpha_i\rangle$ with probability w_i . The fractional population of the states $|\alpha_i\rangle$ satisfy the normalization condition

$$\sum_i w_i = 1$$

The pure states $|\alpha_i\rangle$ are properly normalized

$$\langle\alpha_i|\alpha_i\rangle = 1 \quad \forall i$$

However, the different $|\alpha_i\rangle$ need not be orthogonal to each other. In fact, the states of different subsystems bear no correlation to each other and therefore need not be orthogonal.

The average of any observable \hat{A} when a large number of measurements is performed on the ensemble is

$$\langle\hat{A}\rangle = \sum_i w_i \langle\alpha_i|\hat{A}|\alpha_i\rangle$$

Notice that $\langle\hat{A}\rangle$ involves the usual quantum mechanical average of the observable \hat{A} in the pure state $|\alpha_i\rangle$ weighted by the probability w_i of finding the systems in $|\alpha_i\rangle$.

2.2.2 A single quasi-closed system and its environment

It is interesting to observe that ensemble averages also appear when one considers the properties of a *single isolated system*, which has interacted with the environment prior to isolation. This is the most relevant situation we encounter in real experiments, since preparing a macroscopic system in a given state requires that some interactions with the environment have taken place, whose details cannot all be controlled. We consider a system, which is characterized by a complete set of orthonormal states $\{|\phi_n\rangle\}$, and its environment, which is characterized by the orthonormal basis set $\{|\chi_m\rangle\}$. The entity ‘system plus environment’ (S \oplus E) is assumed to be *strictly isolated*. Thus, according to the laws of quantum mechanics, S \oplus E must be in a well-defined pure state of the form

$$|\Psi\rangle = \sum_{mn} a_{mn} |\chi_m\rangle |\phi_n\rangle,$$

even though we are unable to know the precise values of the expansion coefficients $a_{nm} \in \mathbb{C}$ ($\sum_{mn} |a_{mn}|^2 = 1$).

We would like to compute the average value of an operator \hat{A} , which concerns some property of the system and thus affects only the variables $|\phi_n\rangle$. To this aim we rewrite $|\Psi\rangle$ in the form

$$|\Psi\rangle = \sum_m |\chi_m\rangle \sum_n a_{mn} |\phi_n\rangle \tag{2.1}$$

and define the normalized states of the system $|\alpha_m\rangle$ and the amplitudes b_m by

$$b_m|\alpha_m\rangle = \sum_n a_{mn}|\phi_n\rangle, \quad (2.2)$$

with

$$\langle\alpha_m|\alpha_m\rangle = 1 \quad \forall m. \quad (2.3)$$

Let us remark that $\langle\alpha_m|\alpha_{m'}\rangle \neq 0$ for $m \neq m'$ in general. Using Eqs. (2.2) and (2.3) one obtains

$$|b_m|^2 = \sum_{n,n'} a_{mn'}^* a_{mn} \langle\phi_{n'}|\phi_n\rangle = \sum_n |a_{mn}|^2.$$

Assuming for simplicity that $b_m \in \mathbb{R}$, we have

$$|\alpha_m\rangle = \frac{\sum_n a_{mn}|\phi_n\rangle}{\sqrt{\sum_n |a_{mn}|^2}}.$$

Notice that $|b_m|^2$ represents the probability of finding the environment in the state $|\chi_m\rangle$, when the system plus environment is in the pure state $|\Psi\rangle$ given by Eq. (2.1). We may actually write

$$|\Psi\rangle = \sum_m b_m |\chi_m\rangle |\alpha_m\rangle.$$

The normalization of $|\Psi\rangle$ and of $|\alpha_m\rangle$ for all m imply

$$\sum_m |b_m|^2 = 1.$$

We may now compute $\langle\Psi|\hat{A}|\Psi\rangle$ for any operator concerning the system, i.e., acting only on the variables $|\phi_n\rangle$ or $|\alpha_m\rangle$:

$$\langle\Psi|\hat{A}|\Psi\rangle = \sum_{m,m'} b_{m'}^* b_m \underbrace{\langle\chi_{m'}|\chi_m\rangle}_{\delta_{mm'}} \langle\alpha_{m'}|\hat{A}|\alpha_m\rangle = \sum_m |b_m|^2 \langle\alpha_m|\hat{A}|\alpha_m\rangle.$$

Remarkably, this takes the same form as the ensemble average with $w_m = |b_m|^2$, even though the state of the system plus environment is pure. Moreover, this expression holds for all times, since the eigenstates of the environment remain orthogonal to each other once the system and environment are decoupled. In practice it is impossible to keep a system decoupled from its environment for a very long time. All idealized isolated systems are in reality quasi-closed at best. One can imagine that the system and the environment interact from time to time. This corresponds to a change or redefinition of the expansion coefficients a_{mn} defining the global state $|\Psi\rangle$. Consequently, the interaction with the environment implies sampling different states $|\alpha_m\rangle$ and weights $w_m = |b_m|^2$. It is precisely the system-environment interaction, however weak, what ultimately explains the establishment of thermodynamic equilibrium.

It is important to note that the weights $w_m = |b_m|^2$ are independent of time as long as the system remains perfectly isolated from the environment, i.e., as long as the system and the environment do not interact. Indeed, a lack of interaction means that the Hamiltonian takes the form $\hat{H} = \hat{H}_{env} + \hat{H}_{sys}$ where \hat{H}_{sys} (\hat{H}_{env}) acts only on the variables of the system (environment). The states $|\phi_n\rangle$ and $|\chi_m\rangle$ remain eigenstates of the system and environment if no interaction is present:

$$\hat{H}_{env}|\chi_m\rangle = E_m|\chi_m\rangle$$

and

$$\hat{H}_{sys}|\phi_n\rangle = \varepsilon_n|\phi_n\rangle$$

Consequently,

$$|\Psi(t)\rangle = \sum_{mn} \underbrace{a_{mn} e^{-\frac{i}{\hbar}(E_m + \varepsilon_n)t}}_{a_{mn}(t)} |\chi_m\rangle |\phi_n\rangle$$

and

$$|b_m(t)|^2 = \sum_n |a_{mn}(t)|^2 = \sum_n |a_{mn}|^2$$

is independent of time. This implies that for a strictly isolated system all the time dependence of a mixed state must be ascribed to the time dependence of the states $|\alpha_m\rangle$. As we shall see, this leads to the Liouville equation of motion of the density operator $\hat{\rho}$.

Summarizing, mixed states or mixed ensembles provide a more general perspective to quantum mechanical systems, which includes pure states as a particular case ($w_m = 1$ for a given m) and which allows us to describe macroscopic systems from two complementary perspectives:

- i) Macroscopic systems can be regarded as a statistical ensemble of a large number of macroscopic subsystems, and
- ii) macroscopic systems can be regarded as quasi-closed systems having a very weak though non-vanishing interaction with the environment.

2.3 Density matrix

The expression for the average value of an operator in a mixed state

$$\langle \hat{A} \rangle = \sum_i w_i \langle \alpha_i | \hat{A} | \alpha_i \rangle$$

has a very clear physical interpretation. However one would like to express $\langle \hat{A} \rangle$ in an invariant form, which clearly separates the factors that depend on the ensemble from those that depend on the property under study. To this aim we introduce the completeness

relation $\sum_n |n\rangle\langle n| = 1$ and obtain

$$\begin{aligned}\langle \hat{A} \rangle &= \sum_i w_i \langle \alpha_i | \hat{A} \left(\sum_n |n\rangle\langle n| \right) | \alpha_i \rangle \\ &= \sum_n \sum_i \langle n | \alpha_i \rangle w_i \langle \alpha_i | \hat{A} | n \rangle = \sum_n \langle n | \left(\sum_i | \alpha_i \rangle w_i \langle \alpha_i | \right) \hat{A} | n \rangle \\ &= \text{Tr} \left\{ \left(\sum_i | \alpha_i \rangle w_i \langle \alpha_i | \right) \hat{A} \right\}.\end{aligned}$$

This can be written as

$$\langle \hat{A} \rangle = \text{Tr} \left\{ \hat{\rho} \hat{A} \right\}, \quad (2.4)$$

where we have introduced the density-matrix operator or density operator (DO)

$$\hat{\rho} = \sum_i | \alpha_i \rangle w_i \langle \alpha_i |.$$

The DO depends only on the considered mixed state, since it is given by the participating pure states $| \alpha_i \rangle$ and the corresponding probabilities or weights w_i . As we shall see, $\hat{\rho}$ defines the mixed state completely, i.e., the knowledge of $\hat{\rho}$ at a time t not only allows to determine the average value $\langle \hat{A} \rangle$ and probability distribution $P_A(a)$ for any observable \hat{A} at the given time t , but it also univocally defines $\hat{\rho}$ at any other subsequent time t' . In the case of mixed states, it takes the role played by the wave function for pure states. Of course, $\hat{\rho}$ is independent of the observable \hat{A} under consideration. Equation (2.4) applies to the average of any observable.

Let us recall some properties of the trace of an operator

$$\text{Tr} \left\{ \hat{A} \right\} = \sum_n \langle n | \hat{A} | n \rangle,$$

where $\{|n\rangle\}$ is a complete orthonormal basis. It is easy to see that

$$\text{Tr} \left\{ \hat{A} \hat{B} \right\} = \text{Tr} \left\{ \hat{B} \hat{A} \right\}$$

for any operators \hat{A} and \hat{B} . This also implies that $\text{Tr}\{\hat{A}\}$ is independent of the orthonormal basis used for performing the sum. If \hat{U} is a unitary transformation ($\hat{U}^\dagger \hat{U} = 1$) such that

$$|u_n\rangle = \hat{U} |n\rangle$$

we have

$$\sum_n \langle u_n | \hat{A} | u_n \rangle = \sum_n \langle n | \hat{U}^\dagger \hat{A} \hat{U} | n \rangle = \text{Tr} \left\{ \hat{U}^\dagger \hat{A} \hat{U} \right\} = \text{Tr} \left\{ \hat{A} \right\}$$

Finally, it is also useful to recall that

$$\text{Tr}\{|\alpha\rangle\langle\alpha|\} = \sum_n |\langle n|\alpha\rangle|^2 = 1$$

for any state $|\alpha\rangle$ whose norm $\langle\alpha|\alpha\rangle = 1$.

The density operator $\hat{\rho}$ has the following important properties:

i) Hermiticity:

$$\hat{\rho}^\dagger = \hat{\rho} \quad \text{since} \quad w_i \in \mathbb{R}$$

ii) Normalized trace:

$$\text{Tr}\{\hat{\rho}\} = 1 \quad \text{since} \quad \sum_i w_i = 1$$

iii) Since $\hat{\rho}^\dagger = \hat{\rho}$, it is diagonalizable. Thus, $\hat{\rho}$ has a spectral representation of the form

$$\hat{\rho} = \sum_k \rho_k |\Psi_k\rangle\langle\Psi_k|$$

where $\hat{\rho}|\psi_k\rangle = \rho_k|\psi_k\rangle$, $\rho_k \in \mathbb{R}$ and $\langle\psi_k|\psi_{k'}\rangle = \delta_{kk'}$. The eigenstates of $\hat{\rho}$ form a complete orthonormal basis. Notice that the weights w_i entering the definition of $\hat{\rho}$ are not necessarily the eigenvalues of $\hat{\rho}$, since in general $\langle\alpha_i|\alpha_j\rangle \neq \delta_{ij}$.

iv) The eigenvalues of $\hat{\rho}$ satisfy $0 \leq \rho_k \leq 1$. Let us first show that $\hat{\rho}$ is positive semi-definite:

$$\langle\Psi|\hat{\rho}|\Psi\rangle = \sum_i w_i |\langle\Psi|\alpha_i\rangle|^2 \geq 0$$

for any state $|\Psi\rangle$. This implies that the eigenvalues $\rho_k \geq 0$. In addition, since $\sum_k \rho_k = \text{Tr}\hat{\rho} = 1$ we must have $\rho_k \leq 1$.

v) The square of $\hat{\rho}$ satisfies

$$\text{Tr}\{\hat{\rho}^2\} \leq 1.$$

This is easily proven by noting that

$$1 = (\text{Tr}\{\hat{\rho}\})^2 = \left(\sum_k \rho_k\right)^2 = \sum_k \rho_k^2 + \sum_{k \neq l} \rho_k \rho_l.$$

Since $\sum_{k \neq l} \rho_k \rho_l \geq 0$, we have

$$\text{Tr}\{\hat{\rho}^2\} = \sum_k \rho_k^2 \leq 1.$$

vi) Pure-state characterization:

$$\text{Tr}\{\hat{\rho}^2\} = 1 \quad \Leftrightarrow \quad \hat{\rho} = |\Psi_k\rangle\langle\Psi_k|$$

for some k . In other words, $\text{Tr}\{\hat{\rho}^2\} = 1$ if and only if $\hat{\rho}$ describes a pure state. The proof follows by noting that

$$\text{Tr}\{\hat{\rho}^2\} = \sum_k \rho_k^2 = 1 \quad \Leftrightarrow \quad \sum_{k \neq l} \rho_k \rho_l = 0 \quad \Leftrightarrow \quad \rho_k \rho_l = 0 \quad \forall k \neq l.$$

Therefore, only one eigenvalue ρ_k can be different from zero, which must then be equal to one. Consequently, $\hat{\rho} = |\psi_k\rangle\langle\psi_k|$ and $\hat{\rho}^2 = \hat{\rho}$. We conclude that $\text{Tr}\{\hat{\rho}^2\}$ allows us to distinguish mixed states from pure states.

- vii) $\text{Tr}\{\hat{\rho}^2\}$ is not only independent of the representation but also *independent of time*. This implies that pure states can never evolve into mixed states, and vice versa, as long as the system is perfectly isolated. The reason for this is that the dynamics of a system, which is perfectly decoupled from the environment, follows the unitary time-evolution operator $\hat{U}(t, t_0)$. The proof is straightforward by noting that

$$\hat{\rho}^2 = \sum_{ij} |\alpha_i\rangle w_i \langle\alpha_i|\alpha_j\rangle w_j \langle\alpha_j|$$

and thus

$$\begin{aligned} \text{Tr}\{\hat{\rho}^2\} &= \sum_{ij} \langle n|\alpha_i\rangle w_i \langle\alpha_i|\alpha_j\rangle w_j \langle\alpha_j|n\rangle \\ &= \sum_{ij} w_i w_j |\langle\alpha_i|\alpha_j\rangle|^2. \end{aligned}$$

Since $\hat{U}(t, t_0)$ is unitary, the time evolution $|\alpha_i, t\rangle = \hat{U}(t, t_0)|\alpha_i, t_0\rangle$ does not modify the scalar products $\langle\alpha_i|\alpha_j\rangle$. Consequently, $\text{Tr}\{\hat{\rho}^2\}$ is independent of time in an isolated system.

Summarizing so far, we may say that the set of all operators of the form

$$\hat{\rho} = \sum_i w_i |\alpha_i\rangle\langle\alpha_i| \tag{2.5}$$

with $\langle\alpha_i|\alpha_i\rangle = 1 \forall i$, and the set of all hermitic, positive semi-definite operators with trace equal one are identical. However, notice that the representation of $\hat{\rho}$ in the form (2.5) is not unique, since the $|\alpha_i\rangle$ are not necessarily orthogonal to each other. This means that mixed states, which may look different in terms of $\{|\alpha_i\rangle\}$ and $\{w_i\}$, may have the same density operators $\hat{\rho}$ and may thus be physically identical. On the other side $\hat{\rho}$ characterizes a physical state fully and univocally concerning the results of every possible measurement. The knowledge of $\hat{\rho}$ defines not only the average value of any observable \hat{A} through

$$\langle\hat{A}\rangle = \text{Tr}\{\hat{\rho}\hat{A}\}$$

but also the average

$$\langle F(\hat{A})\rangle = \text{Tr}\{\hat{\rho}F(\hat{A})\}$$

of any function of it.

Consequently, $\hat{\rho}$ defines the probability density distribution $P_A(a)$ that the result of a measurement of \hat{A} yields the value a . This probability density is given by

$$P_A(a) = \langle \delta(a - \hat{A}) \rangle = \text{Tr} \left\{ \hat{\rho} \delta(a - \hat{A}) \right\}. \quad (2.6)$$

Therefore, the knowledge of $\hat{\rho}$ gives the most complete experimentally accessible information on any observable.

Exercise 2.7: Show the validity of the above expression (2.6) for $P_A(a)$ knowing that, by definition, the probability distribution $P_A(a)$ satisfies $\langle F(\hat{A}) \rangle = \int P_A(a) F(a) da$ for any function F of the observable \hat{A} .

From a more mathematical perspective it is interesting to note that the set of all density operators is *convex*. Let $\hat{\rho}_1$ and $\hat{\rho}_2$ be two density operators describing two different (mixed) states. Then the operator

$$\hat{\rho} = \alpha_1 \hat{\rho}_1 + \alpha_2 \hat{\rho}_2$$

with $\alpha_1 + \alpha_2 = 1$ ($\alpha_i \geq 0$) is also a density operator describing a possible mixed state.

Pure states are particular cases of the more general concept of mixed states, which can be perfectly described with density operators. It is easy to see that a density operator $\hat{\rho}$ corresponds to a pure state if and only if *any* of the following equivalent conditions is satisfied:

- i) There is a state $|\alpha\rangle$ such that $\hat{\rho} = |\alpha\rangle\langle\alpha|$.
- ii) $\hat{\rho}^2 = \hat{\rho}$,
- iii) $\text{Tr} \{ \hat{\rho}^2 \} = 1$,
- iv) $\text{Tr} \{ \hat{\rho} \ln \hat{\rho} \} = \langle \ln \hat{\rho} \rangle = 0$, or
- v) $\hat{\rho}$ cannot be written as the combination of two different $\hat{\rho}_1$ and $\hat{\rho}_2$, i.e., there exist no density operators $\hat{\rho}_1 \neq \hat{\rho}_2$ such that $\hat{\rho} = \alpha_1 \hat{\rho}_1 + \alpha_2 \hat{\rho}_2$ with $\alpha_1 + \alpha_2 = 1$ and $\alpha_i > 0$.

Exercise 2.8: Prove the above mentioned equivalent characterizations of a pure state.

2.4 Time dependence of the density operator of isolated systems

We consider here the time dependence of the density operator of strictly isolated systems, for which we know that the weights w_i of the states $|\alpha_i\rangle$ building the mixed state are

independent of time. The quantum states $|\alpha_i\rangle$ follow the Schrödinger equation. We consider $\hat{\rho}(t)$ in the Schrödinger picture:

$$\hat{\rho}_s(t) = \sum_i w_i |\alpha_i, t\rangle \langle \alpha_i, t|,$$

where the Schrödinger kets $|\alpha_i, t\rangle = \hat{U}(t, t_0)|\alpha_i, t_0\rangle$ satisfy the Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} |\alpha_i, t\rangle = \hat{H} |\alpha_i, t\rangle$$

or equivalently

$$-i\hbar \frac{\partial}{\partial t} \langle \alpha_i, t| = \langle \alpha_i, t| \hat{H}$$

The Hamilton operator $\hat{H}(t)$ may depend on time. It is of course hermitic, since $|\alpha_i\rangle$ preserves its norm as a function of time. We then have

$$i\hbar \frac{\partial \hat{\rho}_s}{\partial t} = \sum_i w_i \left(\hat{H} |\alpha_i, t\rangle \langle \alpha_i, t| - |\alpha_i, t\rangle \langle \alpha_i, t| \hat{H} \right)$$

or equivalently

$$i\hbar \frac{\partial \hat{\rho}_s}{\partial t} = [\hat{H}, \hat{\rho}_s]. \quad (2.7)$$

This is known as Liouville's equation. It describes the time dependence of the operator $\hat{\rho}_s$ in the *Schrödinger picture*, which originates in the time dependence of the Schrödinger kets $|\alpha_i\rangle$. As already said, the density operator plays the role of the wave function for pure states in the more general case of mixed states. It depends on time in the Schrödinger picture and is, as we shall see, independent of time in the Heisenberg picture. It is worth noting that Eq. (2.7) is not the equation of motion of an operator $\hat{A}_H(t)$ in the Heisenberg picture: $\partial \hat{A}_H(t) / \partial t = (i/\hbar)[\hat{H}, \hat{A}_H(t)]$. It is the analog (or extension) of the Schrödinger equation $i\hbar \partial |\Psi\rangle / \partial t = \hat{H} |\Psi\rangle$, which holds pure states $|\Psi\rangle$, in the case of mixed states.

An important consequence can be drawn from the fact that Liouville's equation (2.7) is a *first-order differential equation*: The knowledge of $\hat{\rho}$ at any time defines $\hat{\rho}$ at all other times. Moreover, knowing that $\hat{\rho}$ allows us to calculate the outcome of any experiment [e.g., by using the probability distribution $P_A(a)$ given by Eq. (2.6)] we conclude that $\hat{\rho}$ defines the state of the system very much like the wave function or ket $|\Psi\rangle$ does in the case of pure states. Although $\hat{\rho}$ is an hermitic operator, it is not an observable. Just as the wave function, it is not directly measurable.

The thermodynamic equilibrium is defined by requiring that the macrostate of the system is independent of time. This is equivalent to requiring that $\frac{\partial \hat{\rho}_s}{\partial t} = 0$, since $\hat{\rho}_s$ defines the state of the system. The Liouville equation implies that in thermodynamic or statistical equilibrium we have

$$[\hat{H}, \hat{\rho}_s] = 0.$$

In this case the eigenstates $|\psi_k\rangle$ of $\hat{\rho}_s$ can be chosen to be eigenstates of \hat{H} . This is a very important hint for our search of the form of $\hat{\rho}_s$ describing equilibrium.

An alternative to Liouville's differential equation is to express $\hat{\rho}_s$ using the time-evolution operator:

$$|\alpha_i, t\rangle = \hat{U}(t, t_0)|\alpha_i, t_0\rangle$$

and

$$\langle\alpha_i, t| = \langle\alpha_i, t_0|\hat{U}^\dagger(t, t_0).$$

In this way we have

$$\hat{\rho}_s(t) = \hat{U}(t, t_0)\hat{\rho}_s(t_0)\hat{U}^\dagger(t, t_0).$$

Notice that the time evolution of the operator $\hat{\rho}_s(t)$ corresponds simply to a unitary transformation ($\hat{U}\hat{U}^\dagger = 1$). This form of $\hat{\rho}_s(t)$ is particularly useful in order to demonstrate the time independence of $\text{Tr}\{\rho\}$, $\text{Tr}\{\rho^2\}$ and of the eigenvalues ρ_k of $\hat{\rho}$.

Exercise 2.9:

- i) Find the density matrix operator $\hat{\rho}_H(t)$ in the Heisenberg picture. Verify that it is independent of time.
- ii) Verify the equivalence of the average values of any operator \hat{A} in the Schrödinger and Heisenberg picture.

Exercise 2.10: The entropy of a mixed state is defined by $S = -k_B\langle\ln\hat{\rho}\rangle = -k_B\sum_k\rho_k\ln\rho_k$, where k_B is the Boltzmann constant having units of energy divided by temperature. Show that S is independent of time for strictly isolated systems (i.e., following Liouville's equation). As we shall see, this implies that an isolated system can never reach equilibrium on its own, i.e., without a however weak interaction with the environment.

Before closing this section it is probably useful to recall that the time derivative of an operator \hat{f} in quantum mechanics is defined by the condition

$$\left\langle\psi\left|\frac{d\hat{f}}{dt}\right|\psi\right\rangle = \frac{d}{dt}\langle\psi|\hat{f}|\psi\rangle$$

for any $|\psi(t)\rangle$. It follows that

$$\frac{d\hat{f}}{dt} = \frac{\partial\hat{f}}{\partial t} + \frac{i}{\hbar}[\hat{H}, \hat{f}].$$

Notice that $d\hat{f}/dt$ is the operator of a new physical observable in the Schrödinger picture (e.g., the velocity $\hat{v}_x = d\hat{x}/dt$). Calculating the time derivative of the density operator we have

$$\frac{d\hat{\rho}_s}{dt} = \frac{\partial\hat{\rho}_s}{\partial t} + \frac{i}{\hbar} [\hat{H}, \hat{\rho}] = 0,$$

where the second equality follows from the Liouville equation (2.7). This means that $\hat{\rho}_s$ is a constant of motion for strictly isolated systems, even though it depends explicitly on time ($\partial\hat{\rho}_s/\partial t \neq 0$) in the most general out-of-equilibrium case. In other words, the propagation of the kets $|\psi(t)\rangle = \hat{U}(t, t_0)|\psi(t_0)\rangle$ and the time dependence of $\hat{\rho}(t) = \hat{U}(t, t_0)\rho(t_0)\hat{U}^\dagger(t, t_0)$ is such that any average $\langle\psi(t)|\hat{\rho}(t)|\psi(t)\rangle$ involving $\hat{\rho}$ or of any function of $\hat{\rho}$ is independent of time (i.e., $\langle\psi(t)|\hat{\rho}(t)|\psi(t)\rangle = \langle\psi(t_0)|\hat{\rho}(t_0)|\psi(t_0)\rangle$ for all t). We can also say that all eigenstates $|\psi_k\rangle$ of $\hat{\rho}(t_0)$ at a given time t_0 remain eigenstates at any subsequent time t . This remarkable conservation law holds in general. In equilibrium we have in addition that $\partial\hat{\rho}/\partial t = 0$ and therefore $[\hat{H}, \hat{\rho}] = 0$. In this case $\hat{\rho}$ is a constant of motion in the usual most strict sense.

2.5 The statistical distribution function

We would like to transpose the concepts of statistical ensembles and density operator to systems which can be described by classical mechanics. A detailed formal derivation of the classical limit of quantum statistical mechanics will be presented later on. The dynamic state of a classical system with s degrees of freedom is known to be defined by its generalized coordinates $q = (q_1, \dots, q_s)$ and conjugated momenta $p = (p_1, \dots, p_s)$. Each point (p, q) represents a state of the entire macroscopic N -particle system ($s = 3N$). As in the quantum case we refer to these states, which contain the most detailed information on the system, as *microstates* or *pure states*. The $2s$ dimensional space containing all the microstates $(p, q) = (p_1, \dots, p_s, q_1, \dots, q_s)$ is known as Γ -space or *phase space* of the system. As in the quantum case it is meaningful to change the perspective of the mechanical description from the deterministic dynamics of a single system in a precisely defined state (p, q) to the dynamics of the probability distribution of a large ensemble of systems distributed throughout all possible microstates. This broader physical situation is described by the *statistical distribution function* or simply *distribution function*

$$\rho(p, q) = \rho(p_1, \dots, p_s, q_1, \dots, q_s),$$

which represents the joint probability-density function for finding the system in the microstate $(p, q) = (p_1, \dots, p_s, q_1, \dots, q_s)$. In other words

$$\rho(p, q) dpdq = \rho(p_1, \dots, p_s, q_1, \dots, q_s) dp^s dq^s$$

is the probability of finding the system in the phase-space volume element $dp^s dq^s$ centered at the point (p, q) . As any joint probability function, $\rho(p, q)$ satisfies $\rho(p, q) \geq 0$ and $\int \rho(p, q) dpdq = 1$. These conditions correspond to $w_i \geq 0$ and $\sum_i w_i = 1$ in the quantum case.

The physical properties of classical systems are given by functions $A(p, q)$ of the generalized coordinates and momenta. The corresponding average value and variance are given by

$$\langle A \rangle = \int dpdq A(p, q) \rho(p, q)$$

and

$$\Delta A^2 = \langle (A - \langle A \rangle)^2 \rangle = \langle A^2 \rangle - \langle A \rangle^2 \geq 0.$$

The motivations for considering statistical ensembles in order to describe macroscopic systems are the same as in the quantum case:

- i) One may regard a macroscopic system as an ensemble of equivalent, statistically independent subsystems each occupying a state (p_i, q_i) . It is the distribution of the microstates of the subsystems throughout Γ space what is described by the joint probability distribution function $\rho(p, q)$.
- ii) Alternatively, one may follow the time evolution of a system, or of any macroscopic part of a larger system, and record the state (p_i, q_i) of the system at a large number of equally spaced instants t_i . From this perspective $\rho(p, q)$ corresponds to the probability of finding the system in the state (p, q) averaged over a large period of time. The average of observables corresponds then to time averages.
- iii) Another complementary perspective is to consider the system as being quasi-closed, i.e., almost perfectly isolated, thus experiencing relatively rare weak interactions with the environment. These lead to changes of state beyond the time dependence of (p, q) corresponding to a strictly isolated system. Again, the actual microstate of the system cannot be known with certainty. Thus, the notions of statistical ensemble and probability distribution $\rho(p, q)$ prevail.

2.6 Time dependence of $\rho(p, q)$: Liouville theorem

We would like now to investigate the time dependence of $\rho(p, q)$ assuming that the system is isolated. More generally, we consider a system that is quasi-closed and analyze its time evolution between two interactions with the environment. In this case, the generalized coordinates of each element in the ensemble follow Hamilton equations

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}$$

and

$$\dot{q}_i = \frac{\partial H}{\partial p_i}$$

for $i = 1, \dots, s$, where $H = H(p, q)$ is the Hamilton function of the isolated system. These equations describe how each microstate, also known as *representative point*, moves in Γ space as a function of time. Knowing that (p, q) defines the mechanical state completely, one may consider any point $[p(t_0), q(t_0)]$ in the trajectory as initial condition. In this way

one always obtains the same distinct trajectory, i.e., the same unique (since deterministic) curve connecting the points (p, q) at subsequent or past instants t . This implies that a trajectory (p, q) in Γ space can never cross itself or another distinct trajectory, since in this case the crossing point (p_0, q_0) would not define the mechanical state unambiguously. A trajectory can be a simple closed loop, which corresponds to a periodic motion. However, neither two trajectories can merge into one, nor a trajectory can evolve into a closed loop. The latter would imply that there is a point with reversed momenta $(-p_0, q_0)$ in Γ space, where the trajectory bifurcates, which again contradicts the deterministic character of classical motion.

In order to demonstrate the last statement we observe that Hamilton's equations are invariant upon time reversal. More precisely, let $[p(t), q(t)]$ be a solution of Hamilton's equations and consider the trajectory $q'(t) = q(2t_0 - t)$ and $p'(t) = -p(2t_0 - t)$, which reverses the path along the same coordinates with reversed momenta. Should H depend explicitly on time, which is of course not the case in an isolated system, one would also need to inverse its time dependence as $H'(p, q, t) = H(p, q, 2t_0 - t)$. Notice that the locations of the points (p, q) and (p', q') are different due to the momentum reversals. It is easy to see that

$$\frac{dq'_i}{dt}(t) = -\frac{dq_i}{dt}(2t_0 - t) = -\frac{\partial H}{\partial p_i}(2t_0 - t) = -\frac{\partial H'}{\partial p_i}(t) = \frac{\partial H'}{\partial p'_i}(t)$$

and

$$\frac{dp'_i}{dt}(t) = \frac{dp_i}{dt}(2t_0 - t) = -\frac{\partial H}{\partial q_i}(2t_0 - t) = -\frac{\partial H'}{\partial q_i}(t) = -\frac{\partial H'}{\partial q'_i}(t),$$

which implies that (p', q') is also a valid trajectory in Γ space. In the reverse-time evolution merging becomes bifurcation. Therefore, we conclude that merging of trajectories in Γ space is not possible. Consequently, the trajectories must preserve their identity at all times. It is therefore reasonable to expect that along the time evolution each point in Γ space will carry along its local probability density $\rho(p, q)$, so that the total derivate $d\rho/dt = 0$. This statement, known as Liouville's theorem, shall be rigorously demonstrated in the following.

2.6.1 Total time derivative of classical observables

As a preliminary calculation, let us consider some physical property f , which is given by a function $f(p, q, t)$ of the coordinates, momenta and time, and compute its total time derivative along the classical trajectory:

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \sum_k \left(\frac{\partial f}{\partial q_k} \dot{q}_k + \frac{\partial f}{\partial p_k} \dot{p}_k \right).$$

Using Hamilton's equations we have

$$\begin{aligned}\frac{df}{dt} &= \frac{\partial f}{\partial t} + \sum_k \left(\frac{\partial H}{\partial p_k} \frac{\partial f}{\partial q_k} - \frac{\partial f}{\partial p_k} \frac{\partial H}{\partial q_k} \right) \\ &= \frac{\partial f}{\partial t} + \{H, f\},\end{aligned}\tag{2.8}$$

where we have introduced the *Poisson bracket*

$$\{f, g\} = \sum_k \left(\frac{\partial f}{\partial p_k} \frac{\partial g}{\partial q_k} - \frac{\partial g}{\partial p_k} \frac{\partial f}{\partial q_k} \right)$$

between any two functions f and g of (p, q) . Poisson brackets are bilinear functions of f and g , which have very similar algebraic properties as the commutator between two operators:

$$\begin{aligned}\{\alpha f_1 + \beta f_2, g\} &= \alpha \{f_1, g\} + \beta \{f_2, g\}, \\ \{f, g\} &= -\{g, f\},\end{aligned}$$

and

$$\{f_1 f_2, g\} = f_1 \{f_2, g\} + \{f_1, g\} f_2.$$

Notice the analogy between Eq. (2.8) and the time derivative of observables in quantum mechanics:

$$\frac{d\hat{f}}{dt} = \frac{\partial \hat{f}}{\partial t} + \frac{i}{\hbar} [\hat{H}, \hat{f}].$$

One may also say that df/dt and $d\hat{f}/dt$ are connected by the correspondence relation $\{H, f\} \leftrightarrow \frac{i}{\hbar} [\hat{H}, \hat{f}]$.

Applying Eq. (2.8) for df/dt to the distribution function ρ we have

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \{H, \rho\},\tag{2.9}$$

which is known as Liouville's equation.

2.6.2 The continuity equation in n dimensions

We would like to extend the continuity equation, well-known from electromagnetism, to the density fields and associated velocity fields in \mathbb{R}^n . Let $\rho(\vec{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ be the density distribution of some fictitious particles moving in \mathbb{R}^n , and let $\vec{v}(\vec{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be the velocity field corresponding to these particles. The current density is then given by $\vec{j} = \rho \vec{v}$. For any arbitrary volume $V \subset \mathbb{R}^n$, the rate of change in the number of particles inside V per unit time can be calculated in two ways:

$$\frac{d}{dt} \left(\int_V \rho(\vec{x}) d^n x \right) = - \int_{S(V)} \vec{j} \cdot \hat{n} da.\tag{2.10}$$

The left-hand side is the straightforward derivative of the number of particles inside V . And the right-hand side is the flux of the current density across the surface $S(V)$ of the volume V . As usual, the normal \hat{n} to the surface points in the outwards direction. Notice that we have taken into account that there are no external sources or drains of particles, so that the change in the number of particles within V can only result from particles actually crossing the surface $S(V)$. Using Gauss' theorem on the right-hand side of Eq. (2.10) we have

$$\int_V \left(\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{j} \right) d^n x = 0 ,$$

where $\vec{\nabla} = \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n} \right)$ stands for the nabla operator in \mathbb{R}^n . Since this holds for any V we obtain

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{j} = 0 . \quad (2.11)$$

This continuity equation simply expresses the conservation of the number of particles underlying the density $\rho(\vec{x})$ and moving according to the velocity field $\vec{v}(\vec{x})$.

2.6.3 Time dependence of $\rho(p, q, t)$

In order to investigate the time dependence of $\rho(p, q, t)$ it is useful to regard $\rho(p, q)$ as the density of representative points in Γ space, whose motion leads to the current density $\vec{j} = \rho \vec{v}$. Since ρ depends on (p, q) the velocity field is given by

$$\vec{v} = (\dot{p}_1, \dot{p}_2, \dots, \dot{p}_s, \dot{q}_1, \dots, \dot{q}_s) \in \mathbb{R}^{2s} .$$

In a more compact form we may write

$$\vec{v} = (\dot{p}, \dot{q}) = \left(-\frac{\partial H}{\partial q}, \frac{\partial H}{\partial p} \right) .$$

It is easy to see that the velocity field \vec{v} associated with the motion of the representative points in Γ space (microstates of the ensemble) is divergenceless:

$$\vec{\nabla} \cdot \vec{v} = \sum_{k=1}^s \left(-\frac{\partial}{\partial p_k} \frac{\partial H}{\partial q_k} + \frac{\partial}{\partial q_k} \frac{\partial H}{\partial p_k} \right) = 0 ,$$

where we have used the definition

$$\vec{\nabla} = \left(\frac{\partial}{\partial p_1}, \dots, \frac{\partial}{\partial p_s}, \frac{\partial}{\partial q_1}, \dots, \frac{\partial}{\partial q_s} \right) .$$

of the nabla operator in Γ space (\mathbb{R}^{2s}). The divergence of the current density $\vec{j} = \rho \vec{v}$ is then simply

$$\vec{\nabla} \cdot \vec{j} = \vec{\nabla} \cdot (\rho \vec{v}) = \vec{v} \cdot \vec{\nabla} \rho + \rho \vec{\nabla} \cdot \vec{v} = \vec{v} \cdot \vec{\nabla} \rho .$$

Developing $\vec{v} \cdot \vec{\nabla} \rho$ we have

$$\vec{\nabla} \cdot \vec{j} = \sum_{k=1}^s \left(\dot{p}_k \frac{\partial \rho}{\partial p_k} + \dot{q}_k \frac{\partial \rho}{\partial q_k} \right) = \sum_{k=1}^s \left(\frac{\partial H}{\partial p_k} \frac{\partial \rho}{\partial q_k} - \frac{\partial \rho}{\partial p_k} \frac{\partial H}{\partial q_k} \right) = \{H, \rho\}. \quad (2.12)$$

Replacing Eq. (2.12) in the the expression (2.9) for the total time derivative of ρ , and using the continuity equation (2.11) we finally obtain

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \{H, \rho\} = \frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \vec{j} = 0.$$

This result, known as *Liouville's theorem*, tells us that the total derivative of ρ vanishes at all times. In other words, each representative point carries the probability density around it all along its trajectory in phase space. The result holds, of course, provided that the time evolution follows from the Hamilton function of an isolated system.

We arrive to the important conclusion that ρ is a constant of motion, even though $\partial \rho / \partial t \neq 0$. One also says that the *flow* of the fluid associated with the motion of the microstates in a statistical ensemble is incompressible. Notice that the flow is always incompressible but not the fluid, since $\partial \rho / \partial t \neq 0$ in general. However, in the particular case of thermodynamic equilibrium we have $\partial \rho / \partial t = 0$, since the very notion of equilibrium implies time independence. Therefore, in equilibrium, we find

$$\{H, \rho\} = 0.$$

In this case the distribution function ρ is a constant of motion in the narrow sense, very much like the total energy, momentum, or angular momentum of a closed system. As in quantum mechanics, this gives us a very important hint in order to derive the actual expression of the equilibrium ρ as a function of (p, q) .

2.6.4 A note on ergodicity and relaxation times

Since $\rho(p, q)$ is the same for all the microstates visited along a phase-space trajectory, one may be tempted to conclude that ρ should be constant within the hypersurface in Γ space containing all the points that are compatible with the given set of constants of motion (e.g., energy, number of particles and volume). This actually holds for *ergodic systems*, i.e., systems in which the representative points (p, q) as a function of time cover the entire accessible phase space. More precisely, the ergodic condition requires that if one waits long enough time, any representative point (p, q) must eventually come arbitrarily close to any other point in the accessible part of the Γ space. It is important to remark that an arbitrarily long time T_{erg} might be required for the system to reach a state that is arbitrarily close to some point (p', q') . For instance, a single particle traveling at the speed of light would need about 10000 years to come within 1 Å of every point in 1 m³. However, this long time T_{erg} has nothing to do with the time that the system needs to reach equilibrium. The latter, known as relaxation time τ_{rel} , is a well defined property of every macroscopic system, which is not only much shorter, but also much more important

physically. τ_{rel} is independent of any particular choice of the initial and target states. However, it depends on the system size. Short range equilibrium is reached in a much shorter time than across long distances. Moreover, different relaxation times often apply to different degrees of freedom (e.g., translational and spin electronic degrees of freedom, lattice-vibrational degrees of freedom, etc.) or to different properties (e.g., the density of particles and the chemical composition in thermally activated chemical reactions).

The above suggested *equal a priori probability of all accessible states* is physically correct. It constitutes in fact the fundamental principle of statistical mechanics. However, its validity does not rely on the ergodic hypothesis but on quantum mechanics and statistical independence arguments in the macroscopic limit.

2.7 Entropy conservation in strictly isolated classical systems

It is interesting to observe that the entropy

$$S = -\langle \ln \rho \rangle = - \int \rho(p, q) \ln[\rho(p, q)] dpdq$$

of the statistical distribution $\rho(p, q)$ of a strictly isolated system, i.e., a system following a deterministic Hamiltonian dynamics for which Liouville theorem holds, is independent of time.

A brute force proof could follow the lines

$$\frac{\partial S}{\partial t} = - \int dpdq \left(\frac{\partial \rho}{\partial t} \ln \rho + \frac{\partial \rho}{\partial t} \right) = - \int dpdq \frac{\partial \rho}{\partial t} \ln \rho,$$

where we have used that $\int dpdq \rho(p, q) = 1$ at all times. However, a far more elegant proof can be achieved by using some previous results on the entropy of continuous probability distributions and the properties of canonical transformations in classical mechanics. Consider the entropy

$$S(\tau) = \int \rho_\tau(p, q) \ln \rho_\tau(p, q) dpdq, \quad (2.13)$$

where $\rho_\tau(p, q)$ is the statistical distribution at time τ . Liouville's theorem, i.e., $d\rho/dt = 0$, implies

$$\rho_\tau(p_\tau, q_\tau) = \rho_0(p_0, q_0), \quad (2.14)$$

where (p_τ, q_τ) follows from (p_0, q_0) as a result of the classical deterministic time evolution. Obviously, (p_τ, q_τ) is a well-defined function of initial state (p_0, q_0) . The transformation

$$\begin{cases} p_\tau = p_\tau(p_0, q_0) \\ q_\tau = q_\tau(p_0, q_0) \end{cases}$$

and its inverse

$$\begin{cases} p_0 = p_0(p_\tau, q_\tau) \\ q_0 = q_0(p_\tau, q_\tau) \end{cases}$$

are canonical transformations, since a displacement in time does not alter the form of Hamilton's equations. Moreover, the Jacobian $J \left[\frac{\partial(p_0, q_0)}{\partial(p_\tau, q_\tau)} \right] = 1$, as in any canonical transformation [see, for instance, L.D. Landau and E.M. Lifshitz, *Mechanics*, 3rd ed. (Elsevier, Amsterdam, 1976), p. 143 ff.]. Instead of averaging $\ln \rho_\tau$ as a function of (p, q) at an arbitrary time τ , we may integrate ρ_0 as a function of the initial point (p_0, q_0) which evolves into (p, q) at time τ . Note that (p_0, q_0) is common to all (p_τ, q_τ) belonging to this trajectory. Replacing Eq. (2.14) in Eq. (2.13) we may then write

$$S(\tau) = \int \rho_0[p_0(p, q), q_0(p, q)] \ln\{\rho_0[p_0(p, q), q_0(p, q)]\} dpdq. \quad (2.15)$$

It is now meaningful to change variables as $p' = p_0(p, q)$ and $q' = q_0(p, q)$. The volume elements, in general related by

$$dp'dq' = J \left[\frac{\partial(p', q')}{\partial(p, q)} \right] dpdq,$$

are in the present case the same since the Jacobian $J = 1$. Replacing with (p', q') in Eq. (2.15) we obtain

$$S(\tau) = \int \rho_0(p', q') \ln[\rho_0(p', q')] dp'dq' = S(0).$$

Therefore, the entropy of a strictly isolated system is not altered by time evolution as in the quantum case.

In order to discuss the consequences of $dS/dt = 0$ in strictly isolated systems, two different situations should be considered. If the system is in equilibrium, $\partial\rho/\partial t = 0$ and $\rho(p, q)$ does not depend explicitly on time at any representative point (p, q) . It is clear that S also remains constant as a function of time. However, if the system is out of equilibrium, with some value of S that is different from the equilibrium one, the above result implies that either no equilibrium is reached, no matter how large the system is, or that S remains constant along the equilibration process. This would be physically wrong and in contradiction with experimental observations. The reason for this apparent conceptual problem is that we have completely neglected the interaction with the environment. Physically, the interaction with the environment can never be avoided over a long period of time, even if the system can be considered to be closed and the allowed physical processes do not involve energy, particle or volume exchange. We conclude, as in the quantum case, that the environment plays a crucial role in the process of reaching equilibrium. These interactions cause changes in $\rho(p, q)$ which do not result from the dynamics of the isolated system alone. In fact, it can be shown, by using time-dependent perturbation theory in quantum mechanics and taking into account the transitions induced by the interaction with the environment, that $dS/dt \geq 0$ in closed systems. Processes in which S is conserved are reversible, while the others, having $dS > 0$, are irreversible (i.e., they cannot occur spontaneously in reverse direction). Consequently, the equilibrium states of isolated systems, not only have $dS/dt = 0$, but are such that the entropy takes the maximum value compatible with the given boundary conditions and constants of motion.

3 Equilibrium statistical ensembles

3.1 Statistical independence of macroscopic subsystems

The purpose of this chapter is to set the foundations of equilibrium statistical mechanics by introducing the basic physical hypothesis behind its validity and by deriving the form of the density operator $\hat{\rho}$ and the classical statistical distribution function $\rho(p, q)$ which define the equilibrium state. To get started we discuss in this section how $\hat{\rho}$ and $\rho(p, q)$ of a macroscopic system can be expressed in terms of the ρ 's of the numerous parts in which it can be divided. We consider a macroscopic system composed of a large number N of particles and divide it in a large but finite number M of subsystems which in the limit of very large N are macroscopic themselves. Since the interface to volume ratio of the subsystems decreases as $1/\sqrt[3]{N}$ and the effective interactions at the interfaces decrease rapidly as a function of distance (limited interaction range) the coupling between the different subsystems becomes less and less important as the size of the system increases ($N \rightarrow \infty$). Therefore, in the macroscopic limit, the microstates of the different subsystems become statistically independent from each other. For any two subsystems 1 and 2, the probability of finding subsystem 1 in a given quantum state $|\alpha_n\rangle$ [or in a classical representative point (p^α, q^α)] is independent of the state $|\beta_m\rangle$ [or (p^β, q^β)] which the subsystem 2 might take. In other words, the subsystem 1 can take a variety of different states $|\alpha_n\rangle$ without taking care of, or having any influence on the subsystem 2.

In the case of isolated systems, the fundamental conservation laws, which apply to the system as a whole, must be taken into account. For example, if we divide an isolated system in M parts, the states of these parts can be regarded as independent provided that the sums of all their energies, linear and angular momenta, number of particles, and any other constants of motion remain constant. This means that the global conservation laws preclude the system from assuming any microstate in which the constants of motion deviate from some predefined values. Thus, the conservation laws limit the number and nature of the accessible states. Consequently, the statistical independence of the different subsystems applies within the vector space defined by the system's accessible states. Notice that these restrictions do not apply to fully open systems, which can exchange constants of motion with the environment following no conservation constraints, or to subsystems which are a small part of a much larger isolated system.

As already discussed in Chapter 1, the statistical independence between any two events α and β means that the conditional probability $p(\alpha|\beta)$ for the event α to occur given the event β is equal to the unconditional probability $p(\alpha)$ of finding α . Knowing that

$$p(\alpha|\beta) = \frac{p(\alpha, \beta)}{p(\beta)},$$

where $p(\alpha, \beta)$ is the joint probability of finding α and β , we obtain that statistically independent events are characterized by the relation

$$p(\alpha, \beta) = p(\alpha)p(\beta).$$

Concerning the mixed state of the joint system 1+2, the statistical independence of the subsystems 1 and 2 implies that the probability density $\rho_{12}(p^\alpha, p^\beta, q^\alpha, q^\beta)$ of finding 1+2 in the state in which subsystem 1 occupies the representative point (p^α, q^α) and subsystem 2 the representative point (p^β, q^β) is given by

$$\rho_{12}(p^\alpha, p^\beta, q^\alpha, q^\beta) = \rho_1(p^\alpha, q^\alpha) \rho_2(p^\beta, q^\beta),$$

where ρ_1 and ρ_2 are the distribution functions of the subparts. The converse is of course also true. In fact, if ρ_{12} is given by the product of two distributions, each depending only on the variables of a subsystem, then the states of the subsystems are statistically independent random variables.

The same simple product relation holds for the density operators in the quantum case. Let

$$\hat{\rho}_{12} = \sum_{mn} w_{mn}^{12} |\alpha_m\rangle |\beta_n\rangle \langle \beta_n| \langle \alpha_m|$$

be the density operator of the 1+2 system, where w_{mn}^{12} is the probability of finding subsystem 1 in state $|\alpha_m\rangle$ and subsystem 2 in state $|\beta_n\rangle$. The statistical independence of the parts 1 and 2 implies that

$$w_{mn}^{12} = w_m^1 w_n^2 \quad \text{with} \quad \sum_m w_m^1 = \sum_n w_n^2 = 1.$$

Consequently,

$$\begin{aligned} \hat{\rho}_{12} &= \left(\sum_m w_m^1 |\alpha_m\rangle \langle \alpha_m| \right) \left(\sum_n w_n^2 |\beta_n\rangle \langle \beta_n| \right) \\ &= \hat{\rho}_1 \hat{\rho}_2. \end{aligned}$$

Note that $\hat{\rho}_1$ ($\hat{\rho}_2$) acts only on the variables of subsystem 1 (2) and therefore $[\hat{\rho}_1, \hat{\rho}_2] = 0$.

3.2 The statistical independence of extensive additive properties

The statistical independence of the subsystems of macroscopic system has far reaching consequences. For example, for any two physical observables \hat{f}_1 and \hat{f}_2 concerning, respectively, the subsystems 1 and 2, the mean value of the product is given by

$$\langle \hat{f}_1 \hat{f}_2 \rangle = \text{Tr}\{\hat{\rho}_{12} \hat{f}_1 \hat{f}_2\} = \left(\sum_m w_m^1 \langle \alpha_m | \hat{f}_1 | \alpha_m \rangle \right) \left(\sum_n w_n^2 \langle \beta_n | \hat{f}_2 | \beta_n \rangle \right) = \langle \hat{f}_1 \rangle_1 \langle \hat{f}_2 \rangle_2.$$

Let us now focus on an additive (also known as extensive) property \hat{f} of a macroscopic system, which we imagine divided in a large number N of statistically independent subsystems. Let \hat{f}_i be the operator of this property in the subsystem i . The additivity of \hat{f} implies that for the whole system we have

$$\hat{F} = \sum_{i=1}^N \hat{f}_i.$$

If the subsystems are equivalent (same composition and size) the probability distribution $P_{\hat{f}_i}(f)$ for the values of \hat{f}_i is the same in all subsystems, i.e.,

$$P_{\hat{f}_i}(f) = \langle \delta(\hat{f}_i - f) \rangle_i = p(f)$$

is independent of i . The joint probability of measuring f_1, \dots, f_n in the various subsystems is then

$$P(f_1, \dots, f_N) = \prod_{i=1}^N p(f_i).$$

Thus, the hypotheses of the central limit theorem are fulfilled, since the values of f in the various subsystems are independent random variables governed by the same probability density function. It follows that the value of \hat{F} in the whole system, the sum of the \hat{f}_i , follows a Gaussian distribution with average

$$\langle \hat{F} \rangle = \sum_{i=1}^N \langle f_i \rangle = N \langle f \rangle$$

and variance

$$\begin{aligned} \langle (\hat{F} - \langle \hat{F} \rangle)^2 \rangle &= \sum_{i=1}^N \langle (\hat{f}_i - \langle \hat{f}_i \rangle)^2 \rangle + \sum_{i \neq j} \underbrace{\langle (\hat{f}_i - \langle \hat{f}_i \rangle) \rangle}_{=0} \underbrace{\langle (\hat{f}_j - \langle \hat{f}_j \rangle) \rangle}_{=0} \\ \langle \Delta F^2 \rangle &= N \langle \Delta f^2 \rangle. \end{aligned}$$

Consequently, the relative fluctuation of \hat{F} in the entire macroscopic system, as measured by the relative mean-square deviation, is given by

$$\frac{\sqrt{\langle \Delta F^2 \rangle}}{\langle \hat{F} \rangle} = \frac{1}{\sqrt{N}} \frac{\sqrt{\langle \Delta f^2 \rangle}}{\langle f \rangle},$$

which vanishes for $N \rightarrow \infty$. One concludes that the relative fluctuation of any additive property of a macroscopic system decreases as $1/\sqrt{N_p}$, where N_p refers to the number of particles in the system. When the body is macroscopic and $N_p \rightarrow \infty$, the extensive quantities practically do not fluctuate in relative terms, even if they are not strictly conserved. The absolute fluctuations $\sqrt{\langle \Delta F^2 \rangle}$ actually *increase* with size, although much more slowly than the average value, such that the relative fluctuations vanish. Notice that this holds even if the system is open with respect to exchanges of F . For instance, if a macroscopic system can exchange energy through a thermal contact, its energy per particle can be regarded as practically conserved in average. Indeed, the probability of observing a value of an additive property F that differs from the average per particle is negligibly small. More precisely, the probability of observing a relative deviation $\Delta f = (F - \langle \hat{F} \rangle) / \langle \hat{F} \rangle$ which is larger than any whatever small amount δ decreases exponentially as N increases, since the standard deviation σ_N of the Gaussian distribution of Δf decreases as $\sigma_N = \sigma / \sqrt{N}$ for $N \gg 1$. In terms of the error function $\text{erf}(x)$ and its

approximation for large x one may write $\text{Prob}(|\Delta f| \geq \delta) = 1 - \text{erf}(\delta/\sqrt{2}\sigma_N) = 1 - \text{erf}(\delta\sqrt{N}/\sqrt{2}\sigma) \simeq (\sqrt{2}\sigma/\sqrt{\pi}\delta\sqrt{N}) \exp\{-\delta^2 N/2\sigma^2\}$, which tends rapidly to zero for $N \rightarrow \infty$.

In conclusion, one may note that the previous considerations hold for any additive property F and for any form of its statistical distribution $p(f)$, provided that it is the same in all subsystems (central limit theorem). This implies that the equilibrium properties of closed and open macroscopic bodies are physically equivalent, even if the different boundary conditions on the microstates accessible to the system, namely, conservation versus exchange of F , certainly lead to different statistical distributions.

3.3 The importance of additive constants of motion

From now on we focus on conservative systems, i.e., systems whose Hamiltonian does not depend on time. Our purpose is to derive the form of $\rho(p, q)$ and $\hat{\rho}$ in thermodynamic equilibrium, i.e., when $\rho(p, q)$ and $\hat{\rho}$ do not depend explicitly on time. Constants of motion play a central role in both classical and quantum mechanics. They are extremely important in statistical mechanics as well.

From mechanics we know that the constants of motion reflect fundamental symmetries of the physical laws, for instance, the homogeneity of space and time and the isotropy of space. They allow us to classify the classical trajectories and the quantum stationary states according to their values and are therefore important for understanding the physical behavior. Having said this, it is clear that it makes no sense to consider all possible constants of motion, since they are infinite. Any linear combination, product or function of constants of motion is obviously a constant of motion too. In fact, it is sufficient to consider a *complete set of linearly independent additive constants of motion*, since all other constants of motion can be expressed in terms of them. In the following we denote by $\{F_i, \text{ with } i = 1, \dots, s\}$ the complete set of linearly independent additive constants of motion of the system. In classical mechanics the observable F_i is a function $F_i(p, q)$ of the generalized coordinates and conjugated momenta of the particles, while in quantum mechanics F_i is characterized by the hermitic operator \hat{F}_i with eigenvalues f_i .

The additive constants of motion are well known from mechanics:

- i) The total energy $H(p, q)$ or \hat{H} , which reflects the homogeneity of time, i.e., the invariance of the physical laws with respect to translations in time,
- ii) the total linear momentum $\vec{P} = (p_x, p_y, p_z)$, which reflects the homogeneity of space, i.e., the invariance with respect to translations of the system as a whole, and
- iii) the total angular momentum $\vec{J} = (J_x, J_y, J_z)$, which reflects the isotropy of space and the invariance with respect to rotations.

To these seven fundamental additive integrals of motion we should add

- iv) the volume V of the system, and

- v) the total number of particles N . In multi-component systems, the number N_i of each type of particle is conserved. Moreover, under certain circumstances (neglect of spin-orbit interactions) the number of electrons with spin up and spin down are separately conserved.

Notice that the last two constants of motion are actually parameters that define the system. They are not operators in the usual sense and in classical mechanics they do not depend on the generalized coordinates (p, q) . Still, in quantum mechanics the particle number \hat{N} becomes an operator when the second-quantization formalism is used.

In almost all cases of experimental interest, one considers macroscopic bodies which are at rest and do not rotate.³ One can therefore imagine the system being in a fixed box or volume which is at rest. In this case \vec{P} and \vec{J} are no longer constants of motion, which leaves us with E , N and V as the only additive linearly independent constants of motion. Whether one only needs to consider E , N and V or a larger set $\{F_i$, with $i = 1, \dots, s\}$ of additive constant of motion (including for example the magnetization \vec{M} , the number of up-spin and down-spin electrons, or the total angular momentum) depends on the problem under study. However, it is in all cases important to realize that any other constant of motion is necessarily a function of the complete set of linearly independent ones.

Constants of motion are characterized by having a vanishing Poisson bracket or commutator with the Hamiltonian. As stated by the Liouville and von Neumann theorems, the distribution function $\rho(p, q)$ and the density operator $\hat{\rho}$ in equilibrium ($\partial\hat{\rho}/\partial t = 0$) fulfill this condition and are thus constants of motion. Furthermore, the statistical independence of subsystems implies that $\ln \hat{\rho}$ is an *additive constant of motion*. Indeed, $\hat{\rho}_{12} = \hat{\rho}_1 \hat{\rho}_2$ and $[\hat{\rho}_1, \hat{\rho}_2] = 0$ imply

$$\ln \hat{\rho}_{12} = \ln(\hat{\rho}_1 \hat{\rho}_2) = \ln \hat{\rho}_1 + \ln \hat{\rho}_2$$

for any two sufficiently large and thus statistically independent parts of a macroscopic system.⁴ Also in the classical limit we have

$$\ln \rho_{12}(p_1, p_2, q_1, q_2) = \ln[\rho_1(p_1, q_1)\rho_2(p_2, q_2)] = \ln[\rho_1(p_1, q_1)] + \ln[\rho_2(p_2, q_2)].$$

Consequently, the equilibrium $\ln \hat{\rho}$ must be a linear combination of the linearly independent additive constants of motion F_1, \dots, F_s (i.e., $F_i \equiv E, \vec{P}, \vec{L}, V, N$):

$$\ln \hat{\rho} = \sum_{i=1}^s \lambda_i \hat{F}_i, \quad (3.1)$$

or in classical mechanics

$$\ln \rho(p, q) = \sum_{i=1}^s \lambda_i F_i(p, q), \quad (3.2)$$

³The reader may wish to examine the Einstein-de Haas effect for a remarkable exception. A. Einstein and W.J. de Haas, Verh. Dtsch. Phys. Ges. **17**, 152 (1915); **18**, 173 (1916).

⁴Notice that $[\hat{\rho}_1, \hat{\rho}_2] = 0$, since the kets of the two subsystems are orthogonal to each other. This is a necessary condition for the validity of the operator relation $\ln(\hat{\rho}_1 \hat{\rho}_2) = \ln \hat{\rho}_1 + \ln \hat{\rho}_2$.

with some real coefficients λ_i . This is the central expression for the density operator in equilibrium. In particular it implies that the equilibrium density operator and distribution function can only depend on the dynamical variables (p, q) through the additive constants of motion. *All microstates having the same additive constants of motion are equally probable!*

The physical meaning of the coefficient λ_i associated with the constant of motion F_i in the linear combinations (3.1) and (3.2) will be discussed in detail later on. Since $\ln \rho$ is an additive or extensive property, i.e., $\ln[\rho(\alpha F_1, \dots, \alpha F_s)] = \alpha \ln[\rho(F_1, \dots, F_s)]$, we may already anticipate that $\lambda_1, \dots, \lambda_s$ do not change when all F_i are scaled by the same factor α . Such properties are called *intensive*. We shall see that they are related to the temperature, pressure and chemical potential of the system.

3.4 The density operator: General formulation

In the following we derive the general form of the density operator describing mixed states in different equilibrium situations ranging from complete isolation—in which case all the additive constants of motion $\hat{F}_1, \dots, \hat{F}_s$ have the same well-defined values $f_1 \dots f_s$ in all microstates—to complete openness—in which case the system can exchange freely all its additive constants of motion with the environment and only the average values $f_1 = \langle \hat{F}_1 \rangle \dots f_s = \langle \hat{F}_s \rangle$ are well defined. The formulation is admittedly somewhat abstract. For instance, the conserved quantities are denoted by \hat{F}_i with $i = 1, \dots, s$, instead of referring to them explicitly as energy E , volume V , particle number N , et cetera. This has the advantage of stressing the universal validity of the approach, making it easier to apply the formalism to any specific situation, for example, spin-polarized or multicomponent systems. Moreover, in this way we emphasize the analogies and differences between the various usually encountered physical situations including the corresponding statistical ensembles. The explicit form of the density operator in the most usual cases of physical interest is given in Sec. 3.5.

3.4.1 The microcanonical ensemble

First of all we consider systems which are closed with respect to all constants of motion F_1, \dots, F_s . In order to obtain $\hat{\rho}$ one could take the exponential of $\ln \hat{\rho}$ as given by Eqs. (3.1) and (3.2) and impose the necessary constraints on the conserved quantities F_1, \dots, F_s . However, in the case of closed systems a much more direct derivation is possible. In isolated systems the constants of motion have precise values f_i for all $i = 1, \dots, s$. Consequently, the probability of finding the system in a microstate having $F_i \neq f_i$ for some i is strictly zero. The classical $\rho(p, q) = 0$ for all (p, q) having $F_i(p, q) \neq f_i$ for some $i \in [1, s]$. The same holds in quantum mechanics: $w_n = 0$ for all states $|\alpha_n\rangle$ for which $\hat{F}_i|\alpha_n\rangle \neq f_i|\alpha_n\rangle$ for some i . In addition, we know that the integral or trace of $\hat{\rho}$ over all possible states must be 1. The only function satisfying these conditions is the delta

function. It follows that the microcanonic density matrix is given by

$$\hat{\rho}_{mc} = \frac{1}{Z_{mc}} \prod_{i=1}^s \delta(\hat{F}_i - f_i), \quad (3.3)$$

where the normalization divisor

$$Z_{mc} = \text{Tr} \left\{ \prod_{i=1}^s \delta(\hat{F}_i - f_i) \right\}, \quad (3.4)$$

ensures that $\text{Tr}\{\hat{\rho}\} = 1$. Z_{mc} is known as the *microcanonical partition function*. In the classical case we have

$$\rho_{mc}(p, q) = \frac{1}{Z_{mc}} \prod_{i=1}^s \delta[F_i(p, q) - f_i] \quad (3.5)$$

with

$$Z_{mc} = \int dp dq \prod_{i=1}^s \delta[F_i(p, q) - f_i]. \quad (3.6)$$

Notice that the mixed state describing equilibrium, i.e., the macrostate, is fully characterized by the very few numbers f_1, \dots, f_s . Out of the 10^{20} degrees of freedom needed to characterize a single microstate, we are left with typically 3, or only 9 variables at most! The statistical ensemble characterized by $\hat{\rho}_{mc}$ or $\rho_{mc}(p, q)$ as given by Eqs. (3.3)–(3.6) is known as the *microcanonical ensemble*. For closed systems, $\hat{\rho}_{mc}$ and Z_{mc} are defined by the values f_1, \dots, f_s of the constants of motion $\hat{F}_1 \dots \hat{F}_s$ in the particular macrostate under consideration.

The expressions given above are valid for any physical system and any set of linearly independent additive constants of motion. They can be significantly simplified in the most usual situation where $\vec{P} = 0$ and $\vec{L} = 0$. For any fixed volume V and number of particles N we may write

$$\hat{\rho}_{mc} = \frac{1}{Z_{mc}} \delta(\hat{H} - E)$$

with the *microcanonical partition function*

$$Z_{mc} = \text{Tr} \left\{ \delta(\hat{H} - E) \right\}.$$

Here we have implicitly assumed that the Hilbert space in which the trace is calculated contains only states having N particles in a volume V . $\hat{\rho}_{mc}$ and Z_{mc} depend thus on E , V and N .

For the sake of completeness, and in view of later comparison, let us mention that one can start from the general expression (3.3) for $\ln \hat{\rho}$ and express the density operator of a closed system as

$$\hat{\rho}_{mc} = e^{\sum_{i=1}^s \lambda_i \hat{F}_i} \prod_{j=1}^s \delta(\hat{F}_j - f_j),$$

where the product of delta operators plays the role of projector ensuring that no microstates involved in $\hat{\rho}_{mc}$ has $\hat{F}_j \neq f_j$ for some j . Since $\hat{F}_j = f_j$ for all j we may replace the operators in the exponential by the corresponding eigenvalues. In this way we obtain

$$\hat{\rho}_{mc} = e^{\sum_{i=1}^s \lambda_i f_i} \prod_{j=1}^s \delta(\hat{F}_j - f_j),$$

which we can write as

$$\hat{\rho}_{mc} = \frac{1}{Z_{mc}} \prod_{i=1}^s \delta(\hat{F}_i - f_i), \quad (3.7)$$

where we have identified the microcanonical partition function

$$Z_{mc} = e^{-\sum_{i=1}^s \lambda_i f_i}.$$

Knowing that $\text{Tr}\{\hat{\rho}_{mc}\} = 1$ and applying it to Eq. (3.7), we recover Eq. (3.4).

3.4.2 The grand canonical ensemble

We now turn our attention to systems that are open with respect to exchange of all additive constants of motion. These systems can be regarded as parts of a much larger system which can be isolated or not. The statistical ensemble describing completely open systems is known as the *grand canonical ensemble*. By completely open we mean that none of the constant of motion $F_1 \dots F_s$ is fixed in the system. For example, energy, particles and even accessible volume can be exchanged with the environment. The role of the environment is only to define the average values $\langle \hat{F}_i \rangle = f_i$ in the system ($i = 1, \dots, s$). In this case there are no restrictions on the values that the constants of motion \hat{F}_i can take in the microstates involved in the mixed state $\hat{\rho}$. Taking the exponential of Eq. (3.1) we obtain the density operator

$$\hat{\rho}_{gc} = e^{\sum_{i=1}^s \lambda_i \hat{F}_i}. \quad (3.8)$$

The grand canonical ensemble or mixed state $\hat{\rho}_{gc}$ is defined by the coefficients $\lambda_1, \dots, \lambda_s$. In particular they determine the average values

$$f_i = \langle \hat{F}_i \rangle = \text{Tr} \left\{ \hat{\rho}_{gc} \hat{F}_i \right\} = \text{Tr} \left\{ e^{\sum_{i=1}^s \lambda_i \hat{F}_i} \hat{F}_i \right\} \quad (3.9)$$

of all the additive constants of motion \hat{F}_i ($i = 1, \dots, s$). Alternatively, one may consider that the averages of \hat{F}_i take some predefined values f_1, \dots, f_s and solve for the $\lambda_1, \dots, \lambda_s$ which satisfy the coupled Eqs. (3.9) for all i .

An important relation between the λ_i follows from Eq. (3.8) and the normalization of $\hat{\rho}_{gc}$:

$$\text{Tr}\{\hat{\rho}_{gc}\} = \text{Tr}\{e^{\sum_{i=1}^s \lambda_i \hat{F}_i}\} = 1. \quad (3.10)$$

Thus, the coefficients λ_i ($i = 1, \dots, s$) are not all independent from each other, even in a completely open system. In this context it is important to keep in mind that the

operators \hat{F}_i describing the linearly independent additive constants of motion (i.e., the Hamiltonian \hat{H} , momentum operator \hat{P} , etc.) do not depend at all on λ_i . Their average values do. Later on, we shall see that the λ_i are intensive (i.e., size-scaling independent) properties of the system such as temperature, pressure or density. The normalization condition (3.10) expresses a relation between all the λ_i , which is known as the *equation of state* of the material.

3.4.3 The canonical ensemble

The *canonical ensemble* describes the most general intermediate situation in which the system is open with respect to the first r additive constants of motion F_1, \dots, F_r and closed with respect to the remaining ones F_{r+1}, \dots, F_s . Concerning the constants of motion F_1, \dots, F_r , for which exchange is allowed, only the corresponding averages f_1, \dots, f_r can be known. As in the grand canonical case there are no restrictions on the values that these constants of motion can take in the ensemble. Concerning the constants of motion F_{r+1}, \dots, F_s , with respect to which no exchange is possible, we must impose, as in the microcanonical case, that they take the precise values f_{r+1}, \dots, f_s . The canonical density operator is therefore given by

$$\hat{\rho}_c = e^{\sum_{i=1}^s \lambda_i \hat{F}_i} \prod_{j=r+1}^s \delta(\hat{F}_j - f_j),$$

where we have used Eq. (3.1) and imposed the lack of fluctuations of $\hat{F}_{r+1}, \dots, \hat{F}_s$ by means of the delta operators. Since $\hat{\rho} = 0$ unless $\hat{F}_j = f_j$ for $r+1 \leq j \leq s$ we may replace the operators $\hat{F}_{r+1}, \dots, \hat{F}_s$ in the exponential by the corresponding eigenvalues and obtain

$$\hat{\rho}_c = e^{\sum_{i=1}^r \lambda_i \hat{F}_i} \prod_{j=r+1}^s \delta(\hat{F}_j - f_j) e^{\sum_{j=r+1}^s \lambda_j f_j}.$$

Using that $\text{Tr}\{\hat{\rho}_c\} = 1$, we can express the multiplying constant as

$$e^{-\sum_{j=r+1}^s \lambda_j f_j} = \text{Tr} \left\{ e^{\sum_{i=1}^r \lambda_i \hat{F}_i} \prod_{j=r+1}^s \delta(\hat{F}_j - f_j) \right\}.$$

This allows us to write the *canonical density operator* in the more compact form

$$\hat{\rho}_c = \frac{1}{Z_r} e^{\sum_{i=1}^r \lambda_i \hat{F}_i} \prod_{j=r+1}^s \delta(\hat{F}_j - f_j), \quad (3.11)$$

where

$$Z_r = \text{Tr} \left\{ e^{\sum_{i=1}^r \lambda_i \hat{F}_i} \prod_{j=r+1}^s \delta(\hat{F}_j - f_j) \right\} \quad (3.12)$$

$$= e^{-\sum_{j=r+1}^s \lambda_j f_j} \quad (3.13)$$

is the *canonical partition function*.

Z_r has a great practical importance in statistical mechanics, since it can be directly related to the thermodynamic potential or state function of the corresponding physical situation. Both Z_r and $\hat{\rho}_c$, as well as the mixed state they represent, are determined by the parameters $\lambda_1, \dots, \lambda_r$ together with the constants of motion f_{r+1}, \dots, f_s that delimit the Hilbert or phase space of accessible microstates [see Eqs. (3.11) and (3.12)]. Neither Z_r nor $\hat{\rho}_c$ depend on the parameters $\lambda_{r+1}, \dots, \lambda_s$ associated with the quantities $\hat{F}_{r+1}, \dots, \hat{F}_s$ with respect to which the system is closed [see Eqs. (3.11) and (3.12)]. One should, however, mention the useful relation $\ln Z_r = -\sum_{j=r+1}^s \lambda_j f_j$, which follows from Eq. (3.13) and which we shall exploit systematically later on.

Once the natural variables $\lambda_1, \dots, \lambda_r$ and f_{r+1}, \dots, f_s of the canonical ensemble are given it is straightforward to determine the mean values

$$\langle F_i \rangle = \text{Tr}\{\hat{\rho}_c \hat{F}_i\} = f_i \quad (3.14)$$

of the constants of motion \hat{F}_i for which fluctuations and exchanges with the environment are possible ($i = 1, \dots, r$). Alternatively, one may consider that the Eqs. (3.14) express the conditions that $\lambda_1, \dots, \lambda_r$ must satisfy in order that the averages of $\hat{F}_1 \dots \hat{F}_r$ take some predefined values f_1, \dots, f_r for the given fixed values f_{r+1}, \dots, f_s of the remaining constants of motion.

The expressions (3.11) and (3.12) for $\hat{\rho}_c$ and Z_r describe the most general density operator in thermal equilibrium. They include the grand canonical ensemble ($r = s$) and the microcanonical ensemble ($r = 0$) as particular cases. In the grand canonical ensemble we have simply

$$Z_s = 1,$$

and in the microcanonical ensemble we have

$$Z_0 = Z_{mc} = \text{Tr} \left\{ \prod_{j=1}^s \delta(\hat{F}_j - f_j) \right\}.$$

3.5 Explicit forms of the density operator

After having discussed the general formulation, we would like to apply the formalism to the most important cases of physical interest. Before doing that let us recall that the additive constants of motion in mechanics are

- i) the energy E given by \hat{H} ,
- ii) the linear momentum \vec{P} ,
- iii) the angular momentum \vec{L} ,
- iv) the number of particles N (eventually N_i for different components) and
- v) the volume V .

Each one of these properties is clearly related to a particular symmetry of the physical laws or to a boundary condition. In the following we focus on the most usual experimental situation in which the system is at rest, so that $\langle \vec{P} \rangle = 0$ and $\langle \vec{L} \rangle = 0$. This is experimentally realized by keeping the system in a fixed container. Under these circumstances the total \vec{P} and \vec{L} are no longer conserved due to the collisions and external constraints at the boundary of the volume. Although the system is open with respect to linear and angular momentum, no net transfer of linear or angular momentum takes place with the environment, since $\langle \vec{P} \rangle = 0$ and $\langle \vec{L} \rangle = 0$. Moreover, symmetry requires the probability of the microstates having a nonvanishing \vec{P} or \vec{L} to be independent of their orientation in space. This can only be achieved if the coefficients λ_i associated with \vec{P} and \vec{L} are all zero [see Eqs. (3.1) and (3.2)]. Thus, the terms involving \vec{P} and \vec{L} disappear altogether from $\hat{\rho}$.

3.5.1 Microcanonical ensemble

The ensemble corresponding to a system which is closed with respect to all remaining additive constants of motion E , N and V , is known as *microcanonical ensemble*. The microcanonical density operator can be written as

$$\hat{\rho}_{mc} = \frac{1}{Z_{mc}} \delta(\hat{H} - E)$$

with

$$Z_{mc} = \text{Tr}\{\delta(\hat{H} - E)\},$$

where it is understood that the Hilbert space on which $\hat{\rho}$ acts, and within which the partition function Z_{mc} is calculated, contains only states having a fixed volume V and a fixed number of particles N . $\hat{\rho}_{mc}$ and Z_{mc} depend thus on E , V and N .

Let us discuss the physical meaning of Z_{mc} . The number of states having an energy lower than E is given by

$$\Sigma(E) = \sum_n \theta(E - E_n) = \text{Tr}\{\theta(E - \hat{H})\}, \quad (3.15)$$

where $\theta(x)$ is the Heaviside function [$\theta(x) = 1$ for $x > 0$ and $\theta(x) = 0$ for $x < 0$] with $d\theta/dx = \delta(x)$. The number of states $\Gamma(E)$ having an energy in the interval $[E, E + \Delta]$ is

$$\Gamma(E) = \Sigma(E + \Delta) - \Sigma(E) = \frac{d\Sigma}{dE} \Delta + O(\Delta^2) = \Omega(E) \Delta + O(\Delta^2),$$

where we have introduced the density of states of the system $\Omega(E) = d\Sigma/dE$ at the energy E and assumed $\Delta \rightarrow 0$. $\Omega(E)$ represents the number of accessible states at the energy E per unit energy, i.e., the number of states per unit energy in the spectrum of the system at the energy E . From Eq. (3.15) it is clear that

$$\Omega(E) = \frac{d\Sigma}{dE} = \text{Tr}\{\delta(\hat{H} - E)\} = Z_{mc}(E).$$

The microcanonical partition function Z_{mc} represents the *density of accessible states at the energy E* for the given total volume V and particle number N . Note that Z_{mc} is a function of E , V and N , which reflects the nature of the system under study through the eigenvalue spectrum of the Hamiltonian \hat{H} and its dependence on V and N .

In the microcanonical ensemble all the states having the same energy E are equally probable. In classical mechanics all the states (p, q) belonging to the hypersurface $H(p, q) = E$ are equally probable. The latter is of course consistent with ρ being a constant of motion and with the ergodic hypothesis. Notice, however, that we didn't invoke any ergodic hypothesis in order to derive $\rho(p, q)$ but simply the concept of statistical independence of the subsystems in the macroscopic limit.

Exercise 3.11: Compute the density of states $\Omega(E)$ of a system of N independent identical particles of mass m in a volume $V = L^D$ in D dimensions. Suppose that each particle has the energy ε_i ($\sum_i \varepsilon_i = E$) and calculate the number of states $\sigma(\varepsilon_i)$ for one particle having the energy $\varepsilon < \varepsilon_i$. Derive then total number of states $\Sigma(\varepsilon_1, \dots, \varepsilon_N)$ for the case when the individual particles have the energies $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$. How would this expression be modified, if you take into account the principle of indistinguishability of identical particles? Maximize $\Sigma(\varepsilon_1, \dots, \varepsilon_N)$ under the total energy constraint $\sum_i \varepsilon_i = E$ and approximate $\Sigma(E)$ by the maximum value. Derive $\Omega(E)$ and analyze the dependence of Σ and Ω on E , V , N and $v = V/N$.

3.5.2 Canonical ensemble

A system having a fixed volume V and a fixed number of particles N , which can exchange energy with environment is described by the *canonical ensemble* in a narrow sense. In these cases one says that the system has a heat contact with the environment. In fact, the energy exchanged in this way, by keeping V and N fixed, is commonly known as *heat*. The canonical density operator is given by

$$\hat{\rho}_c = \frac{1}{Z_c} e^{-\beta \hat{H}}$$

with

$$Z_c = \text{Tr}\{e^{-\beta \hat{H}}\},$$

where we have implicitly assumed that the Hilbert space in which the trace is calculated consists of all the microstates having N particles in the volume V . Thus, $\hat{\rho}_c$ and Z_c depend on β , V and N . The parameter β defines, together with N and V , the average

energy E of the system:

$$\begin{aligned}
E(\beta) &= \langle \hat{H} \rangle = \text{Tr}\{\hat{\rho}_c \hat{H}\} \\
&= \frac{\text{Tr}\{\hat{H} e^{-\beta \hat{H}}\}}{\text{Tr}\{e^{-\beta \hat{H}}\}} \\
&= -\frac{\partial}{\partial \beta} \ln Z_c.
\end{aligned}$$

It is important to note that the converse is also true. The condition that the average energy $\langle \hat{H} \rangle$ takes the value E , for the given V and N , defines β univocally, since $E(\beta) = \langle \hat{H} \rangle$ is a strictly monotonically decreasing function of β :

$$\begin{aligned}
\frac{\partial E}{\partial \beta} &= \frac{\partial \langle \hat{H} \rangle}{\partial \beta} = -\frac{\text{Tr}\{\hat{H}^2 e^{-\beta \hat{H}}\} \text{Tr}\{e^{-\beta \hat{H}}\} - (\text{Tr}\{\hat{H} e^{-\beta \hat{H}}\})^2}{(\text{Tr}\{e^{-\beta \hat{H}}\})^2} \\
&= -\langle \hat{H}^2 \rangle + \langle \hat{H} \rangle^2 = -\langle (\hat{H} - \langle \hat{H} \rangle)^2 \rangle = -\Delta H^2 < 0.
\end{aligned}$$

The variance ΔH^2 is always strictly positive and tends to zero only for $\beta \rightarrow \infty$. Thus, the relation between E and β is bijective.

In the canonical ensemble the total energy E of the system is not fixed, since the system is open with respect to energy exchange. In the corresponding mixed state, only the average value $\langle \hat{H} \rangle$ is well defined. The statistical distribution $w(E)$ of E , i.e., the probability density $w(E)$ of finding the energy E in the system, is given by

$$\begin{aligned}
w(E) &= \langle \delta(E - \hat{H}) \rangle \\
&= \frac{1}{Z_c} \text{Tr}\{e^{-\beta \hat{H}} \delta(E - \hat{H})\} = \frac{1}{Z_c} \text{Tr}\{e^{-\beta E} \delta(E - \hat{H})\} \\
&= \frac{1}{Z_c} e^{-\beta E} \Omega(E). \tag{3.16}
\end{aligned}$$

In words, the probability of measuring the value E of the total energy is equal to the probability $e^{-\beta E}/Z_c$ of finding the system in a microstate of energy E multiplied by the number of states $\Omega(E)$ having this energy. Again, in equilibrium, all microstates having the same energy are equally probable.

The canonical partition function Z_c can be readily related to the microcanonical par-

tition function $Z_{mc}(E) = \Omega(E)$ by noting that

$$\begin{aligned}
Z_c &= \text{Tr}\{e^{-\beta\hat{H}}\} \\
&= \text{Tr}\left\{e^{-\beta\hat{H}} \int dE \delta(E - \hat{H})\right\} \\
&= \int dE \text{Tr}\left\{e^{-\beta\hat{H}} \delta(E - \hat{H})\right\} \\
&= \int dE e^{-\beta E} \text{Tr}\{\delta(E - \hat{H})\} \\
&= \int dE e^{-\beta E} \Omega(E). \tag{3.17}
\end{aligned}$$

Notice that this could have been directly inferred from Eq. (3.16) by recalling that the probability density $w(E)$ is normalized.

In the canonical ensemble the total energy E of the system does not have well defined value. It is therefore very interesting to analyze its probability distribution $w(E)$ in order to assess quantitatively the importance of energy fluctuations. In the following we will show that $w(E)$ is very sharply peaked at the average value. To this aim let us first note that the density of states $\Omega(E)$ of a macroscopic system increases extremely rapidly with E . We already know from our discussion the of statistical independence of subsystems that $\ln \Omega$ is an additive property. We may therefore write

$$\ln \Omega(E) = N \ln \left[\omega \left(\frac{E}{N}, \frac{V}{N} \right) \right] \quad \Leftrightarrow \quad \Omega(E) = \omega \left(\frac{E}{N}, \frac{V}{N} \right)^N,$$

where $\omega(\varepsilon, v)$ is some increasing function of the energy per particle $\varepsilon = E/N$ representing the density of states per particle or per subsystem. We conclude that the probability density distribution $w(E)$ for the total energy E is the product of the very rapidly increasing function $\Omega(E)$ and the very rapidly decreasing function $e^{-\beta E}$ [see Eq. (3.16)]. Consequently, $w(E)$ shows an extremely narrow maximum. In addition, we already know, from the central limit theorem, that the relative fluctuation $\sqrt{\Delta H^2}/\langle H \rangle \propto 1/\sqrt{N} \rightarrow 0$ for $N \rightarrow \infty$.

Furthermore, in the macroscopic limit ($N \rightarrow \infty$) we may identify the average energy $E = \langle \hat{H} \rangle$ with the most probable energy \bar{E} , i.e., with the energy at which $w(E)$ has its maximum (saddle point integration). Indeed, the average energy is given by

$$\begin{aligned}
E = \langle \hat{H} \rangle &= \int E' w(E') dE' \\
&= \int E' \frac{e^{-\beta E'} \Omega(E')}{Z_c} dE' \xrightarrow{N \rightarrow \infty} \int E' g(E' - \bar{E}) dE' = \bar{E},
\end{aligned}$$

where we have used that $g(E' - \bar{E}) = e^{-\beta E'} \Omega(E')/Z_c$ has a very narrow peak at \bar{E} for $N \rightarrow \infty$, which satisfies $\int g(E' - \bar{E}) dE' = 1$ and converges to a Gaussian distribution centered at \bar{E} .

Consequently, the average E can be obtained from the condition

$$\left. \frac{\partial w(E)}{\partial E} \right|_{E=\bar{E}} = 0.$$

Knowing that

$$\frac{\partial w}{\partial E} = \frac{1}{Z_c} \left[-\beta e^{-\beta E} \Omega(E) + e^{-\beta E} \frac{\partial \Omega}{\partial E} \right],$$

we have

$$\frac{\partial w}{\partial E} = 0 \Leftrightarrow -\beta \Omega + \frac{\partial \Omega}{\partial E} = 0 \Leftrightarrow \beta = \frac{1}{\Omega(E)} \frac{\partial \Omega}{\partial E},$$

which implies

$$\beta = \left. \frac{\partial \ln \Omega}{\partial E} \right|_{N,V}.$$

The fact that the average energy $\langle \hat{H} \rangle$ and the most probable energy \bar{E} coincide in the macroscopic limit ($N \rightarrow \infty$) allows us to directly obtain β as a function of E , without having to invert the equation $E(\beta) = \text{Tr}\{\hat{\rho}_c \hat{H}\}$. Since $\Omega(E)$ is in most cases an increasing function of E , we usually have $\beta > 0$, although there are some remarkable exceptions.

The spectrum of macroscopic systems has a lower bound, the ground-state energy E_0 , and is usually unbound for high energies. The density of states is an increasing function of E and therefore β is in most cases positive. Notice that the trace giving the partition function converges only if the spectrum is bounded at least on one side. If the spectrum has only a lower bound, the trace converges only for $\beta > 0$. There are, however, systems in which the spectrum is bounded on both sides, i.e., in which there is maximum achievable energy E_{max} . An example of bounded spectrum is found in spin systems. In these cases $\Omega(E)$ increases first with E , and then decreases as we approach the upper bound. Thus, for energies close to the E_{max} one finds $\beta < 0$. As a simple example, consider the Ising model with nearest neighbor interactions in a one dimensional chain:

$$\hat{H} = -J \sum_{i=1}^N s_i s_{i+1}$$

with $s_i = \pm 1$ and $J > 0$. In this case one has $E_0 = -JN$ (parallel spins or ferromagnetic order) and $E_{max} = JN$ (alternating spins or antiferromagnetic order).

3.5.3 Grand Canonical ensemble

The equilibrium state of systems which are open with respect to energy and particle exchange is described by the *grand canonical ensemble*. Examples of such systems are metals kept at some given electrostatic potential, where electrons can be freely exchanged,

or chemical solutions exchanging ions with the electrodes. The corresponding density operator is given by

$$\hat{\rho}_{gc} = \frac{1}{Z_{gc}} e^{-\beta(\hat{H} - \sum_i \mu_i \hat{N}_i)}$$

with the partition function

$$Z_{gc} = \text{Tr} \left\{ e^{-\beta(\hat{H} - \sum_i \mu_i \hat{N}_i)} \right\}.$$

The operator \hat{N}_i counts the number of particles of each component i in the system. Both ρ_{gc} and Z_{gc} depend on β , μ_i and V . Consequently, the parameters β and μ_i define the average values of the total energy

$$\begin{aligned} E = \langle \hat{H} \rangle &= \text{Tr} \{ \hat{\rho}_{gc} \hat{H} \} \\ &= -\frac{\partial}{\partial \beta} \ln Z_{gc} \end{aligned} \quad (3.18)$$

and of the particle numbers

$$\begin{aligned} N_i = \langle \hat{N}_i \rangle &= \text{Tr} \{ \hat{\rho}_{gc} \hat{N}_i \} \\ &= \frac{1}{\beta} \frac{\partial}{\partial \mu_i} \ln Z_{gc}. \end{aligned} \quad (3.19)$$

Let us recall that all average values and traces are calculated by taking into account all microstates with arbitrary energy and particle number within a fixed volume V . For simplicity, we focus on one-component systems in the following.

It is easy to express the grand canonical partition function Z_{gc} in terms of the canonical one

$$Z_c(\beta, N) = \text{Tr} \{ e^{-\beta \hat{H}} \}_N,$$

which corresponds to a fixed number of particles N . Splitting the grand canonical trace in separate sums over all states having the same N , we have

$$Z_{gc}(\beta, \mu) = \sum_{N=0}^{+\infty} e^{\beta \mu N} \text{Tr} \{ e^{-\beta \hat{H}} \}_N = \sum_{N=0}^{+\infty} e^{\beta \mu N} Z_c(\beta, N). \quad (3.20)$$

Some authors, such as K. Huang, introduce the fugacity $z = e^{\beta \mu}$, which allows us to express the grand canonical partition function

$$Z_{gc} = \sum_{N=0}^{\infty} z^N Z_c(N)$$

as a power series of z . Although quite practical in some cases, we shall not follow this notation, since it spoils the correspondence between λ_i and f_i .

The probability of finding an energy E and a number of particles N is given by

$$\begin{aligned}
w(E, N) &= \langle \delta(E - \hat{H}) \delta(N - \hat{N}) \rangle \\
&= \frac{1}{Z_{gc}} \text{Tr} \{ \delta(E - \hat{H}) \delta(N - \hat{N}) \} e^{-\beta(E - \mu N)} \\
&= \frac{1}{Z_{gc}} e^{-\beta(E - \mu N)} \Omega_N(E).
\end{aligned} \tag{3.21}$$

where $\Omega_N(E)$ is the usual microcanonic density of states per unit energy when the system contains precisely N particles in the volume V . Notice the analogy with the energy distribution $w(E)$ in the canonical ensemble. In the present case we see that the probability density $w(E, N)$ is equal to the probability $e^{-\beta(E - \mu N)} / Z_{gc}$ of finding the system in a microstate with energy E and N particles multiplied by the number $\Omega_N(E)$ of states with N particles at the energy E per unit energy. In the macroscopic limit, $w(E, N)$ becomes an extremely narrow probability distribution around the average values. Since the central limit theorem applies (E and N are additive properties) the relative fluctuations of E and N become irrelevant for $N \rightarrow \infty$. Therefore, the average values $\langle \hat{H} \rangle$ and $\langle \hat{N} \rangle$ coincide with the most probable ones \bar{E} and \bar{N} .

The maximum of $w(E, N)$ is achieved when its gradient vanishes. From Eq. (3.21) one obtains

$$\left. \frac{\partial w}{\partial E} \right|_N = \frac{1}{Z_{gc}} \left(-\beta \Omega_N + \frac{\partial \Omega_N}{\partial E} \right) e^{-\beta(E - \mu N)}$$

and

$$\left. \frac{\partial w}{\partial N} \right|_E = \frac{1}{Z_{gc}} \left(\beta \mu \Omega_N + \frac{\partial \Omega_N}{\partial \mu} \right) e^{-\beta(E - \mu N)}.$$

Setting the partial derivatives equal to zero yields the important equilibrium relations

$$\beta = \left. \frac{\partial \ln \Omega}{\partial E} \right|_{N, V} \tag{3.22}$$

and

$$\mu = - \frac{1}{\beta} \left. \frac{\partial \ln \Omega}{\partial N} \right|_{E, V}. \tag{3.23}$$

These equations allow us to determine β and μ directly for any given energy and number of particles. In this way, inverting Eqs. (3.18) and (3.19), which give $\langle \hat{H} \rangle$ and $\langle \hat{N} \rangle$ as a function of β and μ , is avoided. Notice that μ is not necessarily negative, as one might be tempted to conclude from Eq. (3.23), since the average energy E , whose dependence on N is not obvious *a priori*, must be kept fixed upon differentiation with respect to N . Adding particles under the constraint of constant total energy does not necessarily

imply that the number of accessible states $\Omega_N(E)$ increases [i.e., $(\partial \ln \Omega / \partial N)|_E$ is not necessarily positive].⁵

The unconstrained probability distribution of the number of particles is given by

$$\begin{aligned} w(N) &= \int w(E, N) dE = \frac{1}{Z_{gc}} e^{\beta \mu N} \int e^{-\beta E} \Omega_N(E) dE \\ &= \frac{Z_c(\beta, N, V)}{Z_{gc}(\beta, \mu, V)} e^{\beta \mu N}, \end{aligned}$$

where we have used Eq. (3.17) for the canonical partition function Z_c . Concerning the dependence of $w(N)$ on N it is important to recall that Z_c also depends on N . In fact, μ and $\partial \ln(Z_c) / \partial N|_{\beta, V}$ always have opposite signs, even if they depend on the physical problem. In fact, one can show that $\mu = -(1/\beta)(\partial \ln Z_c / \partial N)|_{\beta, V}$. Thus, the product of Z_c and the exponential factor yields, as expected, a narrow distribution of N around its average value, which in particular ensures the integrability of $w(N) \rightarrow 0$ for $N \rightarrow \infty$. Furthermore, the convergence of the sum of all $w(N)$, namely, $\sum_N w(N) = 1$, is guaranteed by Eq. (3.20).

Exercise 3.12: Show that

$$\frac{\partial \langle \hat{N} \rangle}{\partial \mu} = \beta (\langle \hat{N}^2 \rangle - \langle \hat{N} \rangle^2).$$

Show that $\langle \hat{N} \rangle = 0$ for $\mu \rightarrow -\infty$ ($\beta > 0$). Explain why for any $\beta > 0$, there is always one and only one solution of the equation $\langle \hat{N} \rangle = N$ for any average number of particles $N > 0$.

The bijective relation between (E, N) and (β, μ)

Finally, before closing this section, we would like to show that in the mapping $(\beta, \mu) \rightarrow (E, N)$ with $N = \langle \hat{N} \rangle$ and $E = \langle \hat{H} \rangle$ is locally invertible for all β and μ . This implies that the solution of the equation $E(\beta, \mu) = E$ and $N(\beta, \mu) = N$ is always unique.

The proof is simpler if we return to the notation used in the general formulation, instead of using the variables β and μ and the operators \hat{H} and \hat{N} , which would otherwise certainly be the more physical choice. Thus, we consider the parameters $\lambda_1 = -\beta$ and $\lambda_2 = \beta\mu$ with corresponding operators $\hat{F}_1 = \hat{H}$ and $\hat{F}_2 = \hat{N}$. In these terms the grand canonical partition function is given by

$$\hat{\rho}_{gc} = \frac{1}{Z_{gc}} e^{\lambda_1 \hat{F}_1 + \lambda_2 \hat{F}_2}$$

⁵As a bus carrying N identical stormtrooper clones fills up, the number of different ways Ω_N in which the seats are occupied by keeping the energy constant (i.e., avoiding that two clones sit on top of each other) surely increases at first when the bus is nearly empty but then decreases when more than half of the seats are already occupied. Fermions show a similar behavior, while non-interacting bosons do not.

with

$$Z_{gc} = \text{Tr}\{e^{\lambda_1 \hat{F}_1 + \lambda_2 \hat{F}_2}\}.$$

It follows that

$$\frac{\partial \ln Z_{gc}}{\partial \lambda_i} = \langle F_i \rangle$$

and

$$\frac{\partial \langle F_i \rangle}{\partial \lambda_i} = \langle \hat{F}_i^2 \rangle - \langle \hat{F}_i \rangle^2$$

for $i = 1, 2$. In addition,

$$\frac{\partial \langle \hat{F}_1 \rangle}{\partial \lambda_2} = \langle \hat{F}_1 \hat{F}_2 \rangle - \langle \hat{F}_1 \rangle \langle \hat{F}_2 \rangle = \frac{\partial \langle \hat{F}_2 \rangle}{\partial \lambda_1}.$$

The Jacobian matrix of the transformation is then given by

$$J = \left(\frac{\partial \langle F_i \rangle}{\partial \lambda_j} \right) = \begin{bmatrix} \langle F_1^2 \rangle - \langle F_1 \rangle^2 & \langle F_1 F_2 \rangle - \langle F_1 \rangle \langle F_2 \rangle \\ \langle F_1 F_2 \rangle - \langle F_1 \rangle \langle F_2 \rangle & \langle F_2^2 \rangle - \langle F_2 \rangle^2 \end{bmatrix},$$

where the off-diagonal elements can be written as

$$\langle \hat{F}_1 \hat{F}_2 \rangle - \langle \hat{F}_1 \rangle \langle \hat{F}_2 \rangle = \left\langle (\hat{F}_1 - \langle \hat{F}_1 \rangle)(\hat{F}_2 - \langle \hat{F}_2 \rangle) \right\rangle.$$

We need to show that the determinant of the Jacobian is never zero, whatever density matrix $\hat{\rho}$ is used for performing the averages. To this aim, we define an appropriate inner product in an *ad hoc* vector space and apply the Cauchy-Schwarz inequality.

We consider s compatible linear operators $\hat{F}_1, \dots, \hat{F}_s$ with $[\hat{F}_i, \hat{F}_j] = 0 \forall i, j$. We define the vector space \mathbb{V} of all the linear combinations of the zero-average operators $\hat{f}_i = \hat{F}_i - \langle \hat{F}_i \rangle$. In this space we introduce the bilinear function

$$\langle \hat{f} | \hat{g} \rangle = \langle \hat{f} \hat{g} \rangle,$$

where the averages are taken with respect to some fixed density matrix $\hat{\rho}$. $\langle f | g \rangle$ has all the properties of an inner product. First, $\langle \hat{f} | \hat{f} \rangle \geq 0 \forall f$ and $\langle \hat{f} | \hat{f} \rangle = 0 \Rightarrow \hat{f} = 0$. Here we have used that $\langle \hat{f}^2 \rangle = 0$ implies that \hat{f} is constant and thus zero. Second, $\langle f | g \rangle$ is bilinear and symmetric, since $[\hat{F}_i, \hat{F}_j] = 0$. Consequently, we can apply the Cauchy-Schwarz inequality

$$\langle f | g \rangle^2 \leq \langle f | f \rangle \langle g | g \rangle,$$

which implies

$$0 \leq (\langle F_1^2 \rangle - \langle F_1 \rangle^2) (\langle F_2^2 \rangle - \langle F_2 \rangle^2) - (\langle F_1 F_2 \rangle - \langle F_1 \rangle \langle F_2 \rangle)^2 = \det(J).$$

The equal sign holds only when the vectors (operators) are linearly dependent, which of course does not hold for \hat{H} and \hat{N} . Since the Jacobian has always the same sign (and is never zero) the change of variables is invertible, which implies the uniqueness of the solution of the equations

$$E(\beta, \mu) = \langle \hat{H} \rangle_{\beta\mu} = E$$

and

$$N(\beta, \mu) = \langle \hat{N} \rangle_{\beta\mu} = N.$$

Exercise 3.13: Show that

$$\frac{\partial \ln Z_{gc}}{\partial \beta} = - \left(\langle \hat{H} \rangle - \mu \langle \hat{N} \rangle \right)$$

and

$$\frac{\partial}{\partial \beta} \left(\langle \hat{H} \rangle - \mu \langle \hat{N} \rangle \right) = - \left[\langle (\hat{H} - \mu \hat{N})^2 \rangle - \langle \hat{H} - \mu \hat{N} \rangle^2 \right] < 0.$$

Since $\langle \hat{H} \rangle - \mu \langle \hat{N} \rangle$ is a monotonically decreasing function of β for all μ , conclude that the equation $\langle \hat{H} \rangle - \mu \langle \hat{N} \rangle = \text{constant}$ always has a unique solution as a function of β . Perform the corresponding analysis concerning the chemical potential μ .

3.5.4 Grand Canonical pressure ensemble

A system which is open with respect to all constants of motion is characterized by the density operator

$$\hat{\rho}_{gcp} = e^{-\beta(\hat{H} - \mu\hat{N} + pV)},$$

where the partition function $\text{Tr}\{\rho_{gcp}\} = 1$. Thus, there is no need for a normalizing factor. The relevant parameters or variables entering ρ_{gcp} are here β , μ and p , which define the average energy $E = \langle \hat{H} \rangle$, average number of particles $\langle \hat{N} \rangle$ and average volume $\langle \hat{V} \rangle$, except for a common multiplicative constant. In fact, β , μ and p cannot define the additive properties E , $\langle \hat{N} \rangle$ and $\langle \hat{V} \rangle$. They are bonded by the condition $\text{Tr}\{\rho_{gcp}\} = 1$, which represents the equation of state of the material under consideration, and are therefore not independent of each other. The variables β , μ and p can only define the ratios or densities of the additive constants of motion, for example, $E/\langle \hat{N} \rangle$ and $\langle \hat{V} \rangle/\langle \hat{N} \rangle$. We shall see that β , μ and p are intensive properties which remain unchanged when all additive properties E , N and V are multiplied or scaled by the same factor [e.g., $p(E, N, V) = p(\alpha E, \alpha N, \alpha V) \quad \forall \alpha > 0$ and similarly for β and μ]. Knowing β , μ and p provides no information on the actual size of the system. Under these circumstances, volume fluctuations come into play. We find this physical situation when we consider a hot air balloon, an air bubble in a liquid, or a warm-air bubble climbing its way up in the atmosphere on a sunny day. This ensemble is known as *grand canonical pressure ensemble*.

The probability density of finding an energy E , number of particles N and volume V is

$$w(E, N, V) = \langle \delta(E - \hat{H}) \delta(N - \hat{N}) \delta(V - \hat{V}) \rangle = e^{-\beta(E - \mu N + pV)} \Omega_{N,V}(E),$$

where $\Omega_{N,V}(E)$ is the density of states for a system containing N particles in a volume V at the energy E . Notice that $\Omega_{N,V}(E)$ represents a density of states per unit energy and unit volume. As the product of delta functions, it has units of inverse energy times

inverse volume. In the same way as the previously considered ensembles, statistical independence implies that $\Omega_{N,V}(E)$ is an exponentially increasing function of E , N and V , whereas the Boltzmann factor decreases exponentially as a function of these variables ($\beta > 0$, $\mu < 0$ and $p > 0$). Therefore, $w(E, N, V)$ has an extremely narrow distribution around the averages $\langle E \rangle$, $\langle N \rangle$ and $\langle V \rangle$, which coincide with the values of E , N and V that maximize $w(E, N, V)$.

3.6 Implicit notation

Throughout the previous discussion of the density operators and partition functions in the different ensembles we made use of a widespread convention, which consists in dropping the δ functions in the expressions for $\hat{\rho}_c$ and Z_r . This is possible, and should lead to no confusion, since we implicitly assumed that the Hilbert space on which $\hat{\rho}_c$ acts, and with respect to which the trace in Z_r is calculated, contains only states $|\psi\rangle$ having well-defined values of the constants of motion:

$$\hat{F}_j|\psi\rangle = f_j|\psi\rangle$$

for $j = r + 1, \dots, s$. Keeping this in mind is, of course, crucial for the calculations. Using this convention, the general expressions for the canonical ensemble take the form

$$\hat{\rho}_c = \frac{1}{Z_r} e^{\sum_{i=1}^r \lambda_i \hat{F}_i},$$

where

$$Z_r = \text{Tr}\{e^{\sum_{i=1}^r \lambda_i \hat{F}_i}\},$$

which is clearly more compact and elegant.

4 Entropy

The entropy is one of the most important concepts in statistical mechanics. It is defined as

$$S = -k_B \langle \ln \hat{\rho} \rangle = -k_B \text{Tr}\{\hat{\rho} \ln \hat{\rho}\},$$

where $-\langle \ln \hat{\rho} \rangle$ is the entropy of the probability distribution of the microstates in the mixed state defined by $\hat{\rho}$ and $k_B = 1,38 \times 10^{-23}$ J/K is the Boltzmann constant. As already discussed in Sec. 1.8, the average $-\langle \ln \hat{\rho} \rangle = S/k_B$ gives a measure of the diversity of the population of microstates or, equivalently, of the lack of information in the corresponding probability distribution. The Boltzmann constant is introduced in order to be consistent with the historical definition of entropy change, which is the energy difference at constant volume and particle number (i.e., reversible heat) between two equilibrium states divided by the absolute temperature T at which the heat transfer occurs. Thus, k_B and S have units of energy divided by temperature. The temperature unit K stands for *Kelvin* after William Thomson, 1st Baron Kelvin (1824–1907) who measured the value of the absolute zero of temperature ($\simeq -273^\circ\text{C}$). A temperature difference of 1 K is the same as 1°C , only the reference zero is different. In the Kelvin or absolute temperature scale, zero corresponds to the lowest attainable temperature value, whose existence was already known to Carnot (1824). The Boltzmann constant is not a fundamental physical constant. Its value follows from the convention that assigns the temperature $T = 273.16$ K to the triple point of water. The choice of the triple point is practical since the temperature at which the three phases coexist (solid, liquid, gas) can be precisely and easily measured. The choice of the temperature value at the triple point ensures consistency with the Celsius scale (0°C corresponds to 273.15 K). Thus, maybe some day one might be able to compute k_B in terms of \hbar , c and the mass and charge of the elementary particles. k_B allows us to relate temperature and energy scales. Roughly speaking, for simple estimates, one may take that $1 \text{ eV} \approx 10^4 \text{ K}$, actually 8620 K, or $10 \text{ K} \approx 1 \text{ meV}$. As we shall see, these rough approximations of k_B are very useful in order to judge which energy differences are attainable or important at a given temperature. In the following, unless explicitly stated otherwise, the temperature always refers to the absolute scale.

If we denote by ρ_ν the eigenvalues of $\hat{\rho}$ we have

$$S = -k_B \sum_{\nu} \rho_{\nu} \ln \rho_{\nu} \geq 0,$$

since $x \ln x \leq 0$ for $0 \leq x \leq 1$. We already know that $S = 0$ only for pure states. S represents the average of the additive constant of motion $-\ln \hat{\rho}$, which we inferred under the assumption of statistical independence of macroscopic subsystems. In previous chapters we have analyzed statistical ensembles (mixed states) corresponding to different physical situations: perfect isolation, exchange of some constants of motion and full openness. The corresponding density operators $\hat{\rho}$ characterizing the various mixed states are obviously different and the differences are significant qualitatively. For instance, in the microcanonical ensemble the total energy is sharply defined, whereas in the canonical

ensemble the total energy fluctuations grow as \sqrt{N} . Nevertheless, we have shown that all the different ensembles yield equivalent descriptions of the equilibrium state and its macroscopic properties, since the relative fluctuations of the constant of motions (e.g., E , N and V) are irrelevant in the macroscopic limit ($N \rightarrow \infty$). One of our first goals in this chapter is to show that calculating the entropy in the various ensembles, according to the corresponding ρ 's, yields the same result, provided that the parameters defining the macroscopic state are the same.

Another question of central interest is the behavior of S upon scaling. For open systems, in the grand canonical ensemble, we know that $\ln \hat{\rho}$ is an additive constant of motion, which can be expressed as

$$\ln \hat{\rho} = \sum_{i=1}^s \lambda_i \hat{F}_i,$$

in terms of the set $\hat{F}_1, \dots, \hat{F}_s$ of linearly independent additive constants of motion. In this case, the additivity of S follows immediately:

$$S = -k_B \langle \ln \hat{\rho} \rangle = -k_B \sum_{i=1}^s \lambda_i \langle \hat{F}_i \rangle = -k_B \sum_{i=1}^s \lambda_i f_i.$$

If the system is not fully open, i.e., if some additive constants of motion are constrained to have well-defined values (i.e., no exchange with the environment), the statistical independence of the subsystems does not strictly hold. Indeed, different subsystems of an isolated system can be regarded as statistically independent, except for the fact that the sum of the conserved additive properties over all subsystems must always take a well-defined value. Therefore, one could doubt on the additivity of S *a priori*. However, we have seen that for macroscopic systems such constraints are not essential, since the density operators of different microcanonical and grand canonical ensembles are equivalent in practice: the fluctuations of the properties F_i ($i = 1, \dots, r$) with respect to which the system is open, are negligible with respect to the average values $\langle F_i \rangle$. For instance, the energy distribution $w(E)$ in the canonical ensemble is extremely sharp, even though E is not strictly conserved. As we shall see, the additivity of S holds in general.

Let us consider the general canonical ensemble, which is open with respect to $\hat{F}_1, \dots, \hat{F}_r$, and closed with respect to $\hat{F}_{r+1}, \dots, \hat{F}_s$. In this case we have

$$\hat{\rho}_r = e^{\sum_{i=1}^s \lambda_i \hat{F}_i} \prod_{j=r+1}^s \delta(\hat{F}_j - f_j), \quad (4.1)$$

which can be written as

$$\hat{\rho}_r = \frac{1}{Z_r} e^{\sum_{i=1}^r \lambda_i \hat{F}_i}, \quad (4.2)$$

with the partition function

$$Z_r = e^{-\sum_{j=r+1}^s \lambda_j f_j} = \text{Tr}\{e^{\sum_{i=1}^r \lambda_i \hat{F}_i}\}. \quad (4.3)$$

The entropy of the mixed state $\hat{\rho}_r$ is then given by

$$S = -k_B \langle \ln \hat{\rho} \rangle = -k_B \left[\sum_{i=1}^r \lambda_i \langle \hat{F}_i \rangle - \ln Z_r \right] = -k_B \sum_{i=1}^s \lambda_i f_i, \quad (4.4)$$

where we have used that $\langle \hat{F}_i \rangle = f_i$ for $1 \leq i \leq r$. The last equality shows that S is equal to a linear combination of all the additive properties f_1, \dots, f_s . Since the λ_i should be intensive properties—in order to ensure the additivity of $\ln \hat{\rho}$ — S is expected to be an additive property for all r , including the microcanonic case ($r = 0$) and the already mentioned grandcanonic case ($r = s$).

Notice that for each r or type of ensemble, S can be always be regarded as a function of the r coefficient $\lambda_1, \dots, \lambda_r$ —which control the average values $\langle \hat{F}_1 \rangle = f_1, \dots, \langle \hat{F}_r \rangle = f_r$ of the constants of motion exchanged with the environment—and of the additive properties f_{r+1}, \dots, f_s —which are the same in all microstates. This follows clearly from the fact that the density operator $\hat{\rho}_r$, from which S is issued, is defined by $\lambda_1, \dots, \lambda_r$ and f_{r+1}, \dots, f_s . For example, in the canonical ensemble, $\hat{\rho}_c$ and S depend on β , N and V . However, later on we will see that it is actually meaningful to regard S as a function of all additive constants of motion f_1, \dots, f_s . For instance, in the most usual case we have $S = S(E, N, V)$. In fact, the additivity of S will be rigorously demonstrated by showing that S is an homogeneous function of grade 1 of these variables.

4.1 Maximum entropy theorem

We would like to establish the fundamental variational principle of statistical mechanics, in terms of which the state of equilibrium can be characterized. To this aim we consider the entropy

$$S[\hat{\rho}] = -k_B \langle \ln \hat{\rho} \rangle = -k_B \text{Tr}\{\hat{\rho} \ln \hat{\rho}\} \quad (4.5)$$

as a functional of the density operator $\hat{\rho}$ defining an arbitrary mixed state in the Hilbert space of our system. The properties of $\hat{\rho}$ are clear: $\hat{\rho}^\dagger = \hat{\rho}$, its eigenvalues ρ_ν satisfy $0 \leq \rho_\nu \leq 1$, and $\text{Tr}\{\hat{\rho}\} = \sum_\nu \rho_\nu = 1$. All $\hat{\rho}$'s are positive-semidefinite, trace-one, hermitic operators. We consider a general canonical ensemble. The underlying Hilbert space is spanned by the states having well-defined constants of motion $\hat{F}_{r+1}, \dots, \hat{F}_s$.

Theorem: The following fundamental property holds. Consider a system which is closed with respect to the constants of motion $\hat{F}_{r+1}, \dots, \hat{F}_s$ which have well-defined values f_{r+1}, \dots, f_s . Among all the the density operators $\hat{\rho}$ satisfying $\langle \hat{F}_i \rangle = \text{Tr}\{\hat{\rho} \hat{F}_i\} = f_i$ for $i = 1, \dots, r$ the one yielding the largest S is the canonical density operator $\hat{\rho}_r$ given by Eqs. (4.2) and (4.3), which describes the state of equilibrium. This is equivalent to the inequality

$$S[\hat{\rho}] \leq S[\hat{\rho}_r]$$

or, using Eqs. (4.4) and (4.5), to

$$S[\hat{\rho}] = -k_B \text{Tr}\{\hat{\rho} \ln \hat{\rho}\} \leq S[\hat{\rho}_r] = -k_B \left[\sum_{i=1}^r \lambda_i f_i - \ln Z_r \right].$$

Thus, the state of equilibrium is the one which maximizes S under the constraints $\langle \hat{F}_i \rangle = f_i$ for $i = 1, \dots, r$, and $\hat{F}_j = f_j$ for $j = r + 1, \dots, s$.

Lemma: In order to prove this important statement we first show the following lemma. Given any two density operators $\hat{\rho}$ and $\hat{\rho}'$, we have

$$\text{Tr}\{\hat{\rho} \ln \hat{\rho}'\} \leq \text{Tr}\{\hat{\rho} \ln \hat{\rho}\}, \quad (4.6)$$

The starting point is the inequality $\ln x \leq x - 1$ for any $x > 0$. This implies that for any hermitic operator \hat{A} with eigenvectors $|n\rangle$ and eigenvalues $a_n \geq 0$, we have

$$\langle \psi | \ln \hat{A} | \psi \rangle = \sum_n |\langle \psi | n \rangle|^2 \ln a_n \leq \sum_n |\langle \psi | n \rangle|^2 (a_n - 1) = \langle \psi | (\hat{A} - 1) | \psi \rangle. \quad (4.7)$$

We consider the diagonal representation of

$$\hat{\rho} = \sum_n \rho_n |n\rangle \langle n|$$

with $\rho_n > 0$ and obtain

$$\text{Tr}\{\hat{\rho} [\ln(\hat{\rho}') - \ln(\hat{\rho})]\} = \sum_n \rho_n \langle n | [\ln(\hat{\rho}') - \ln(\rho_n)] | n \rangle = \sum_n \rho_n \langle n | \ln \left(\frac{\hat{\rho}'}{\rho_n} \right) | n \rangle,$$

where we have used that $\rho_n > 0$. The operators $\hat{\rho}'/\rho_n$ are all positive semi-definite, since $\hat{\rho}'$ has positive or zero eigenvalues and $\rho_n > 0$. Therefore, using (4.7) we have

$$\langle n | \ln \left(\frac{\hat{\rho}'}{\rho_n} \right) | n \rangle \leq \langle n | \left(\frac{\hat{\rho}'}{\rho_n} - 1 \right) | n \rangle$$

for all n . It follows that

$$\text{Tr}\{\hat{\rho} (\ln \hat{\rho}' - \ln \hat{\rho})\} \leq \sum_n \rho_n \langle n | \left(\frac{\hat{\rho}'}{\rho_n} - 1 \right) | n \rangle = \sum_n \langle n | \hat{\rho}' | n \rangle - \text{Tr}\{\hat{\rho}\} = \text{Tr}\{\hat{\rho}'\} - 1 = 0.$$

This proves the inequality (4.6).

Proof: We turn now our attention to the entropy $S[\hat{\rho}]$ of an arbitrary density operator $\hat{\rho}$. Using the inequality (4.6) with $\hat{\rho}$ arbitrary and $\hat{\rho}' = \hat{\rho}_r$ we have

$$S[\hat{\rho}] = -k_B \text{Tr}\{\hat{\rho} \ln \hat{\rho}\} \leq -k_B \text{Tr}\{\hat{\rho} \ln \hat{\rho}_r\} = -k_B \text{Tr} \left\{ \hat{\rho} \left(\sum_{i=1}^r \lambda_i \hat{F}_i - \ln Z_r \right) \right\},$$

where we have replaced, on the right hand side, the explicit form of the canonic density operator $\ln \hat{\rho}_r = \sum_{i=1}^r \lambda_i \hat{F}_i - \ln Z_r$ given by Eq. (4.2). Since all the mixed states $\hat{\rho}$ have the same averages $\langle \hat{F}_i \rangle = \text{Tr}\{\hat{\rho} \hat{F}_i\} = f_i$ for $i = 1, \dots, r$, we finally obtain

$$S[\hat{\rho}] \leq -k_B \sum_{i=1}^s \lambda_i f_i = S[\hat{\rho}_r], \quad (4.8)$$

which proves the maximum entropy theorem. Notice that the logarithm of the density operator at equilibrium $\hat{\rho}_r$ is a linear combination of additive constants of motion. Therefore, the average $\langle \ln \hat{\rho}_r \rangle = \text{Tr}\{\hat{\rho} \ln \hat{\rho}_r\}$ is the same for all the density operators $\hat{\rho}$ having the same average values $\langle \hat{F}_i \rangle = f_i$ of the properties \hat{F}_i with respect to which the system is open ($1 \leq i \leq r$) and the same constants of motion f_j with respect to which the system is closed ($r+1 \leq j \leq s$).

We have already discussed the meaning of $-\langle \ln \hat{\rho} \rangle = -\sum_n \rho_n \ln \rho_n$ for an arbitrary probability distribution ρ_n , where $0 \leq \rho_n \leq 1$ and $\sum_n \rho_n = 1$. We have seen that $-\langle \ln \hat{\rho} \rangle$ gives a measure of the degree of disorder in the distribution, or equivalently, a measure of the deviation of the mixed states described by $\hat{\rho}$ from a pure state ($S = 0$ only for pure states). Moreover, S always increases as the number of microstates participating in the mixed states increases since a larger number of participating states reduces the average of ρ_n and $\ln \rho_n$ ($\sum_n \rho_n = 1$). We have also interpreted $-\langle \ln \hat{\rho} \rangle$ as a measure of the degree of uncertainty (or lack of knowledge) that we have on the mixed state of the system. With this in mind, the principle of maximum entropy tells us that among all the states of a system satisfying $\langle F_i \rangle = f_i$ for $i = 1, \dots, r$ and $\hat{F}_i \equiv f_i$ for $i = r+1, \dots, s$, the equilibrium state is the one with the highest degree of disorder, the most far from pure, and the one about which we know less. This fundamental characteristic of the equilibrium state is most plausible from a physical perspective. Indeed, one can actually grasp the state of equilibrium in this way.

4.2 The approach to equilibrium

An alternative perspective to the principle of maximum entropy can be achieved by investigating the time dependence of the entropy $S = -k_B \langle \ln \hat{\rho} \rangle$ of a system in an arbitrary mixed state $\hat{\rho}$ along the spontaneous process of reaching equilibrium. We already know that the interactions with the environment play a central role in reaching thermodynamic equilibrium. In fact, we have seen that when the system is strictly isolated, the total time derivative of $\hat{\rho}$ vanishes and the entropy, given by its eigenvalues, is independent of time. The main reason behind this remarkable time independence—besides the fundamental invariance of inner products in quantum mechanical time evolution—is that the probabilities P_n of all the microstates $|n\rangle$ (also referred to as weights w_n) are independent of time, as long as the system does not interact with the environment. In this section we would like to discuss how P_n depends on time when the system evolves towards equilibrium and how this conditions the time evolution of the entropy in spontaneous processes. To this aim we will treat the interactions with the

environment as stochastic processes, which are uncorrelated in time. This assumption will lead us to the *master equation*. Subsequently, with $P_n(t)$ in hand, the fundamental symmetry of quantum mechanical transition probabilities will be exploited in order to derive the time dependence of the entropy. This will provide us with a complementary dynamical perspective to the maximum entropy theorem.⁶

4.2.1 The master equation

We consider a system in an arbitrary mixed state, not necessarily in equilibrium, which is described by the density operator

$$\hat{\rho} = \sum_m |m\rangle P_m \langle m|.$$

The states $|m\rangle$ are assumed here to be the eigenstates of $\hat{\rho}$ with $\langle m|n\rangle = \delta_{mn}$ and $\sum_m P_m = 1$. The interactions with the environment are described by some external, yet unspecified perturbation \hat{V} . Since the interactions are uncontrolled, the changes of state of the system cannot be predicted with certainty. We shall describe them as a stochastic process, which is defined as a time-indexed or time-ordered succession of states

$$m_1 t_1, m_2 t_2, \dots, m_{n-1} t_{n-1}$$

at times $t_1 < t_2 < \dots < t_{n-1}$. Moreover, we assume that the transitions are uncorrelated in time, i.e., the probability for the system to be in the state m_k at time t_k is completely determined by the state m_{k-1} which the system had at time t_{k-1} . Such processes are called *Markovian* and the succession of states m_1, m_2, \dots, m_{n-1} that the system undergoes is usually referred to as a *Markov chain*. The assumed lack of any memory or inertial effects implies that the time evolution is entirely controlled by the transition probability

$$p(m, t' | n, t), \quad (4.9)$$

which represents the conditional probability of finding the system in the state m at time t' , knowing that it was in state n at time t . Clearly, when $t' = t$ we have

$$p(m, t | n, t) = \delta_{mn} \quad \forall t. \quad (4.10)$$

Moreover, the normalization condition requires

$$\sum_m p(m, t' | n, t) = 1 \quad \forall t' > t, \quad (4.11)$$

irrespectively of the state n in which the system is at time t . For small time differences $\Delta t = t' - t$, it is meaningful to perform a linear expansion and to introduce the transition rate or rate constant k_{mn} representing the transition probability per unit time from n to m :

$$p(m, t + \Delta t | n, t) = k_{mn} \Delta t + O(\Delta t^2), \quad (4.12)$$

⁶The main ideas behind this section are borrowed from Feynman's book [3].

for all $m \neq n$. The probability of staying in the same state m at $t + \Delta t$ is obtained from the normalization condition (4.11):

$$p(m, t + \Delta t | m, t) = 1 - \sum_{n \neq m} k_{nm} \Delta t + O(\Delta t^2). \quad (4.13)$$

Notice that the linear expansions (4.12) and (4.13) of the conditional probabilities satisfy Eqs. (4.10) and (4.11) to first order in Δt . The dynamics of the probability distribution $\{P_m\}$, in a Markov processes is entirely controlled by the transition rates k_{mn} .

The probability of finding the system in the state m at time $t + \Delta t$ is equal to the sum of the probabilities of being in any state n (including $n = m$) at time t multiplied by the transition probability $p(m, t + \Delta t | n, t)$ from the state n to the state m . This is written as

$$P_m(t + \Delta t) = \sum_n p(m, t + \Delta t | n, t) P_n(t).$$

Replacing (4.12) for $m \neq n$ and (4.13) for $m = n$, we have

$$\begin{aligned} P_m(t + \Delta t) &= \sum_{n \neq m} k_{mn} \Delta t P_n(t) + \left(1 - \sum_{n \neq m} k_{nm} \Delta t \right) P_m(t) \\ &= P_m(t) \Delta t + \sum_{n \neq m} k_{mn} P_n(t) \Delta t - \sum_{n \neq m} k_{nm} P_m(t) \Delta t, \end{aligned}$$

which implies

$$\frac{dP_m}{dt} = \sum_{n \neq m} (k_{mn} P_n - k_{nm} P_m). \quad (4.14)$$

The first term on the right-hand side, which is positive and thus increases P_m , represents the probability of ending up in the state m coming from any other state $n \neq m$, while the second negative terms correspond to the probability of making a transition to any other state n , starting from m . This simple, physically transparent relation plays a central role in stochastic dynamics. It is known as the *master equation*.

4.2.2 Time dependence of the entropy

In order to be able to determine the time dependence of P_m and of S we need some information on the transition rates k_{mn} . To this aim we assume in the following that the system is closed with respect to all its additive conserved quantities, i.e., that no energy, particles or volume are exchanged as a result of the interaction with the environment. This corresponds to the microcanonical ensemble. The transition rates k_{mn} are derived from time-dependent perturbation theory in quantum mechanics.⁷ Given a constant

⁷See, for example, J.J. Sakurai, *Modern Quantum Mechanics*, (Addison-Wesley, Reading, 1994) p. 327 ff.

perturbation \hat{V} , with matrix elements $V_{nm} = \langle n|\hat{V}|m\rangle$, the transition probability per unit time from an unperturbed state $|m\rangle$ to a state $|n\rangle$ is given by the Fermi golden rule

$$k_{nm} = \frac{2\pi}{\hbar} |V_{nm}|^2 \delta(E_n - E_m), \quad (4.15)$$

Despite the fact that the details of the interaction \hat{V} between the system and its environment are not known in general, the most important information for us is the microscopic reversibility

$$k_{nm} = k_{mn}. \quad (4.16)$$

This symmetry always holds at the quantum level, between every pair of states $|m\rangle$ and $|n\rangle$, since \hat{V} is hermitic ($V_{nm} = V_{mn}^*$). It tells us that at a microscopic level the probability for a transition from state $|m\rangle$ to $|n\rangle$ is the same as the probability for a transition from $|n\rangle$ to $|m\rangle$. Further statistical considerations, concerning the number of accessible states of the environment when the system is in the final state $|m\rangle$, are not important at this stage since all microstates have the same energy ($E_n = E_m$).

We may finally turn our attention to the entropy

$$S = -k_B \langle \ln \hat{\rho} \rangle = -k_B \sum_m P_m \ln P_m$$

and calculate

$$\frac{dS}{dt} = -k_B \sum_m \left(\frac{dP_m}{dt} \ln P_m + \frac{dP_m}{dt} \right).$$

Taking into account that

$$\sum_m P_m = 1 \quad \Rightarrow \quad \sum_m \frac{dP_m}{dt} = 0,$$

we have

$$\begin{aligned} \frac{dS}{dt} &= -k_B \sum_m \frac{dP_m}{dt} \ln P_m \\ &= -k_B \sum_m \sum_{n \neq m} (k_{mn} P_n - k_{nm} P_m) \ln P_m \\ &= -k_B \sum_{m \neq n} k_{mn} (P_n - P_m) \ln P_m \\ &= -k_B \sum_{m < n} k_{mn} (P_n - P_m) (\ln P_m - \ln P_n). \end{aligned}$$

Since $\ln(x)$ is an increasing function of x the signs of $(P_n - P_m)$ and $(\ln P_m - \ln P_n)$ are always opposite. This implies

$$\frac{dS}{dt} \geq 0 \quad \forall t. \quad (4.17)$$

One concludes that the entropy of a closed system always increases along the process of reaching equilibrium, i.e., along any spontaneous process. The equilibrium state of a closed system can only be the one with the maximum entropy.

Notice that the previous derivation of the dynamical entropy increase is intimately related to the microscopic reversibility $k_{nm} = k_{mn}$. Intuitively, it is also clear that symmetric rates always lead to an increase of the diversity or disorder in the probability distribution $\{P_m\}$. Consider for example a pure state having $P_n = 1$ and $P_m = 0$ for all other m . Since $P_m = 0$ for $m \neq n$ and $P_n = 1$, the transition rate k_{mn} (from n to m) will lead to an increase of P_m . However, P_m is very small at the beginning. Therefore, the probability for the system to go back from m to n is very small, even though the conditional probability for a transition from m to n , i.e., the rate k_{nm} is the same as k_{mn} . Instead, the system will make transitions from m to all other states l , for which $k_{lm} \neq 0$, and so on, until the maximum diversity or disorder in the probability distribution $\{P_n\}$, i.e., maximum S is reached.

It is worth noting that the master equation also applies to the stochastic dynamics of open systems, for which $k_{nm} \neq k_{mn}$ and the equilibrium state does not correspond to the probability distribution giving the maximum entropy. Of course, $dP_m/dt = 0$ and $dS/dt = 0$ once equilibrium is reached. A simple transparent relation implying equilibrium is the *detailed balance condition* $k_{mn}P_n = k_{nm}P_m$ for all n and m , as one can easily verify by substituting in Eq. (4.14). Although this is not the only mathematical possibility of reaching $dP_m/dt = 0$ for all m , it can be shown to be the only one consistent with microscopic reversibility (Kolmogorov's criterion).

A particularly relevant, often encountered physical situation occurs when the system can exchange energy with a much larger environment which plays the role of a thermal bath and which, being very large, always has a well defined temperature T [$1/T = (\partial S/\partial E)_{N,V}$]. In this case one can apply Fermi's golden rule given by Eq. (4.15), and the resulting reversibility and energy conservation predicted by quantum mechanics, by regarding the system plus bath as a single closed unit. It follows that $k_{mn}/k_{nm} = e^{-\beta(\varepsilon_m - \varepsilon_n)}$, where ε_n and ε_m denote the energies of the system in the respective states and $\beta = 1/k_B T$. Clearly, the transition rates are not symmetric. For every pair of states, the downward transitions —towards reducing the energy of the system and thus increasing that of the bath— are more favorable. A reduction of the entropy of the system along the spontaneous dynamics cannot be excluded under these circumstances. This occurs, for example, along the process of folding proteins into their natural state, or simply when a hot system is plunged in a cold environment. Of course, the entropy of the proteins plus the surrounding liquid or tissue certainly increases along the process. We shall later on see that in these cases the maximum entropy principle is conveniently adapted to yield another variational principle, namely, the principle of minimal free energy $F = E - TS$, where E refers to the average energy and T to the absolute temperature. From a dynamical perspective the master equation implies $dF/dt \leq 0$ for all t and for all initial probability distributions. Notice that minimizing F is equivalent to the general maximum entropy theorem discussed in Sec. 4.1 when the average energy E is fixed.

4.2.3 A simple example of irreversible entropy growth

The principle of maximum entropy (either in its variational or dynamical form) allows us to predict and understand the evolution of a closed systems from a nonequilibrium situation towards the equilibrium state. For example, if two systems 1 and 2 are brought in thermal contact, i.e., if they are allowed to exchange heat (i.e., energy without volume or particle exchange) the energy will flow from the system having the lowest $\partial S/\partial E$ to the system having the largest $\partial S/\partial E$ since in this way $S = S_1 + S_2$ is increased. Incidentally, we conclude that $\partial S/\partial E$ provides a universal, device-independent means of defining *temperature*. As we shall see, $\partial S/\partial E = 1/T$, where T is the absolute temperature of the system. Thus, energy flows spontaneously only from the hotter to the cooler system, in agreement with our usual thermodynamic notion of temperature. Equilibrium is reached only when $\partial S/\partial E$ and thus the temperatures in both systems are equal.

4.3 Thermodynamic equivalent descriptions

We have learned that systems having a given set of additive constants of motion f_1, \dots, f_s are described by different density operators depending on the degree of isolation, i.e., on whether f_i is a truly conserved property with a well-defined value ($\hat{F}_i = f_i$) or whether f_i is actually the average value $\langle \hat{F}_i \rangle = f_i$ of a property, which allows exchanges and thus fluctuates. However, we have also seen that these descriptions are essentially equivalent, since the probability of observing a significant relative fluctuation tends to zero as $1/\sqrt{N}$ as the size of the system grows. We therefore conclude that the various mixed states or density operators $\hat{\rho}_r$ satisfying $\langle \hat{F}_i \rangle = f_i$ for $i = 1, \dots, r$ and $\hat{F}_i = f_i$ for $i = r + 1 \dots, s$ are thermodynamically equivalent descriptions of the same macroscopic state. More specifically, the microcanonical, canonical and grand canonical descriptions of a system with given E , N and V are equivalent, regardless of whether E is strictly conserved or just $E = \langle \hat{H} \rangle$, for example.

We have thus different possibilities of characterizing a macroscopic state, either by mean of the additive observables f_1, \dots, f_s , or by any of the sets of s variables

$$\lambda_1, \dots, \lambda_r, f_{r+1}, \dots, f_s,$$

provided that the λ_i yield $\langle \hat{F}_i \rangle = f_i$ for $i = 1, \dots, r$. The variables $\lambda_1, \dots, \lambda_r, f_{r+1}, \dots, f_s$ define

$$\hat{\rho}_r = \frac{1}{Z_r} e^{\sum_{i=1}^r \lambda_i \hat{F}_i}$$

and with that all observables. Knowing that only s quantities are needed to define the macrostate, it is clear that there must be relations allowing to determine s of them in terms of the others, for example, expressing $\lambda_1, \dots, \lambda_s$ in terms of f_1, \dots, f_s . Establishing these relations is important for at least two reasons. First, they are expected to be universal, i.e., valid for any system in any of its states. And second, they provide us with a means of assessing the physical meaning and properties of the parameters $\lambda_1, \dots, \lambda_s$.

One could in principle attempt to express λ_i as a functions of f_i by solving the coupled equations

$$f_i = \langle \hat{F}_i \rangle = \text{Tr}\{\hat{\rho}_r \hat{F}_i\}. \quad (4.18)$$

for $i = 1, \dots, r$. However, this pathway is problem specific and in most cases infeasible. It depends on the specific form of \hat{H} . Nevertheless, the entropy and the partition function Z_r provide us with a much simpler way to solve the problem.

From the definition of entropy we have

$$\begin{aligned} S &= -k_B \langle \ln \hat{\rho}_r \rangle = -k_B \left[\sum_{i=1}^r \lambda_i \langle \hat{F}_i \rangle - \ln Z_r \right] \\ &= -k_B \left[\sum_{i=1}^r \lambda_i f_i - \ln Z_r \right] \end{aligned} \quad (4.19)$$

This holds for all r including the microcanonical case ($r = 0$), where

$$S = k_B \ln Z_0 = k_B \ln \Omega,$$

and the grand canonical case ($r = s$) where

$$S = -k_B \sum_{i=1}^s \lambda_i \langle \hat{F}_i \rangle = -k_B \sum_{i=1}^s \lambda_i f_i.$$

It is easy to see that all these different ways to compute S , using different ensembles, yield the same result, provided that the variables defining the macrostate are the same. We know that the partition function is given by

$$Z_r = \text{Tr}\{e^{\sum_{i=1}^r \lambda_i \hat{F}_i} \prod_{j=r+1}^s \delta(f_j - \hat{F}_j)\} = e^{-\sum_{j=r+1}^s \lambda_j f_j}. \quad (4.20)$$

Substituting (4.20) in (4.19) we obtain

$$S = -k_B \left[\sum_{i=1}^r \lambda_i f_i + \sum_{j=r+1}^s \lambda_j f_j \right] = -k_B \sum_{i=1}^s \lambda_i f_i, \quad (4.21)$$

where we have used that $\langle \hat{F}_i \rangle = f_i$ for $1 \leq i \leq r$.

In order to characterize the entropy, or any partition function Z_r , we are in principle free to use any set of s independent variables. Nevertheless, it is meaningful to describe Z_r in terms of $\lambda_1, \dots, \lambda_r, f_{r+1}, \dots, f_s$, since these are the variables which define the corresponding physical situation or ensemble and which can be controlled directly in experiment. For instance, in the microcanonical ensemble the system is closed and the natural variables which have well defined values are E , N and V . In the canonical ensemble the energy can fluctuate and the natural variables having precise values are β

or T , N and V . Finally, in the grand canonical ensemble the variables that one controls are β , μ and V , energy and particle number fluctuate. A further important reason for regarding $\lambda_1, \dots, \lambda_r, f_{r+1}, \dots, f_s$ as the natural variables of Z_r is the very useful relation

$$\frac{\partial \ln Z_r}{\partial \lambda_i} = \langle \hat{F}_i \rangle = f_i, \quad (4.22)$$

valid for $i = 1, \dots, r$ in any canonical ensemble. Here it is implicitly understood that all other variables of Z_r , namely, λ_j for $j \neq i$ ($1 \leq j \leq r$) and f_{r+1}, \dots, f_s , are kept constant upon performing the partial derivation with respect to λ_i .

In the case of the entropy we have noticed that $S = k_B \ln Z_0$ is (apart from a multiplicative constant) equal to the logarithm of the microcanonical partition function Z_0 , which is entirely defined by the additive constants of motion f_1, \dots, f_s . Therefore, it is meaningful to regard

$$S = S(f_1, \dots, f_s)$$

as a function of the additive constants of motion. In the most usual explicit case we have

$$S = S(E, V, N) = k_B \ln[\Omega(E, N, V)].$$

We would like to understand the dependence of S on its fundamental variables f_1, \dots, f_s for all possible ensembles, from the microcanonical ($r = 0$) to the grand canonical one ($r = s$), by computing $\frac{\partial S}{\partial f_j}$ keeping all the other f_k fixed. In this way we would be able to obtain the entropy change associated with any change in the macrostate as

$$dS = \sum_i \frac{\partial S}{\partial f_i} df_i.$$

For example, one could write

$$dS = \frac{\partial S}{\partial E} dE + \frac{\partial S}{\partial V} dV + \frac{\partial S}{\partial N} dN.$$

In order to calculate the partial derivatives of $S = -k_B \langle \ln \hat{\rho}_r \rangle$ with respect to f_j (keeping all other f_k with $k \neq j$ fixed) we need to consider the cases $j \leq r$ and $j > r$ separately. First, starting from Eq. (4.19) for $1 \leq j \leq r$, we have

$$\begin{aligned} \left. \frac{\partial S}{\partial f_j} \right|_{f_{k \neq j}} &= -k_B \left[\lambda_j + \sum_{i=1}^r f_i \frac{\partial \lambda_i}{\partial f_j} - \sum_{i=1}^r \frac{\partial \ln Z_r}{\partial \lambda_i} \frac{\partial \lambda_i}{\partial f_j} \right] \\ &= -k_B \lambda_j - k_B \sum_{i=1}^r \frac{\partial \lambda_i}{\partial f_j} (f_i - f_i) \\ &= -k_B \lambda_j. \end{aligned} \quad (4.23)$$

Second, for $r < j \leq s$ we have

$$\left. \frac{\partial S}{\partial f_j} \right|_{f_{k \neq j}} = -k_B \left[\sum_{i=1}^r f_i \frac{\partial \lambda_i}{\partial f_j} - \left. \frac{\partial \ln Z_r}{\partial f_j} \right|_{f_{k \neq j}} \right]. \quad (4.24)$$

It is important to notice that the λ_i with $1 \leq i \leq r$ depend on all f_i , including the f_j in the range $r + 1 \leq j \leq s$, since the λ_i need to adapt to any change of f_j in order to keep the other averages f_1, \dots, f_r fixed. The dependence of $\ln Z_r$ on f_j for f_i fixed ($i \neq j$) has therefore two sources. On the one hand we $\partial\lambda_i/\partial f_j$ for $1 \leq i \leq r$, in order to keep the averages f_1, \dots, f_r fixed, and, on the other hand, we have the straightforward dependence of Z_r on the additive constant of motion f_j for $r + 1 \leq j \leq s$. Developing the last term in Eq. (4.24) accordingly we have

$$\begin{aligned} \left. \frac{\partial \ln Z_r}{\partial f_j} \right|_{f_{k \neq j}} &= \sum_{i=1}^r \frac{\partial \ln Z_r}{\partial \lambda_i} \frac{\partial \lambda_i}{\partial f_j} + \frac{\partial \ln Z_r}{\partial f_j} \\ &= \sum_{i=1}^r f_i \frac{\partial \lambda_i}{\partial f_j} + \frac{\partial \ln Z_r}{\partial f_j}, \end{aligned}$$

where, as usual, we implicitly assume that the other natural variables of Z_r , namely, $\lambda_1, \dots, \lambda_r$ and f_{r+1}, \dots, f_s , are kept constant in the partial derivative of the right hand side. The second line follows from Eq. (4.22). Replacing in Eq. (4.24) we have

$$\frac{\partial S}{\partial f_j} = -k_B \left[\sum_{i=1}^r f_i \frac{\partial \lambda_i}{\partial f_j} - \sum_{i=1}^r f_i \frac{\partial \lambda_i}{\partial f_j} - \frac{\partial \ln Z_r}{\partial f_j} \right],$$

where the first two terms cancel out. Moreover, recalling that $\ln Z_r = -\sum_{j=r+1}^s \lambda_j f_j$ we have $\frac{\partial \ln Z_r}{\partial f_j} = -\lambda_j$ for $r < j \leq s$. Thus, we finally obtain

$$\left. \frac{\partial S}{\partial f_j} \right|_{f_{i \neq j}} = k_B \left. \frac{\partial \ln Z_r}{\partial f_j} \right|_{\lambda_i} = -k_B \lambda_j. \quad (4.25)$$

In conclusion, we have thus reached the most important result

$$\left. \frac{\partial S}{\partial f_j} \right|_{f_i} = -k_B \lambda_j \quad (4.26)$$

for all $j = 1, \dots, s$. This holds for all types of ensembles, from the completely closed microcanonic one ($r = 0$), over the partly closed canonic ones ($1 \leq r < s$) to the completely open grandcanonic one ($r = s$).

Example: Consider a single-component system with constants of motion $f_1 = E$, $f_2 = N$ and $f_3 = V$. In the grand canonical pressure ensemble, where the system is open with respect to all f_i , i.e., E , N and V , we have

$$\hat{\rho}_{gc} = e^{-\beta(\hat{H} - \mu\hat{N} + pV)}.$$

The derivatives of $S = S(E, N, V)$ are

$$\left. \frac{\partial S}{\partial E} \right|_{N, V} = -k_B(-\beta) = \frac{1}{T}, \quad (4.27)$$

$$\left. \frac{\partial S}{\partial N} \right|_{E,V} = -k_B(\beta\mu) = -\frac{\mu}{T} \quad (4.28)$$

and

$$\left. \frac{\partial S}{\partial V} \right|_{E,N} = -k_B(-\beta p) = \frac{p}{T}. \quad (4.29)$$

Once we identify $\lambda_1 = -\beta = -1/k_B T$, $\lambda_2 = \beta\mu = \mu/k_B T$, and $\lambda_3 = -\beta p = -p/k_B T$, we may express any entropy change resulting from any reversible changes in E , N and V as

$$dS = \frac{1}{T}dE - \frac{\mu}{T}dN + \frac{p}{T}dV. \quad (4.30)$$

The reader familiar with thermodynamics has probably already recognized that this equation hides the first and second principles. In conclusion, the entropy allows us to calculate the parameters $\lambda_1, \dots, \lambda_r$ directly from its derivatives with respect to f_1, \dots, f_r without needing to solve the equations $\langle \hat{F}_i \rangle(\lambda_1, \dots, \lambda_r) = f_i$ for $1 \leq i \leq r$.

As an important corollary of Eq. (4.26) we can demonstrate the additivity of S . Recalling Eq. (4.21),

$$S = -k_B \sum_{i=1}^s \lambda_i f_i, \quad (4.31)$$

and replacing in it Eq. (4.26) we conclude that S satisfies the Euler condition

$$S = \sum_{i=1}^s \frac{\partial S}{\partial f_i} f_i. \quad (4.32)$$

Applying Euler's homogeneous function theorem (see section 4.3.1) we conclude that the entropy is an *homogeneous function of degree 1* and thus satisfies

$$S(\alpha f_1, \dots, \alpha f_s) = \alpha S(f_1, \dots, f_s) \quad (4.33)$$

for all $\alpha > 0$. This important property means that if one changes the values of all additive constants of motion f_i by the same scaling factor α , then the entropy also changes by the same factor α . This kind of properties are said to be *extensive*.

Consider now two isolated systems A and B with additive constants of motion f_i^A and f_i^B . The entropy of the ensemble, before allowing any exchange of f_i , is $S(f_1^A, \dots, f_s^A) + S(f_1^B, \dots, f_s^B)$, since S is additive. If now exchanges are allowed, the pair AB (which is in general out of equilibrium at the beginning) will evolve towards equilibrium by exchanging f_i keeping the sum $f_i^A + f_i^B$ constant. Since equilibration occurs spontaneously, the entropy must increase (or remain the same) during this process. Therefore, the entropy of the ensemble $S(f_1^A + f_1^B, \dots, f_s^A + f_s^B)$ at equilibrium satisfies

$$S(f_1^A, \dots, f_s^A) + S(f_1^B, \dots, f_s^B) \leq S(f_1^A + f_1^B, \dots, f_s^A + f_s^B).$$

The equal sign holds when the two systems were already in equilibrium before being put together.

In order to discuss the previous statement in more detail, suppose that the ratios between all the f_i 's in both systems are the same :

$$\frac{f_i^A}{f_1^A} = \frac{f_i^B}{f_1^B} \Leftrightarrow \frac{f_i^A}{f_i^B} = \frac{f_1^A}{f_1^B} = \alpha \Leftrightarrow f_i^A = \alpha f_i^B \quad \forall i.$$

It then follows that

$$S(f_i^A) + S(f_i^B) = S(\alpha f_i^B) + S(f_i^B) = (\alpha + 1) S(f_i^B) = S[(\alpha + 1)f_i^B] = S(f_i^A + f_i^B).$$

One concludes that if the ratios f_i/f_j are kept constant, the entropy is just proportional to the size of the system.

What about the converse? What relation holds between the intensive properties λ_i of two systems A and B , for which the entropy of the equilibrated ensemble AB is simply equal to the sum of the entropies before contact? In this case

$$S(f_i^A) + S(f_i^B) = S(f_i^A + f_i^B)$$

implies

$$\frac{\partial S}{\partial f_i^A} = \frac{\partial S}{\partial f} \Big|_{f=f_i^A+f_i^B} = \frac{\partial S}{\partial f_i^B} \Rightarrow \lambda_i^A = \lambda_i^B$$

for all i . We conclude that the equal sign holds if and only if the parts were already in equilibrium before exchanges of the additive constants of motion are allowed.

The coefficients λ_i defining $\hat{\rho}_r$ and Z_r are partial derivatives of S , which is an homogeneous function of first degree:

$$\frac{\partial S}{\partial f_j} = -k_B \lambda_j.$$

Therefore, they are homogeneous functions of degree zero, i.e.,

$$\lambda_i(\alpha f_1, \dots, \alpha f_s) = \lambda_i(f_1, \dots, f_s). \quad (4.34)$$

Such properties are said to be *intensive*. They do not change when the extensive properties are scaled, i.e., changed by keeping the ratios f_i/f_j constant. The λ_i depend only on $s - 1$ independent ratios, for example, on $f_2/f_1, f_3/f_1, \dots, f_s/f_1$. That the λ_i are not all independent is consistent with the fact that they satisfy the equation of state given by the normalization condition $\text{Tr}\{\rho_{gcp}\} = 1$.

Examples of the intensive properties λ_i are the temperature T or $\beta = 1/k_B T$, the chemical potential μ , and the pressure p . Note that any function of intensive properties is also intensive. In particular β , μ and p depend on the intensive variables E/N and V/N .

In order to summarize a few comments are due:

- i) The knowledge of the parameters $\lambda_1, \dots, \lambda_s$ gives us no information on the size of the system. For instance, knowing the temperature and pressure gives no information of the number of atoms or the volume.

- ii) The λ_i with $i = 1, \dots, s$ cannot all be independent of each other, i.e., they cannot all be varied at will, since they depend on only $s - 1$ densities, such as $f_2/f_1, f_3/f_1, \dots, f_s/f_1$. Therefore, T, p and μ cannot be chosen at will. As an example, consider a non interacting classical gas, for which we have $pV = Nk_B T$ or $p\frac{V}{N} = k_B T$.
- iii) The relation between the λ_i 's depends of course on the type of system that we are studying, i.e., on its Hamiltonian, composition and interactions, not on the size of the system. This relation is known as equation of state of the material. It is a result of quantum mechanics and statistical mechanics. It cannot be obtained by thermodynamic arguments.
- iv) The dependence of the λ_i among each other has already been mentioned, when introducing the grand canonical density operator

$$\hat{\rho}_{gc} = e^{\sum_{i=1}^s \lambda_i \hat{F}_i},$$

which satisfies

$$Z_{gc} = \text{Tr}\{e^{-\sum_{i=1}^s \lambda_i \hat{F}_i}\} = 1.$$

In fact, the condition $Z_{gc} = 1$ defines the relation between the intensive quantities $\lambda_1, \dots, \lambda_s$ known as equation of state. Note that Z_{gc} depends only on λ_i , once the operators \hat{F}_i giving the additive constants of motion in the system under study have been specified. In the most usual case, these are the Hamiltonian \hat{H} , the number particles N , and the volume V .

The logarithm of the partition functions Z_r is also extensive. This can be seen easily by noting that

$$k_B \ln Z_r = S + k_B \sum_{i=1}^r \lambda_i f_i \quad (4.35)$$

from Eq. (4.19), or by the relation

$$\ln Z_r = - \sum_{j=r+1}^s \lambda_j f_j, \quad (4.36)$$

taking into account that the λ_j are intensive properties [see Eq. (4.34)]. This means that

$$\ln Z_r(\lambda_1 \dots \lambda_r, \alpha f_{r+1}, \alpha f_s) = \alpha \ln Z_r(\lambda_1 \dots \lambda_r, f_{r+1} \dots f_s). \quad (4.37)$$

From Eq. (4.25) we know that

$$\frac{\partial \ln Z_r}{\partial f_j} = -\lambda_j \quad (4.38)$$

for $r + 1 \leq j \leq s$. Replacing in Eq. (4.36) we have

$$\ln Z_r = \sum_{j=r+1}^s f_j \frac{\partial \ln Z_r}{\partial f_j},$$

which is consistent with Euler's theorem, since Z_r is extensive with respect to f_{r+1}, \dots, f_s .

As a final consequence of the extensiveness of the logarithm of the partition functions, it is interesting to revisit the dependence of the density of states $\Omega(E)$ on system size. We know that $S(E, V, N) = k_B \ln \Omega(E)$ is an extensive property, which implies

$$\ln[\Omega(E, V, N)] = N \ln \left[\Omega\left(\frac{E}{N}, \frac{V}{N}, 1\right) \right].$$

Denoting $\omega\left(\frac{E}{N}, \frac{V}{N}\right) = \Omega\left(\frac{E}{N}, \frac{V}{N}, 1\right)$ we have

$$\Omega(E, V, N) = \left[\omega\left(\frac{E}{N}, \frac{V}{N}\right) \right]^N, \quad (4.39)$$

which confirms that the density of states is an extremely rapidly increasing function of the system size.

4.3.1 Euler theorem for homogeneous functions

Let $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous differentiable function. We say that ϕ is homogeneous of degree k if

$$\phi(\alpha \vec{x}) = \alpha^k \phi(\vec{x}).$$

Euler's theorem states that $\phi(x_1, \dots, x_n)$ is homogeneous of degree k if and only if

$$\sum_{i=1}^n \frac{\partial \phi}{\partial x_i} x_i = k \phi. \quad (4.40)$$

Proof: We assume that ϕ is an homogeneous function of degree k , i.e., $\phi(\alpha \vec{x}) = \alpha^k \phi(\vec{x})$. Differentiation with respect to α then yields

$$\frac{d\phi(\alpha \vec{x})}{d\alpha} = k \alpha^{k-1} \phi(\vec{x}).$$

In addition we know that in general

$$\frac{d\phi(\alpha x_1, \dots, \alpha x_n)}{d\alpha} = \sum_i \frac{\partial \phi}{\partial x_i}(\alpha \vec{x}) x_i. \quad (4.41)$$

Consequently,

$$\sum_i \frac{\partial \phi}{\partial x_i}(\alpha \vec{x}) x_i = k \alpha^{k-1} \phi(\vec{x}).$$

Setting $\alpha = 1$, we have

$$\sum_i \frac{\partial \phi}{\partial x_i}(\vec{x}) x_i = k \phi(\vec{x}),$$

which proves the statement.

In order to demonstrate the converse, we start from Eq.(4.40) at the point $\alpha\vec{x}$, namely,

$$\alpha \sum_i \frac{\partial \phi_i}{\partial x_i}(\alpha\vec{x}) x_i = k\phi(\alpha\vec{x}),$$

and compare it with the general relation (4.41). In this way we obtain

$$\alpha \frac{d}{d\alpha} \phi(\alpha\vec{x}) = k\phi(\alpha\vec{x}).$$

The solution of this first-order linear differential equation in the variable α is $\phi(\alpha\vec{x}) = A\alpha^k$, where A is independent of α . Setting $\alpha = 1$ we have $A = \phi(\vec{x})$ and $\phi(\alpha\vec{x}) = \alpha^k \phi(\vec{x})$, which means that $\phi(\vec{x})$ is homogeneous of degree k .

Corollary:

If $\phi: \mathbb{R}^n \rightarrow \mathbb{R}$ is an homogeneous function of degree k , then the partial derivatives $\partial\phi/\partial x_j$ are all homogeneous functions of degree $k - 1$. The homogeneity of $\phi(\vec{x})$ implies

$$\begin{aligned} \sum_{i=1}^n x_i \frac{\partial \phi}{\partial x_i} &= k\phi \\ \Rightarrow \sum_{i=1}^n x_i \frac{\partial}{\partial x_i} \left(\frac{\partial \phi}{\partial x_j} \right) + \frac{\partial \phi}{\partial x_j} &= k \frac{\partial \phi}{\partial x_j} \\ \Rightarrow \sum_{i=1}^n x_i \frac{\partial}{\partial x_i} \left(\frac{\partial \phi}{\partial x_j} \right) &= (k - 1) \frac{\partial \phi}{\partial x_j}. \end{aligned}$$

This means that $\partial\phi/\partial x_j$ is homogeneous of degree $k - 1$ for all j .

4.4 Thermodynamic potentials: General formulation

Energy plays a central role in any mechanical theory and statistical mechanics is no exception. This manifests itself in the physical importance of the canonical and grand canonical ensembles and in the convention used for denoting the intensive variables λ_i . The first and most important observable is $\hat{F}_1 = \hat{H}$ and the corresponding parameter is $\lambda_1 = -\beta$, so that the canonical density operator is written as

$$\hat{\rho}_c = \frac{1}{Z_c} e^{-\beta \hat{H}}$$

with

$$Z_c = \text{Tr}\{e^{-\beta \hat{H}}\}.$$

From Eq. (4.26) we have

$$\left. \frac{\partial S}{\partial E} \right|_{N,V} = k_B \beta.$$

In order to interpret β physically we consider two previously isolated subsystems 1 and 2, with energies E_1 and E_2 which are brought into thermal contact. The ensemble 1 + 2 remains isolated but the subsystems can exchange energy between them. If any energy transfer occurs in the process of reaching a new equilibrium situation, for example, if subsystem 1 changes its energy to $E'_1 = E_1 + \Delta E$ the total $E = E_1 + E_2 = E'_1 + E'_2$ must remain constant, since the ensemble 1 and 2 is isolated. Thus, $E'_2 = E_2 - \Delta E$. The additivity of the entropy implies then that for any given ΔE we have

$$S = S_1(E'_1) + S_2(E'_2) = S_1(E_1 + \Delta E) + S_2(E_2 - \Delta E).$$

The change of entropy associated with a small energy transfer ΔE is therefore given by

$$\Delta S = \frac{\partial S_1}{\partial E} \Delta E - \frac{\partial S_2}{\partial E} \Delta E = k_B (\beta_1 - \beta_2) \Delta E + O(\Delta E^2).$$

Since ΔS must necessarily be positive or zero, we have

$$\Delta E > 0 \Leftrightarrow \beta_1 > \beta_2,$$

or equivalently,

$$\Delta E > 0 \Leftrightarrow \frac{1}{\beta_1} < \frac{1}{\beta_2}.$$

Thus, the energy is absorbed by the system having the smallest $1/\beta$ and equilibrium is reached when $\beta_1 = \beta_2$, i.e., when $\Delta S = 0$.

Notice that the total entropy change $\Delta S = \Delta S_1 + \Delta S_2$ is always positive. However, this is not necessarily true for the subparts. For example, for $\beta_1 > \beta_2$, we have $\Delta S_1 = k_B \beta_1 \Delta E > 0$ and $\Delta S_2 = -k_B \beta_2 \Delta E < 0$. The total entropy increases since the entropy of the cooler subsystem increases more than the decrease of the entropy of the hotter subsystem. Therefore, $1/\beta$ has the properties of a temperature in thermodynamic sense. We have therefore defined the absolute temperature as

$$T = \frac{1}{k_B \beta} \quad \Leftrightarrow \quad \beta = \frac{1}{k_B T}. \quad (4.42)$$

The other intensive variables are redefined with respect to β . Denoting them as α_i , they are given by

$$\lambda_i = -\beta \alpha_i,$$

or equivalently,

$$-k_B \lambda_i = \frac{\alpha_i}{T}$$

for $i = 2, \dots, s$. For $i = 1$ we already have $\lambda_1 = -\beta$ with $\alpha_1 = 1$ and $f_1 = E$. As we shall see, the intensive variables α_i are physically more appealing than the λ_i 's. For example, $\alpha_2 = -\mu$ is the opposite of the chemical potential, which measures the energy changes associated with changes in the particle number, and $\alpha_3 = p$ is the pressure, which measures the energy changes associated with changes on volume..

Consequently, we write

$$\hat{\rho}_r = \frac{1}{Z_r} e^{-\beta(\hat{H} + \sum_{i=2}^r \alpha_i \hat{F}_i)} \quad (4.43)$$

with

$$Z_r = \text{Tr}\{e^{-\beta(\hat{H} + \sum_{i=2}^r \alpha_i \hat{F}_i)}\}. \quad (4.44)$$

Furthermore, we have seen that $\ln Z_r$ is an extensive property from which a number of important information of the equilibrium state can be derived, particularly in the form of derivatives with respect to λ_i and f_i . For instance, we know that

$$-\frac{\partial \ln Z_r}{\partial \beta} = \langle \hat{H} \rangle = E$$

and

$$\frac{\partial \ln Z_r}{\partial \lambda_i} = \langle \hat{F}_i \rangle = f_i.$$

In terms of α_i , this is equivalent to

$$-\frac{1}{\beta} \frac{\partial \ln Z_r}{\partial \alpha_i} = f_i \quad (4.45)$$

for $i = 2, \dots, r$. It is therefore very useful to introduce the *thermodynamic potential*

$$\Phi_r = -\frac{1}{\beta} \ln Z_r = -k_B T \ln Z_r \quad (4.46)$$

corresponding to the canonical ensemble, which is open with respect to r additive constants of motion $\hat{F}_1, \dots, \hat{F}_r$ and closed with respect to $\hat{F}_{r+1}, \dots, \hat{F}_s$. Indeed, in terms of ϕ_r we simply have

$$\frac{\partial \Phi_r}{\partial \alpha_i} = f_i, \quad (4.47)$$

for $i = 2, \dots, r$.

The functions Φ_r are named thermodynamic potentials because their partial derivatives give direct access to the most fundamental thermodynamic properties. As for Z_r , the natural variables of Φ_r are $T, \alpha_2, \dots, \alpha_r$ and f_{r+1}, \dots, f_s . It is therefore important to analyze how Φ_r depends on these variables in the various ensembles. We have already seen that the dependence on $\alpha_2, \dots, \alpha_r$ is given by Eq. (4.47). Concerning the additive quantities f_j with $j = r+1, \dots, s$, we have

$$\frac{\partial \Phi_r}{\partial f_j} = -\frac{1}{\beta} \frac{\partial \ln Z_r}{\partial f_j} = \frac{1}{\beta} \lambda_j = -\alpha_j. \quad (4.48)$$

Finally, in order to quantify the dependence on temperature we need to compute $\partial \Phi_r / \partial T$ keeping all other variables $\alpha_2, \dots, \alpha_r$ and f_{r+1}, \dots, f_s constant. To this aim it is useful to note the following simple relation

$$T \frac{\partial}{\partial T} = -\beta \frac{\partial}{\partial \beta}. \quad (4.49)$$

We then have

$$\begin{aligned} T \left. \frac{\partial \Phi_r}{\partial T} \right|_{\alpha_i, f_j} &= -\beta \frac{\partial}{\partial \beta} \left(-\frac{1}{\beta} \ln Z_r \right) = \beta \frac{\partial}{\partial \beta} \left(\frac{1}{\beta} \ln Z_r \right) \\ &= \beta \left[-\frac{1}{\beta^2} \ln Z_r + \frac{1}{\beta} \sum_{i=1}^r \frac{\partial \ln Z_r}{\partial \lambda_i} \frac{\partial \lambda_i}{\partial \beta} \right]. \end{aligned}$$

Since $\lambda_1 = -\beta$ and $\lambda_i = -\beta \alpha_i$ for $i = 2, \dots, r$, and recalling that $\frac{\partial \ln Z_r}{\partial \lambda_i} = f_i$ for $1 \leq i \leq r$ ($f_1 = E$) we have

$$\begin{aligned} T \left. \frac{\partial \Phi_r}{\partial T} \right|_{\alpha_i, f_j} &= -k_B T \ln Z_r - E - \sum_{i=2}^r f_i \alpha_i \\ &= \Phi_r - E - \sum_{i=2}^r \alpha_i f_i. \end{aligned} \tag{4.50}$$

This expression can be remarkably simplified once Φ_r is expressed in terms of E , S and the intensive variables $T, \alpha_2, \dots, \alpha_r$. From the definition (4.46) of Φ_r and the expression (4.19) for S we have

$$\begin{aligned} \Phi_r &= -k_B T \ln Z_r = -TS - k_B T \sum_{i=1}^r \lambda_i f_i \\ &= E - TS + \sum_{i=2}^r \alpha_i f_i. \end{aligned} \tag{4.51}$$

This equation, often taken as a definition of Φ_r , expresses Φ_r in terms of E , S and the pairs of intensive and extensive properties α_i and f_i with respect to which the system is open ($2 \leq i \leq r$). It shows that Φ_r is an extensive property. In addition, an interesting alternative expression of Φ_r can be obtained by exploiting the relation of $\ln Z_r$ and the variables with respect to which the system is closed:

$$\ln Z_r = - \sum_{j=r+1}^s \lambda_j f_j.$$

It follows straightforwardly that

$$\begin{aligned} \Phi_r &= -\frac{1}{\beta} \ln Z_r = \sum_{j=r+1}^s \frac{\lambda_j}{\beta} f_j \\ &= - \sum_{j=r+1}^s \alpha_j f_j. \end{aligned} \tag{4.52}$$

We may now focus our attention back on the temperature dependence of Φ_r and its derivative with respect to T . Replacing in it Eq. (4.51) in Eq. (4.50) we obtain

$$T \frac{\partial \Phi_r}{\partial T} = E - TS + \sum_2^r \alpha_i f_i - E - \sum_2^r \alpha_i f_i$$

and therefore

$$\frac{\partial \Phi_r}{\partial T} = -S. \quad (4.53)$$

This important relation allows us to obtain S directly as the partial derivative of the thermodynamic potential Φ_r with respect to T .

For the sake of completeness we give here the derivatives of S with respect to f_i in terms of the newly introduced physical magnitudes T and α_i . Eq. (4.26) reads

$$\frac{\partial S}{\partial E} = \frac{\partial S}{\partial f_1} = -k_B \lambda_1 = k_B \beta = \frac{1}{T}$$

and

$$\frac{\partial S}{\partial f_i} = -k_B \lambda_i = k_B \beta \alpha_i = \frac{\alpha_i}{T}$$

for $i \geq 2$.

4.4.1 Legendre transformations

The different thermodynamic potentials Φ_r are related through Legendre transformations. In order to obtain $\Phi_r = \Phi_r(T, \alpha_2, \dots, \alpha_r, f_{r+1}, \dots, f_s)$, which is a function of α_r , from $\Phi_{r-1} = \Phi_{r-1}(T, \alpha_2, \dots, \alpha_{r-1}, f_r, \dots, f_s)$, which is a function of f_r , we first need to replace f_r by α_r in Φ_{r-1} . Therefore, one should first solve

$$\frac{\partial \Phi_{r-1}}{\partial f_r}(T, \alpha_2, \dots, \alpha_{r-1}, f_r, \dots, f_s) = -\alpha_r$$

in order to obtain

$$f_r = f_r(T, \alpha_2, \dots, \alpha_r, f_{r+1}, \dots, f_s).$$

The thermodynamic potential is then given by

$$\Phi_r(T, \alpha_2, \dots, \alpha_r, f_{r+1}, \dots, f_s) = \Phi_{r-1}(T, \alpha_2, \dots, \alpha_{r-1}, f_r, \dots, f_s) + \alpha_r f_r \dots \quad (4.54)$$

It is easy to verify that Φ_r , constructed in this way, has the same partial derivatives than Φ_{r-1} , given by Eq. (4.46) and therefore coincides with it. In fact, from Eq. (4.54) we have

$$\frac{\partial \Phi_r}{\partial T} = \frac{\partial \Phi_{r-1}}{\partial T} + \underbrace{\frac{\partial \Phi_{r-1}}{\partial f_r}}_{-\alpha_r} \cdot \frac{\partial f_r}{\partial T} + \alpha_r \frac{\partial f_r}{\partial T} = -S$$

and

$$\frac{\partial \Phi_r}{\partial \alpha_i} = \frac{\partial \Phi_{r-1}}{\alpha_i} + \underbrace{\frac{\partial \Phi_{r-1}}{\partial f_r}}_{-\alpha_r} \cdot \frac{\partial f_r}{\partial \alpha_i} + \alpha_r \frac{\partial f_r}{\partial \alpha_i} = f_i$$

for $2 \leq i \leq r-1$. Moreover,

$$\frac{\partial \Phi_r}{\partial \alpha_r} = \underbrace{\frac{\partial \Phi_{r-1}}{\partial f_r}}_{-\alpha_r} \cdot \frac{\partial f_r}{\partial \alpha_r} + f_r + \alpha_r \frac{\partial f_r}{\partial \alpha_r} = f_r$$

and finally

$$\frac{\partial \Phi_r}{\partial f_j} = \underbrace{\frac{\partial \Phi_{r-1}}{\partial f_r}}_{-\alpha_r} \cdot \frac{\partial f_r}{\partial f_j} + \underbrace{\frac{\partial \Phi_{r-1}}{\partial f_j}}_{-\alpha_j} + \alpha_r \frac{\partial f_r}{\partial f_j} = -\alpha_j$$

for $r+1 \leq j \leq s$.

Thermodynamic potentials of closed systems

Before closing this section we would like to discuss the thermodynamic potentials corresponding to the microcanonical ensemble, which cannot be obtained from Eq. (4.46) or Eq. (4.51). In this case the extensive properties are E , N and V and in addition the entropy S . These four properties are of course related since E , N and V alone define the macroscopic state.

From Eq. (4.51) we have

$$F = \Phi_1 = E - TS.$$

The canonical potential Φ_1 , usually denoted by F , is known as *Helmholtz free energy*. The energy, a function of S and the remaining constants of motion f_2, \dots, f_s , is then given by

$$E = E(S, f_2, \dots, f_s) = TS + \Phi_1. \quad (4.55)$$

In order to obtain an explicit expression for E in terms of S, f_2, \dots, f_s we calculate the partial derivatives and take advantage of the extensivity of E . One thus obtains

$$\left. \frac{\partial E}{\partial S} \right|_{f_i} = T + S \frac{\partial T}{\partial S} + \underbrace{\frac{\partial \Phi_1}{\partial T}}_{-S} \frac{\partial T}{\partial S} = T. \quad (4.56)$$

and

$$\left. \frac{\partial E}{\partial f_i} \right|_{S, f_j \neq i} = S \frac{\partial T}{\partial f_i} + \underbrace{\frac{\partial \Phi_1}{\partial f_i}}_{-\alpha_i} + \underbrace{\frac{\partial \Phi_1}{\partial T}}_{-S} \frac{\partial T}{\partial f_i} = -\alpha_i \quad (4.57)$$

for $i = 2, \dots, s$. We may thus write

$$dE = TdS - \sum_{i=2}^s \alpha_i df_i. \quad (4.58)$$

The extensiveness of E (Euler theorem) implies that

$$E = TS - \sum_{i=2}^s \alpha_i f_i. \quad (4.59)$$

The same result could have been obtained from the general expression (4.31) for the entropy

$$S = -k_B \sum_{i=1}^s \lambda_i f_i = \frac{1}{T} E + \sum_{i=2}^s \frac{\alpha_i}{T} f_i, \quad (4.60)$$

or even by combining Eqs. (4.51) and (4.52) for $r = 1$.

5 Thermodynamic properties

The purpose of this chapter is to discuss the thermodynamic potentials and derived thermodynamic properties of practical interest by applying the general formulation developed in the previous chapter.

5.1 Thermodynamic potentials: special cases

5.1.1 Energy E and Entropy S

We consider a system composed of particles of the same kind. The extensive constants of motion which characterize the macroscopic state are the energy E , the number of particles N and the volume V . The other important extensive property of statistical interest is the entropy S . The relation between these four properties can be expressed as

$$S = S(E, N, V) \quad \text{or} \quad E = E(S, N, V).$$

The partial derivatives are the intensive properties known as temperature

$$\left. \frac{\partial E}{\partial S} \right|_{V, N} = T, \quad (5.1)$$

chemical potential

$$\left. \frac{\partial E}{\partial N} \right|_{S, V} = \mu, \quad (5.2)$$

and pressure

$$\left. \frac{\partial E}{\partial V} \right|_{S, N} = -p. \quad (5.3)$$

The physical meaning of the intensive properties becomes clear if one considers the total differential of the energy

$$dE = TdS - pdV + \mu dN. \quad (5.4)$$

$TdS = \delta Q_R$ is the change in energy by constant volume (isochoric process) and constant number of particles. This is known as *reversible heat*.

The *chemical potential* μ represents the change in energy when a particle is added to the system by keeping the volume and the entropy constant. In other words, μ is the energy that a particle must bring ($\mu > 0$) or that the system must release ($\mu < 0$) when a particle is added in order to keep the entropy unchanged at constant volume V .

Finally, p is the *pressure* since $\delta W = -pdV$ is the change of the energy of the system associated with a reversible volume change dV , which is equal to the work done by the external forces without involving any heat or particle exchange ($dS = 0$ and $dN = 0$).

Processes without volume change ($dV = 0$) are known as *isochoric*. When the entropy is constant we call them *isoentropic*. Processes without heat exchange ($\delta Q = 0$) are

known as *adiabatic*. Processes in which the system is always in thermodynamic equilibrium are called *reversible*. Thus, reversible and adiabatic means isentropic since in this case $TdS = \delta Q_R = 0$.

The extensiveness of E implies

$$E = TS - pV + \mu N \quad (5.5)$$

We can also express $S = S(E, N, V)$ as a function of the mechanical constant of motion:

$$S = \frac{1}{T}E - \frac{\mu}{T}N + \frac{p}{T}V \quad (5.6)$$

and

$$dS = \frac{1}{T}dE - \frac{\mu}{T}dN + \frac{p}{T}dV. \quad (5.7)$$

The energy and the entropy are known as *thermodynamic potentials*, since the thermodynamic properties such as T , p and μ , which characterize the macroscopic state of the system, can be obtained as their partial derivatives with respect to the variables that define them. As all other thermodynamic potentials, they are *state functions*, since their values depend only on the actual macroscopic state of the system and not on the way by which this state has been achieved. In contrast, for example, the exchanged heat and the delivered work are not state functions.

5.1.2 Helmholtz free energy F

In addition, there are other four potentials which have great practical significance. The first one is the Helmholtz free energy

$$F = \Phi_1 = -k_B T \ln Z_c = -k_B T \ln \left[\text{Tr} \{ e^{-\beta \hat{H}} \} \right], \quad (5.8)$$

which is a function of T , N and V . Using the general expression (4.51) of the thermodynamic potentials for $r = 1$ we have

$$F = E - TS. \quad (5.9)$$

The partial derivatives are obtained by applying Eqs. (4.48) and (4.53):

$$\left. \frac{\partial F}{\partial T} \right|_{V,N} = -S, \quad (5.10)$$

$$\left. \frac{\partial F}{\partial N} \right|_{T,V} = \mu \quad (5.11)$$

$$\text{and} \quad (5.12)$$

$$\left. \frac{\partial F}{\partial V} \right|_{T,N} = -p. \quad (5.13)$$

This implies

$$dF = -S dT + \mu dN - p dV . \quad (5.14)$$

Applying Eq. (4.52), or recalling that F is extensive and using Euler's theorem, we can express F as a linear combination of its natural extensive variables V and N :

$$F = \mu N - pV . \quad (5.15)$$

It is clear that all these expressions could have been derived by applying the general Legendre transformation formalism developed in the previous section to the present specific cases, as is usually done in most thermodynamic textbooks. For instance, if one considers physical situations in which T , instead of S , is the variable under control, one performs a Lagrange transformation on E by subtracting TS . The resulting thermodynamic potential is the Helmholtz free energy $F = E - TS$, whose differential form reads

$$dF = dE - TdS - SdT = TdS - pdV + \mu dN - TdS - SdT = -SdT - pdV + \mu dN .$$

This is equivalent to Eqs. (5.10)–(5.13). The extensiveness of F , which follows from the extensiveness of E and S and the intensiveness of T , implies Eq. (5.15).

In order to provide a physical interpretation to F we consider a process in which the volume is changed keeping the temperature and the particle number constant. In this case we have

$$dF = -pdV$$

since $dT = 0$ and $dN = 0$. Thus, dF represents the available energy, or free energy, in isothermal processes. Note that the corresponding change in internal energy $dE = TdS - pdV$ is equal to $dF + TdS$ for $dT = 0$. As the system expands at constant T , heat has to be absorbed in order to keep the temperature constant ($TdS > 0$). Therefore, $dE > dF$ for $dV > 0$ and $dT = 0$ ($dF < 0$). In general $|dF| > |dE|$ for $dT = 0$.

Taking into account that

$$F = -k_B T \ln \left[\text{Tr} \left\{ e^{-\beta \hat{H}} \right\} \right]$$

we may write

$$Z_c = e^{-\beta F} \quad (5.16)$$

and

$$\hat{\rho}_c = \frac{1}{Z_c} e^{-\beta \hat{H}} = e^{\beta(F - \hat{H})} = e^{-\beta(\hat{H} - F)} . \quad (5.17)$$

From a statistical perspective, F can be obtained, as all the other thermodynamic potentials, directly from the corresponding partition function.

We may also verify that the usual statistical definition of entropy

$$S = -k_B \langle \ln \hat{\rho}_c \rangle \quad (5.18)$$

in the canonical ensemble and its well-known physical interpretation coincide with the thermodynamic concept of entropy, since Eqs. (5.18) and Eq. (5.9) are actually the same. Indeed, replacing Eq. (5.17) for $\hat{\rho}_c$ in Eq. (5.18) we have

$$S = -k_B \langle \ln \hat{\rho}_c \rangle = k_B \beta \langle \hat{H} - F \rangle = \frac{E - F}{T}.$$

5.1.3 Enthalpy H

The *enthalpy* is the appropriate thermodynamic potential Φ_r when the macroscopic state is defined by the entropy S , particle number N and pressure p . It is obtained by means of a Legendre transformation on $E = E(S, N, V)$ in order to replace V by p as the natural variable:

$$H(S, N, p) = E + pV, \quad (5.19)$$

which implies

$$\begin{aligned} dH &= T dS + \mu dN - p dV + p dV + V dp \\ &= T dS + \mu dN + V dp. \end{aligned} \quad (5.20)$$

Again, the extensiveness of H implies

$$H = TS + \mu N. \quad (5.21)$$

In an isobaric process ($dp = 0$) keeping the number of particles constant ($dN = 0$) the exchanged *reversible heat* $\delta Q_R = T dS$ corresponds to the change in enthalpy. Therefore H is sometimes also called *heat function*. Notice that if $\delta Q_R > 0$, usually work has to be done by the system in order to keep the pressure constant ($dV > 0$ for $\alpha > 0$). Thus, the change in energy is smaller than the enthalpy change: $dE = T dS - p dV < dH = T dS$ for $dp = 0$, $dN = 0$ and $dV > 0$. In general $|dH| > |dE|$ for $dp = 0$ ($\alpha > 0$).

5.1.4 Free enthalpy G

The *free enthalpy* $G = G(T, N, p)$ is obtained by changing the variables of the thermodynamic potential to T , N and p . This is achieved by a Legendre transformation starting either from F or from H :

$$G = F + pV = H - TS, \quad (5.22)$$

which implies

$$\begin{aligned} dG &= -S dT + \mu dN - p dV + p dV + V dp \\ &= -S dT + \mu dN + V dp. \end{aligned} \quad (5.23)$$

In terms of N , the only extensive natural variable of G , we have

$$G = \mu N. \quad (5.24)$$

Therefore, in a system having just one kind of particles, the free enthalpy per particle G/N is equal to the chemical potential μ .

5.1.5 Grand canonical potential Φ

Another important thermodynamic potential is the *grand canonical potential*

$$\Phi = -k_B T \ln \left[\text{Tr} \left\{ e^{-\beta(\hat{H} - \mu\hat{N})} \right\} \right] = -k_B T \ln Z_{gc}, \quad (5.25)$$

which is directly obtained from the grand canonical partition Z_{gc} . The natural variables of Φ are T , μ and V . In this ensemble the volume is fixed, while energy and particle-number fluctuations are possible. From the general relations (4.47) and (4.53) we have

$$\left. \frac{\partial \Phi}{\partial T} \right|_{\mu, V} = -S \quad (5.26)$$

$$\left. \frac{\partial \Phi}{\partial \mu} \right|_{T, V} = -N \quad (5.27)$$

$$\left. \frac{\partial \Phi}{\partial V} \right|_{T, \mu} = -p. \quad (5.28)$$

Consequently,

$$d\Phi = -S dT - p dV - N d\mu \quad (5.29)$$

and

$$\Phi = -pV = F - \mu N = E - TS - \mu N. \quad (5.30)$$

The density of Φ per unit volume is equal to minus the pressure: $\Phi/V = -p$.

From Eq. (5.25) it follows that

$$Z_{gc} = e^{-\beta\Phi}. \quad (5.31)$$

We may thus write

$$\hat{\rho}_{gc} = e^{\beta(\Phi - \hat{H} + \mu\hat{N})} = e^{-\beta(\hat{H} - \mu\hat{N} - \Phi)}. \quad (5.32)$$

5.1.6 Deriving properties in the grand canonical pressure ensemble

If all additive constants of motion can be exchanged with the environment, there are no restrictions to be imposed to the system's Hilbert space. In this case no normalization factor enters the definition of the density operator $\hat{\rho}_{gcp} = e^{-\beta(\hat{H} - \mu\hat{N} + p\hat{V})}$. One may say that the partition function $Z_{gcp} = \text{Tr}\{\hat{\rho}_{gcp}\} = 1$. Consequently, no thermodynamic potential can be derived from it. Moreover, since all the variables defining the ensemble (β , μ and p) are intensive properties, the size of the system is not defined. Only the densities (e.g., E/N and V/N) are known. Nevertheless, the normalization condition gives us access to the equation of state of the material under consideration.

In order to show how the equation of state comes about we express Z_{gcp} in terms of the grand canonical potential Φ as follows

$$Z_{gcp} = \text{Tr}\{e^{-\beta(\hat{H}-\mu\hat{N}+p\hat{V})}\} \quad (5.33)$$

$$= \lim_{V_m \rightarrow \infty} \frac{1}{V_m} \int_0^{V_m} e^{-\beta pV} \text{Tr}\{e^{-\beta(\hat{H}-\mu\hat{N})}\}_V dV \quad (5.34)$$

$$= \lim_{V_m \rightarrow \infty} \frac{1}{V_m} \int_0^{V_m} e^{-\beta pV} Z_{gc}(\beta, \mu, V) dV \quad (5.35)$$

$$= \lim_{V_m \rightarrow \infty} \frac{1}{V_m} \int_0^{V_m} e^{-\beta[pV+\Phi(T,\mu,V)]} dV, \quad (5.36)$$

where V_m stands for the maximum allowed volume. Since the grand canonical potential is an extensive property we have $\Phi \propto V$. Consequently,

$$Z_{gcp} = 1 \quad \Leftrightarrow \quad pV = -\Phi \quad (5.37)$$

for all V . Otherwise, Z_{gcp} would diverge or vanish. This condition coincides with the Eq. (5.30) derived in the framework of the grand canonical ensemble. Thus, the normalization condition on Z_{gcp} allows us to establish the relation between Φ and p . Conversely, even if the physical situation under study involves volume fluctuations, one may still work in the grand canonical ensemble (Sec. 5.1.5) since Eq. (5.30) guarantees the normalization of Z_{gcp} . Finally, notice that the relation

$$p = -\frac{\Phi}{V} \quad (5.38)$$

comprises only intensive properties and is therefore independent of the system size, which is given by V or N . It represents the material-specific relation known as equation of state.

5.2 Derived thermodynamic properties

5.2.1 Heat capacities and specific heats

Besides the thermodynamic potentials, it is important to consider a number of derived thermodynamic properties of physical interest. One of the first questions to be asked, in order to characterize the thermal properties of a material, is how much heat δQ_R must be transferred to the system in a reversible form in order to obtain a certain temperature change ΔT . This property, known as *heat capacity*, is defined by

$$C = \frac{\delta Q_R}{dT} = T \frac{dS}{dT}. \quad (5.39)$$

Let us recall that $TdS = dE$ when $dV = 0$ and $dN = 0$. The amount of heat needed in order to obtain a given temperature change ΔT depends of course on the type of process that one considers. For example, for a given $\Delta T > 0$, the heat $\delta Q_R > 0$ absorbed in a

transformation at constant pressure is always larger than the heat required for the same temperature change at constant V , since in the former case the system will expand and do work in the environment.

Keeping the number of particles N constant, one possibility is to consider a transformation at constant volume (isochoric process) in which case we have

$$C_{V,N} = C_V = T \left. \frac{\partial S}{\partial T} \right|_{V,N} = \left. \frac{\partial E}{\partial T} \right|_{V,N}. \quad (5.40)$$

The other most common situation is to keep the pressure constant (isobaric process) in which case we have

$$C_{p,N} = C_p = T \left. \frac{\partial S}{\partial T} \right|_{p,N} = \left. \frac{\partial H}{\partial T} \right|_{p,N}. \quad (5.41)$$

Notice that the heat capacities, as defined by Eqs. (5.39), (5.40) and (5.41), are extensive properties. Material specific values per particle or per mole, are known as *specific heat capacity* or simply *specific heat*.

5.2.2 Compressibilities

A further important question is to quantify the change in the volume of the system associated with a given reversible change in pressure. Again, such a volume change depends on the conditions imposed along the process. One therefore defines the *isothermal compressibility* as

$$\kappa_{T,N} = \kappa_T = - \left. \frac{1}{V} \frac{\partial V}{\partial p} \right|_{T,N} \quad (5.42)$$

and the *adiabatic compressibility* as

$$\kappa_{S,N} = \kappa_S = - \left. \frac{1}{V} \frac{\partial V}{\partial p} \right|_{S,N}. \quad (5.43)$$

In the former the system is compressed at constant temperature (for example, in contact with a thermal bath) and in the latter by precluding any heat exchange. Notice that κ is an intensive property. When the compression is done at constant T , the system can release internal energy to the environment (assuming $\Delta p > 0$). This is not possible in an isentropic process, in which case the temperature will increase, thus rendering any volume reduction more difficult. Consequently, we expect in general that $\kappa_T > \kappa_S$.

5.2.3 Thermal expansion

A further property of general interest is the *thermal expansion coefficient*

$$\alpha = \left. \frac{1}{V} \frac{\partial V}{\partial T} \right|_{p,N}. \quad (5.44)$$

This is an intensive property which measures the relative volume change $\Delta V/V = \alpha\Delta T$ resulting from a given temperature change at constant pressure. Note that α can be positive (the most usual case) or negative. Examples of negative α are found in water close to the freezing point ($T < 3,984$ °C) as it approaches the transition to ice, but also in several semiconductors.

5.2.4 Charge compressibility

Finally, it is interesting to quantify how the number of particles changes when the chemical potential is changed. By analogy with the volume compressibility κ discussed in Sec. 5.2.2, we define the *charge compressibility*

$$\kappa_c = \frac{1}{N} \left. \frac{\partial N}{\partial \mu} \right|_{V,T}, \quad (5.45)$$

which is also known as charge susceptibility. The latter designation, borrowed from the analogy with the magnetic susceptibility, seems less appropriate, since one should keep the term charge susceptibility to indicate changes in the charge distribution resulting from the action of some external field, which usually depends on the wave vector \vec{q} and frequency ω . The charge compressibility can be regarded as the many-body density of states at the chemical potential μ , since it measures how many states are occupied (emptied) by particles entering (leaving) the system when μ is increased (decreased): $\Delta N = \kappa_c N \Delta \mu$. The charge compressibility κ_c is often used to identify the occurrence of metal-insulator transitions since $\kappa_c > 0$ is characteristic of a metallic behavior whereas $\kappa_c = 0$ indicates the presence of a gap in the energy spectrum.

5.2.5 Assessing fluctuations

The above considered properties, which have a very clear thermodynamical interpretation, provide us with a remarkable insight into the statistical microscopic *fluctuations* of the constants of motion E , V and N , which are present in the equilibrium macroscopic mixed states. Let us first of all consider the canonical ensemble where V and N are fixed. Recalling that

$$E = \langle \hat{H} \rangle = \frac{\text{Tr} \left\{ e^{-\beta \hat{H}} \hat{H} \right\}}{\text{Tr} \left\{ e^{-\beta \hat{H}} \right\}}$$

we have

$$C_V = \left. \frac{\partial E}{\partial T} \right|_{V,N} = -\frac{1}{T} \beta \left. \frac{\partial E}{\partial \beta} \right|_{V,N} = \frac{1}{k_B T^2} \left(\langle \hat{H}^2 \rangle - \langle \hat{H} \rangle^2 \right).$$

We conclude that the heat capacity at constant volume

$$C_{V,N} = \frac{(\Delta H)^2}{k_B T^2} \geq 0 \quad (5.46)$$

gives a measure of the energy fluctuations. A larger specific heat implies stronger fluctuations of the energy at a given temperature. The extensivity of C_V implies that $c_V = C_V/N$ and C_V/E are independent of system size. Thus, the standard deviation $\sqrt{(\Delta H)^2} \propto \sqrt{N}$. We recover the known result

$$\frac{\sqrt{(\Delta H)^2}}{\langle \hat{H} \rangle} \propto \frac{1}{\sqrt{N}},$$

which means that the microcanonical and canonical ensembles are equivalent for $N \rightarrow \infty$.

In order to analyze the compressibility we turn to the grand canonical pressure ensemble and write

$$V = \langle \hat{V} \rangle = \frac{\text{Tr} \left\{ \hat{V} e^{-\beta(\hat{H} - \mu \hat{N} + p \hat{V})} \right\}}{\text{Tr} \left\{ e^{-\beta(\hat{H} - \mu \hat{N} + p \hat{V})} \right\}},$$

where we have formally introduced the operator \hat{V} measuring the volume occupied by the system. It follows that

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

and

$$\kappa_{T,N} = \frac{1}{k_B T} \frac{(\Delta V)^2}{V} \geq 0. \quad (5.47)$$

For a given T , a larger compressibility implies that the relative volume fluctuations are more important. The intensiveness of κ implies vanishing relative volume fluctuations for $N \rightarrow \infty$:

$$\frac{\sqrt{(\Delta V)^2}}{V} = \sqrt{\frac{k_B T \kappa_{T,N}}{V}} \propto \frac{1}{\sqrt{N}}.$$

In order to relate the charge compressibility to particle-number fluctuations in the grand canonical ensemble we start from

$$N = \frac{\text{Tr} \left\{ e^{-\beta(\hat{H} - \mu \hat{N})} \hat{N} \right\}}{\text{Tr} \left\{ e^{-\beta(\hat{H} - \mu \hat{N})} \right\}}$$

and obtain

$$\frac{\partial N}{\partial \mu} = \beta \left(\langle \hat{N}^2 \rangle - \langle \hat{N} \rangle^2 \right) = \frac{(\Delta N)^2}{k_B T}.$$

The charge compressibility is then given by

$$\kappa_c = \frac{1}{N} \left. \frac{\partial N}{\partial \mu} \right|_{T,V} = \frac{1}{k_B T} \frac{(\Delta N)^2}{N} \geq 0. \quad (5.48)$$

The intensiveness of κ_c implies that the relative fluctuations of the particle number

$$\frac{\sqrt{(\Delta N)^2}}{N} \propto \frac{1}{\sqrt{N}}$$

vanish for $N \rightarrow \infty$.

5.3 Minimum free-energy theorem

In section 4.1 we have demonstrated a fundamental variational principle for the entropy

$$S[\hat{\rho}] = -k_B \langle \ln \hat{\rho} \rangle = -k_B \text{Tr} \{ \hat{\rho} \ln \hat{\rho} \}$$

regarded as a functional of the density matrix $\hat{\rho}$. The maximum entropy theorem states that all the density operators $\hat{\rho}$, which comply with the normalization condition $\text{Tr} \hat{\rho} = 1$, and yield same average values the additive constants of motion f_1, \dots, f_r , under the constraint of fixed values of the other constants of motion f_{r+1}, \dots, f_s , always satisfy the inequality

$$S[\hat{\rho}] \leq S[\hat{\rho}_r], \quad (5.49)$$

where $\hat{\rho}_r$ is the corresponding equilibrium canonical density operator. It is important to be clear about the constraints imposed to $\hat{\rho}$ in order that (5.49) holds:

$$\langle \hat{F}_i \rangle = \text{Tr} \{ \hat{\rho} \hat{F}_i \} = f_i \quad \text{for } i = 1, \dots, r$$

and

$$\hat{F}_i \equiv f_i \quad \text{for } i = r + 1, \dots, s.$$

For example, in a grand canonical ensemble energy and particles can be exchanged but the volume is conserved ($r = 2$). The macrostate is in this case characterized by the values of the average energy $E = \langle \hat{H} \rangle$, the average number of particles $N = \langle \hat{N} \rangle$, and the fixed volume V . The maximum entropy theorem says that among all the mixed states $\hat{\rho}$ having $E = \langle \hat{H} \rangle$ and $N = \langle \hat{N} \rangle$ in a volume V , the equilibrium one corresponds to the $\hat{\rho}$ yielding the largest $S[\hat{\rho}] = -k_B \langle \ln \hat{\rho} \rangle$. This important result implies the second law of thermodynamics. It allows us to predict the sense of evolution of processes that occur spontaneously. In particular we have shown, taking advantage of the additivity of S that

$$S(f_1^A, \dots, f_s^A) + S(f_1^B, \dots, f_s^B) \leq S(f_1^{A+B}, \dots, f_s^{A+B})$$

when two systems A and B are brought into contact and are allowed to exchange additive constants of motion (e.g., E and N).

We seek for an appropriate variational principle for the canonical and grand canonical ensembles, in particular in order to understand how to picture the equilibrium state that is reached under the constraint of constant intensive properties, such as temperature and chemical potential. To be explicit we focus on the grand canonical ensemble with a fixed volume V , temperature T and chemical potential μ . We consider the grand canonical potential

$$\begin{aligned} \Phi[\hat{\rho}] &= E[\hat{\rho}] - \mu N[\hat{\rho}] - T S[\hat{\rho}] \\ &= \text{Tr} \left\{ \hat{\rho} \left(\hat{H} - \mu \hat{N} + T k_B \ln \hat{\rho} \right) \right\} \end{aligned} \quad (5.50)$$

as a functional of $\hat{\rho}$. The parameters T and μ characterize the ensemble and are therefore fixed, as well as the volume V . The following *minimum free-energy theorem* holds. The

minimum of $\Phi[\hat{\rho}]$ among all the density operators $\hat{\rho}$, satisfying $\text{Tr}\{\hat{\rho}\} = 1$, is achieved when $\hat{\rho}$ is equal to the equilibrium grand canonical density operator

$$\hat{\rho}_{gc} = \frac{1}{Z_{gc}} e^{-\beta(\hat{H} - \mu\hat{N})}$$

corresponding to the given T , μ and V . This can be written as

$$\Phi[\hat{\rho}_{gc}] \leq \Phi[\hat{\rho}] \quad \forall \hat{\rho} \text{ satisfying } \text{Tr}\{\hat{\rho}\} = 1, \quad (5.51)$$

or as

$$\min_{\substack{\hat{\rho} \\ \text{Tr}\{\hat{\rho}\}=1}} \{\Phi[\hat{\rho}]\} = \Phi[\hat{\rho}_{gc}]. \quad (5.52)$$

The proof is simple taking advantage of the maximum entropy theorem. However, before applying the latter, we need to keep in mind that $S[\hat{\rho}]$ is maximal only among all the $\hat{\rho}$ yielding a given $\langle\hat{H}\rangle$, $\langle\hat{N}\rangle$ and V . We therefore perform the minimization in Eq. (5.52) in two steps:

$$\min_{\hat{\rho}} \{\Phi[\hat{\rho}]\} = \min_{E,N} \left\{ \min_{\substack{\hat{\rho} \rightarrow \langle\hat{H}\rangle=E \\ \langle\hat{N}\rangle=N}} \left[\text{Tr} \left\{ \hat{\rho} \left(\hat{H} - \mu\hat{N} + T k_B \ln \hat{\rho} \right) \right\} \right] \right\}.$$

The inner minimization runs over all the $\hat{\rho}$ yielding some given average values of $E = \langle\hat{H}\rangle$ and $N = \langle\hat{N}\rangle$, while the outer minimization removes this constraint by minimizing over all possible values of E and N . Since all the $\hat{\rho}$ within the outer minimization yield the same $\langle\hat{H}\rangle = E$ and $\langle\hat{N}\rangle = N$ for the given volume V , we replace them by E and N , and write

$$\min_{\hat{\rho}} \{\Phi[\hat{\rho}]\} = \min_{E,N} \left\{ E - \mu N + \min_{\substack{\hat{\rho} \rightarrow \langle\hat{H}\rangle=E \\ \langle\hat{N}\rangle=N}} \left[T k_B \text{Tr} \left\{ \hat{\rho} \ln \hat{\rho} \right\} \right] \right\}.$$

The maximum entropy theorem implies that

$$\min_{\substack{\hat{\rho} \rightarrow \langle\hat{H}\rangle=E \\ \langle\hat{N}\rangle=N}} \left[T k_B \text{Tr} \left\{ \hat{\rho} \ln \hat{\rho} \right\} \right] = -T S(E, N, V),$$

where $S(E, N, V)$ is the entropy corresponding to the average energy E , particle number N and volume V . Moreover, the maximum S is achieved by the grand canonical density operator

$$\rho_{gc} = \frac{1}{Z_{gc}} e^{-\tilde{\beta}(\hat{H} - \tilde{\mu}\hat{N})}, \quad (5.53)$$

where $\tilde{\beta} = \tilde{\beta}(E, N, V)$ and $\tilde{\mu} = \tilde{\mu}(E, N, V)$ are the inverse temperature and chemical potential yielding the given averages E and N for the fixed volume V . In order to

determine the optimal $\tilde{\beta}$ and $\tilde{\mu}$ we must perform the minimization with respect to E and N :

$$\min_{\hat{\rho}} \{\Phi[\hat{\rho}]\} = \min_{E, N} \{E - \mu N - T S(E, N, V)\}.$$

Denoting by $F = E - \mu N - T S(E, V, N)$ the grand canonical potential which corresponds to $\tilde{\beta}$ and $\tilde{\mu}$, it is clear that the optimal E and N are given by the usual extremal conditions

$$\left. \frac{\partial F}{\partial E} \right|_{N, V} = 1 - T \left. \frac{\partial S}{\partial E} \right|_{N, V} = 0 \quad \Leftrightarrow \quad \left. \frac{\partial S}{\partial E} \right|_{N, V} = \frac{1}{T}$$

and

$$\left. \frac{\partial F}{\partial N} \right|_{E, N} = -\mu - T \left. \frac{\partial S}{\partial N} \right|_{E, V} = 0 \quad \Leftrightarrow \quad \left. \frac{\partial S}{\partial N} \right|_{E, V} = -\frac{\mu}{T}.$$

Since $\partial S/\partial E = k_B \tilde{\beta} = 1/\tilde{T}$ and $\partial S/\partial N = -k_B \tilde{\beta} \tilde{\mu} = -(\tilde{\mu}/\tilde{T})$, we conclude that the minimum is achieved when $\tilde{T} = T$ and $\tilde{\mu} = \mu$. Replacing these optimal values on Eq. (5.53) proves the theorem.

We have thus reached the important conclusion that any process occurring spontaneously at a given temperature and chemical potential will evolve in order to minimize the grand canonical potential $\Phi[\hat{\rho}]$. This result has been obtained by maximizing the entropy under the constraint of fixed $\langle \hat{H} \rangle$, $\langle \hat{N} \rangle$ and V . Conversely, the fact that the equilibrium state at constant T , μ and V is achieved by minimizing the grand canonical potential $\Phi[\hat{\rho}]$ leads to the maximum entropy principle. It is easy to see that the same minimal property discussed here in the case of $\Phi[\hat{\rho}]$, also holds for the Helmholtz free energy $F[\hat{\rho}]$, regarded as a functional of $\hat{\rho}$, keeping the temperature, particle number and volume constant.

6 Thermodynamic relations

In the last chapter we have introduced a variety of thermodynamic properties (thermodynamic potentials, heat capacities, compressibilities, etc.) which allow us to describe different aspects of macroscopic equilibrium states, as well as the processes connecting them. Taking into account that there are far more properties of physical interest than the few extensive or intensive magnitudes needed for characterizing the equilibrium state, it is most reasonable to attempt to establish general relations between them. Moreover, we know that the thermodynamic properties fulfill very specific conditions, for example, the extensiveness of the potentials or the intensiveness of pressure, temperature and chemical potential. Establishing the relations among the thermodynamic properties is important in order to derive the values of new properties from a smaller set of known ones. Furthermore, the relations can also be used to test the validity of models and approximations of the microscopic theory of matter. It is the purpose of this section to derive a number of thermodynamic relations and to discuss the most important methods used to derive them.

6.1 Duhem-Gibbs relations for thermodynamic potentials

The first and more fundamental interdependences of thermodynamic properties are the Duhem-Gibbs relations among the thermodynamic potentials. They follow from the extensive character of the potentials Φ_r as a function of their additive natural variables f_i , from the corresponding Euler conditions $\Phi_r = \sum_i \frac{\partial \Phi_r}{\partial f_i} f_i$, or from the Legendre transformations relating the potentials. Examples of this kind are

$$\begin{aligned} E &= TS - pV + \mu N, \\ F &= E - TS = \mu N - pV, \end{aligned}$$

and

$$\Phi = E - TS - \mu N = -pV.$$

6.2 Intensive nature of the derivatives of extensive properties

Another source of thermodynamic relations is the intensive character of properties such as p , μ and T . For example, if we consider the pressure $p = p(T, V, N)$ in a canonical ensemble, we know that

$$p(T, V, N) = p(T, \alpha V, \alpha N). \quad (6.1)$$

Taking the derivative with respect to α we have

$$0 = \left. \frac{\partial p}{\partial V} \right|_{T, N} V + \left. \frac{\partial p}{\partial N} \right|_{T, V} N, \quad (6.2)$$

which allows us to exchange $\partial p/\partial V$ by $\partial p/\partial N$.

The same reasoning can be applied to other intensive properties. For instance, in the case of the chemical potential

$$\mu(T, V, N) = \mu(T, \alpha V, \alpha N), \quad (6.3)$$

we obtain

$$0 = \left. \frac{\partial \mu}{\partial V} \right|_{T, N} V + \left. \frac{\partial \mu}{\partial N} \right|_{T, V} N, \quad (6.4)$$

with the help of which we can exchange $\partial \mu/\partial V$ by $\partial \mu/\partial N$.

6.3 Integrability of the differential of thermodynamic potentials

We know that the thermodynamic potentials are state functions, whose values do not depend on the previous history of the system. Therefore, the change in a state function f (typically a thermodynamic potential) between any two states 1 and 2 is independent of the details of the process involved in going from 1 to 2. Mathematically, this means that

$$f_2 - f_1 = \int_1^2 df$$

is independent of the integration path, which is usually stated by requiring that

$$\oint df = 0$$

for any closed path. Writing the differential form of f as

$$df = f_x dx + f_y dy,$$

the integrability condition, also known as Schwarz theorem, reads

$$\frac{\partial f_x}{\partial y} = \frac{\partial f_y}{\partial x}$$

or

$$\frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right) = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right). \quad (6.5)$$

Once applied to the thermodynamic potentials, this simple condition on the partial derivatives becomes a very useful tool for revealing important connections between the rates of change of different thermodynamic properties. The identities derived in this way are usually known as Maxwell relations.

As a first example let us consider the differential of the Helmholtz free energy

$$dF = -SdT - pdV + \mu dN$$

from which we conclude that

$$\left. \frac{\partial S}{\partial V} \right|_{T,N} = \left. \frac{\partial p}{\partial T} \right|_{V,N}. \quad (6.6)$$

An analogous relation can be derived from the differential of the grand canonical potential

$$d\Phi = -S dT - p dV - N d\mu,$$

namely,

$$\left. \frac{\partial S}{\partial V} \right|_{T,\mu} = \left. \frac{\partial p}{\partial T} \right|_{V,\mu}. \quad (6.7)$$

The only difference between (6.6) and (6.7) is that the derivatives are now taken for a fixed μ instead of a fixed N .

There are three pairs of conjugated extensive-intensive variables: (S, T) , (V, p) and (N, μ) . By picking one variable from each pair, we can therefore construct 8 different thermodynamic potentials, which depend on the chosen variables. In each case there are three possible pairs of variables whose derivatives can be exchanged. In the examples above we considered the potentials $F = F(T, V, N)$ and $\Phi = \Phi(T, V, \mu)$ and the variables V and T . We have therefore $3 \times 8 = 24$ different thermodynamic relations of this kind. Actually, they can be reduced to 12, if we ignore which is the third variable that is kept fixed on both sides of the equality. Thus, instead of Eqs. (6.6) and (6.7), one could simply write

$$\left. \frac{\partial S}{\partial V} \right|_T = \left. \frac{\partial p}{\partial T} \right|_V$$

regardless of whether N or μ is kept fixed. The reader is encouraged to explore and derive some of these relations and to compare their results with those reported in the literature.

Some relations are more interesting than others. For instance, starting from

$$dH = TdS + Vdp + \mu dN,$$

one obtains the relation

$$\left. \frac{\partial T}{\partial p} \right|_{S,N} = \left. \frac{\partial V}{\partial S} \right|_{p,N}, \quad (6.8)$$

which seems hard to interpret physically *a priori*, although it already looks more appealing when expressed in terms of the inverse functions as

$$\left. \frac{\partial p}{\partial T} \right|_{S,N} = \left. \frac{\partial S}{\partial V} \right|_{p,N}. \quad (6.9)$$

An interesting relation is obtained by considering the free enthalpy

$$dG = -S dT + V dp + \mu dN.$$

In this case we obtain

$$\left. \frac{\partial S}{\partial p} \right|_T = - \left. \frac{\partial V}{\partial T} \right|_p = -\alpha V, \quad (6.10)$$

where $\alpha = \frac{1}{V} \left. \frac{\partial V}{\partial T} \right|_p$ is the thermal expansion coefficient. In this way we relate the isothermal pressure dependence of S to α .

From the differential $dF = -SdT - pdV + \mu dN$ follows

$$\left. \frac{\partial \mu}{\partial V} \right|_{N,T} = - \left. \frac{\partial p}{\partial N} \right|_{V,T}. \quad (6.11)$$

The right hand side can be transformed by replacing $\partial p/\partial N$ by $\partial p/\partial V$ using (6.2) [i.e., $N\partial p/\partial N + V\partial p/\partial V = 0$]. Thus we have

$$\begin{aligned} \left. \frac{\partial \mu}{\partial V} \right|_{N,T} &= \left. \frac{\partial p}{\partial V} \right|_{N,T} \frac{V}{N} \\ &= \frac{1}{N} \frac{V}{\left. \frac{\partial p}{\partial V} \right|_{N,T}} \\ &= - \frac{1}{N} \frac{1}{\kappa_{T,N}}. \end{aligned} \quad (6.12)$$

Since μ is also an intensive property, we can replace $\partial\mu/\partial V$ by $\partial\mu/\partial N$ and relate the latter to the charge compressibility κ_c as

$$\left. \frac{\partial \mu}{\partial V} \right|_{N,T} = - \left. \frac{\partial \mu}{\partial N} \right|_{V,T} \frac{N}{V} = - \frac{1}{V} \frac{N}{\left. \frac{\partial N}{\partial \mu} \right|_{V,T}} = - \frac{1}{V} \frac{1}{\kappa_c}. \quad (6.13)$$

Finally, combining (6.12) and (6.13), we conclude that

$$\kappa_{T,N} = \frac{V}{N} \kappa_c = \frac{V}{N^2} \left. \frac{\partial N}{\partial \mu} \right|_{V,T}, \quad (6.14)$$

which relates the volume compressibility to the charge compressibility, and allows us to determine $\kappa_{T,N}$ in the framework of the grand canonical ensemble, where the natural variables are (T, V, μ) .

One can also derive interesting relations for the derivatives of the thermodynamic potentials with respect to variables that are not the natural ones or by keeping those variables constant. As an example, let us consider the energy

$$E = TS - pV + \mu N,$$

with its differential

$$dE = T dS - p dV + \mu dN,$$

and look for $\partial E/\partial V$ keeping the temperature and particle number constant. We may then write

$$\begin{aligned}\frac{\partial E}{\partial V}\Big|_{T,N} &= T\frac{\partial S}{\partial V}\Big|_{T,N} - p \\ &= T\frac{\partial p}{\partial T}\Big|_{V,N} - p.\end{aligned}\tag{6.15}$$

In the first step we have used that a change in volume V at constant T and N causes an energy change $dE = TdS - pdV$, provided that dS also refers to the entropy change at constant T and N . In the second step we have used the condition on the crossed derivatives in $dF = -SdT + \mu dN - pdV$.

Alternatively, we may start from the Duhem-Gibbs relation $E = TS + \mu N - pV$ and take the derivative with respect to V straightforwardly:

$$\frac{\partial E}{\partial V}\Big|_{T,N} = T\frac{\partial S}{\partial V}\Big|_{T,N} - \frac{\partial p}{\partial V}\Big|_{T,N} V - p + \frac{\partial \mu}{\partial V}\Big|_{T,N} N,\tag{6.16}$$

which surely looks more complicated at first. However, from dF and by changing $\partial p/\partial N$ into $\partial p/\partial V$ we have

$$\frac{\partial \mu}{\partial V}\Big|_{T,N} = -\frac{\partial p}{\partial N}\Big|_{T,V} = \frac{\partial p}{\partial V}\Big|_{T,N} \frac{V}{N}.$$

Replacing in (6.16) brings us back to Eq. (6.15), as it should. Many other relations can be derived in a similar way (e.g., for $\frac{\partial F}{\partial V}\Big|_{S,N}$).

6.4 Jacobi-determinant manipulations

A large number of relations in thermodynamics are obtained by changing an extensive variable (e.g., V) by its conjugated intensive variable (e.g., p) in partial derivatives. This concerns both exchanging derivatives with respect to these variable or changing the variables which are kept fixed, for example, from $\frac{\partial E}{\partial V}\Big|_{T,N}$ to $\frac{\partial E}{\partial p}\Big|_{T,N}$, or from $\frac{\partial E}{\partial T}\Big|_{V,N}$ to $\frac{\partial E}{\partial T}\Big|_{p,N}$. This kind of manipulations can be performed in a systematic way by using the properties of the 2×2 Jacobi determinant.

Let us first recall the definition and some useful properties of the Jacobian. Given two functions of the variables u and v , namely, $f, g : \mathbb{R}^2 \rightarrow \mathbb{R}$, $f = f(u, v)$ and $g = g(u, v)$, the Jacobi determinant is defined by

$$\frac{\partial(f, g)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{vmatrix} = \frac{\partial f}{\partial u} \frac{\partial g}{\partial v} - \frac{\partial g}{\partial u} \frac{\partial f}{\partial v}.\tag{6.17}$$

The change of sign of the determinant when two columns or rows are interchanged implies

$$\frac{\partial(f, g)}{\partial(u, v)} = -\frac{\partial(g, f)}{\partial(u, v)} = -\frac{\partial(f, g)}{\partial(v, u)} = \frac{\partial(g, f)}{\partial(v, u)}.\tag{6.18}$$

In addition we have

$$\frac{\partial(f, v)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ 0 & 1 \end{vmatrix} = \frac{\partial f}{\partial u} \Big|_v = \frac{\partial(v, f)}{\partial(v, u)}. \quad (6.19)$$

If the variables u and v are themselves functions $u = u(x, y)$ and $v(x, y)$ of other variables x and y , we can apply the usual chain rule for the partial derivatives:

$$\begin{aligned} \frac{\partial f}{\partial x} &= \frac{\partial f}{\partial u} \frac{\partial u}{\partial x} + \frac{\partial f}{\partial v} \frac{\partial v}{\partial x}, \\ \frac{\partial f}{\partial y} &= \frac{\partial f}{\partial u} \frac{\partial u}{\partial y} + \frac{\partial f}{\partial v} \frac{\partial v}{\partial y} \end{aligned}$$

and similarly for g . These relations can be written in a compact matrix form as

$$\begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{pmatrix} \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}.$$

Taking determinants on both sides we have

$$\frac{\partial(f, g)}{\partial(x, y)} = \frac{\partial(f, g)}{\partial(u, v)} \cdot \frac{\partial(u, v)}{\partial(x, y)}. \quad (6.20)$$

If the mapping $(u, v) \rightarrow (f, g)$ is bijective, we can invert it, i.e., solve the equations $f = f(u, v)$ and $g = g(u, v)$ for u and v . The functions $u(f, g)$ and $v(f, g)$ obtained in this way satisfy

$$f(u(f', g'), v(f', g')) = f'$$

and

$$g(u(f', g'), v(f', g')) = g'.$$

We can therefore write

$$1 = \frac{\partial(f, g)}{\partial(f, g)} = \frac{\partial(f, g)}{\partial(u, v)} \cdot \frac{\partial(u, v)}{\partial(f, g)}. \quad (6.21)$$

As a first application of the Jacobi determinant method we consider $\frac{\partial p}{\partial T} \Big|_V$ and try to relate it with other properties by changing V by p . We thus write

$$\begin{aligned} \frac{\partial p}{\partial T} \Big|_V &= \frac{\partial(p, V)}{\partial(T, V)} = \frac{\partial(p, V)}{\partial(p, T)} \frac{\partial(p, T)}{\partial(T, V)} \\ &= -\frac{\partial(p, V)}{\partial(p, T)} \frac{\partial(p, T)}{\partial(V, T)} \\ &= -\frac{\partial V}{\partial T} \Big|_p \frac{\partial p}{\partial V} \Big|_T \\ &= \frac{V\alpha}{-\frac{\partial V}{\partial p} \Big|_T} = \frac{\alpha}{\kappa_T}. \end{aligned} \quad (6.22)$$

As a physically transparent byproduct, notice that the sign of $\frac{\partial p}{\partial T}\big|_V$ is the same as the sign of the thermal expansion coefficient α .

A further interesting application of the Jacobi determinant method is to establish the relation between the heat capacities C_p and C_V by changing p by V as the variable that is kept constant:

$$C_p = T \frac{\partial S}{\partial T}\bigg|_p = T \frac{\partial(S, p)}{\partial(T, p)} = T \frac{\partial(S, p)}{\partial(S, V)} \frac{\partial(S, V)}{\partial(T, p)},$$

where we have replaced p by V , keeping S constant [$\partial(T, p) \rightarrow \partial(S, V)$]. On the right-hand side we recognize the adiabatic compressibility κ_S since

$$\frac{\partial(S, p)}{\partial(S, V)} = \frac{1}{\frac{\partial V}{\partial p}\big|_S} = -\frac{1}{V\kappa_S}.$$

In a second step we intend to find C_V by replacing $\partial(T, p)$ by $\partial(T, V)$ in the denominator. In this way we obtain

$$C_p = T \left(-\frac{1}{V\kappa_S} \right) \frac{\partial(S, V)}{\partial(T, V)} \frac{\partial(T, V)}{\partial(T, p)}.$$

Noting that

$$T \frac{\partial(S, V)}{\partial(T, V)} = T \frac{\partial S}{\partial T}\bigg|_V = C_V \quad \text{and} \quad \frac{\partial(T, V)}{\partial(T, p)} = \frac{\partial V}{\partial p}\bigg|_T = -V\kappa_T,$$

we have

$$C_p = C_V \frac{\kappa_T}{\kappa_S} \quad \text{or} \quad \frac{C_p}{C_V} = \frac{\kappa_T}{\kappa_S}. \quad (6.23)$$

There are also some Jacobi determinants which take a simple closed form in terms of known thermodynamic properties and which can be used in order to establish new relations. For example:

$$\frac{\partial(S, V)}{\partial(T, p)} = \begin{vmatrix} \frac{\partial S}{\partial T}\big|_p & \frac{\partial S}{\partial p}\big|_T \\ \frac{\partial V}{\partial T}\big|_p & \frac{\partial V}{\partial p}\big|_T \end{vmatrix} = \begin{vmatrix} C_p \frac{1}{T} & -\alpha V \\ V\alpha & -V\kappa_T \end{vmatrix} = -\frac{V\kappa_T}{T} C_p + \alpha^2 V^2, \quad (6.24)$$

where we have used that

$$dG = -SdT + Vdp - \mu dN \quad \Rightarrow \quad \frac{\partial S}{\partial p}\bigg|_T = -\frac{\partial V}{\partial T}\bigg|_p = -\alpha V.$$

Using the relation (6.24) we can rewrite C_V as

$$\begin{aligned} C_V &= T \frac{\partial S}{\partial T}\bigg|_V = T \frac{\partial(S, V)}{\partial(T, V)} = T \frac{\partial(S, V)}{\partial(T, p)} \frac{\partial(T, p)}{\partial(T, V)} \\ &= (-V\kappa_T C_p + T\alpha^2 V^2) \left(-\frac{1}{V\kappa_T} \right) = C_p - T \frac{\alpha^2 V}{\kappa_T}, \end{aligned}$$

and obtain the important relation

$$C_p - C_V = T \frac{\alpha^2 V}{\kappa_T} \geq 0. \quad (6.25)$$

Combining (6.23) and (6.25) we have

$$1 - \frac{C_V}{C_p} = 1 - \frac{\kappa_S}{\kappa_T} = T \frac{\alpha^2 V}{\kappa_T C_p}$$

and

$$\kappa_T - \kappa_S = T \frac{\alpha^2 V}{C_p} \geq 0. \quad (6.26)$$

These relations show that only three of the five quantities C_V , C_p , κ_T , κ_S and α are actually independent. The present analysis is just one example of how thermodynamic relations can be used in order to derive results on new properties and to gain insight into the correlations between various macroscopic observables, which can be verified by comparing independent measurements. As a byproduct we should also mention the inequalities

$$C_p \geq C_V \geq 0 \quad (6.27)$$

and

$$\kappa_T \geq \kappa_S \geq 0, \quad (6.28)$$

which can be easily understood physically.

The number of thermodynamic relation that one may conceive is almost inexhaustible. We shall not pursue with further examples here, but rather derive those needed in the context of the specific problems to be discussed below.

Exercise 6.14: Starting from the definition of κ_S show that $\kappa_S = \frac{C_V}{C_p} \kappa_T$ by using simple Jacobi transformations.

6.5 Measuring the absolute temperature scale

In the thermodynamic relations discussed in the previous section the absolute temperature scale T appears explicitly, either as the parameter which is varied, or which is kept constant, or as a multiplicative factor. This means that the derived equations would not hold if T were replaced by an arbitrary empirical temperature scale. One may therefore use these relations in order to determine the absolute temperature $T = T(\theta)$ as a function of an empirical temperature θ satisfying a one-to-one correspondence with T . For example, θ can be the voltage across a thermocouple, the length of a Hg filament, and the like.

In order to illustrate this remarkable possibility, we consider three related properties, namely, the Joule-Thompson coefficient μ_{JT} , the heat capacity C_p and the expansion coefficient α . In each case we distinguish between the true exact values of the properties, which are denoted by μ_{JT} , C_p and α , and the values obtained using an empirical (laboratory specific) temperature θ , which are denoted by $\tilde{\mu}_{JT}$, \tilde{C}_p and $\tilde{\alpha}$, respectively.

i) The Joule-Thompson coefficient

$$\mu_{JT} = \left. \frac{\partial T}{\partial p} \right|_H = \left. \frac{dT}{d\theta} \frac{\partial \theta}{\partial p} \right|_H = \tilde{\mu}_{JT} \frac{dT}{d\theta} \quad (6.29)$$

measures the temperature change following an isenthalpic pressure change. On the left hand side of (6.29) we have the true Joule-Thompson coefficient while on the right hand side we have the measured one according to $\theta = \theta(T)$.

ii) The heat capacity at constant pressure is given by

$$C_p = \left. \frac{\partial H}{\partial T} \right|_p = \left. \frac{\partial H}{\partial \theta} \right|_p \frac{d\theta}{dT} = \tilde{C}_p \frac{1}{\frac{dT}{d\theta}}. \quad (6.30)$$

iii) Finally, the expansion coefficient is given by

$$\alpha = \frac{1}{V} \left. \frac{\partial V}{\partial T} \right|_p = \frac{1}{V} \left. \frac{\partial V}{\partial \theta} \right|_p \frac{d\theta}{dT} = \tilde{\alpha} \frac{1}{\frac{dT}{d\theta}}. \quad (6.31)$$

We would like to express μ_{JT} in terms of known quantities including in particular C_p and α :

$$\mu_{JT} = \left. \frac{\partial T}{\partial p} \right|_H = \frac{\partial(T, H)}{\partial(p, H)} = \frac{\partial(T, H)}{\partial(T, p)} \underbrace{\frac{\partial(T, p)}{\partial(p, H)}}_{-1/C_p} = -\frac{\left. \frac{\partial H}{\partial p} \right|_T}{C_p}. \quad (6.32)$$

In order to replace $\left. \frac{\partial H}{\partial p} \right|_T$ by involving α we use two different Maxwell relations. From $dH = TdS + Vdp + \mu dN$ we have

$$\left. \frac{\partial H}{\partial p} \right|_{T, N} = T \left. \frac{\partial S}{\partial p} \right|_T + V, \quad (6.33)$$

and from $dG = -SdT + Vdp + \mu dN$ we have

$$-\left. \frac{\partial S}{\partial p} \right|_T = \left. \frac{\partial V}{\partial T} \right|_p. \quad (6.34)$$

Thus, we have

$$\left. \frac{\partial H}{\partial p} \right|_T = T \left. \frac{\partial S}{\partial p} \right|_T + V = V - T \left. \frac{\partial V}{\partial T} \right|_p = V(1 - T\alpha)$$

and therefore

$$\mu_{JT} = \frac{V}{C_p} (T\alpha - 1). \quad (6.35)$$

Replacing (6.29), (6.30) and (6.31) in (6.35) we finally obtain

$$\begin{aligned}\tilde{\mu}_{JT} \frac{dT}{d\theta} &= \frac{V}{\tilde{C}_p} \frac{dT}{d\theta} \left(T \frac{d\tilde{\alpha}}{dT} - 1 \right) \\ \left(\tilde{\mu}_{JT} + \frac{V}{\tilde{C}_p} \right) \frac{dT}{d\theta} &= \frac{VT\tilde{\alpha}}{\tilde{C}_p} \\ \frac{d \ln T}{d\theta} &= \frac{V\tilde{\alpha}}{\tilde{\mu}_{JT}\tilde{C}_p + V}.\end{aligned}$$

We conclude that the measurement of $\tilde{\mu}_{JT}$, \tilde{C}_p and $\tilde{\alpha}$ at a volume V , by using an arbitrary temperature scale and an arbitrary substance, allows us to determine the absolute temperature except for a multiplicative constant (i.e., an additive constant in $\ln T$). This constant can only be fixed by a convention, which defines the unit of the absolute temperature. In this context it is important to notice that all thermodynamic relations are invariant with respect to a uniform scaling transformation $T' = aT$, where a is a constant. The adopted convention is such that 1 K is the temperature of the triple point of water divided by 273.16, or equivalently, that the triple point of water is 273.16 K. Since in the Celsius scale the triple-point temperature is 0.01 °C, the absolute zero corresponds to -273.15 °C.

7 Thermodynamic processes

Thermodynamic processes are present in our everyday life, for example, in a number of processes in the atmosphere and in many technological applications (internal combustion engines, refrigerators, etc.). In many cases the process is cyclic. One has a working substance (typically a gas) which undergoes a cyclic process repeatedly. Very often, the substance is replaced by a new initial substance at the beginning of the cycle (e.g., the vaporized water in a steam engine or the fuel in an internal combustion engine). Since the principles of cyclic processes are not affected by this, we shall ignore the replacement and assume that the particle number N is constant. Thus, two thermodynamic variables [e.g., (p, V) or (S, T)] are enough to characterize the equilibrium state of the substance. We shall further assume that the processes are quasi static, so that equilibrium can be taken for granted. Irreversible processes are discussed in a forthcoming section.

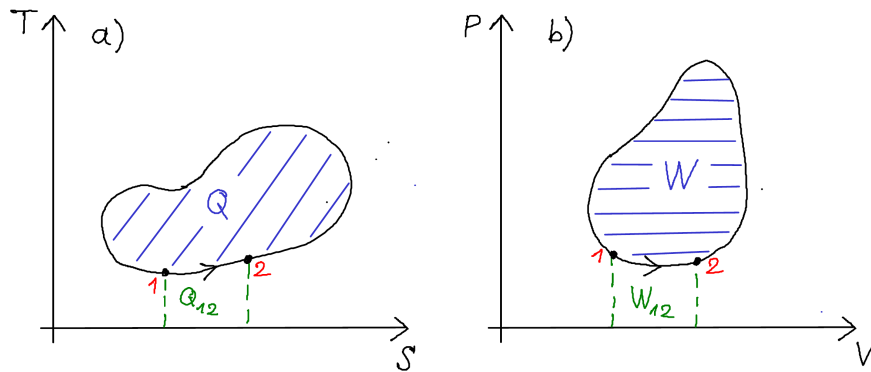


Figure 1: (a) Representation of a cyclic process in a TS -diagram. The area Q_{12} under the $T(S)$ curve gives the absorbed heat during the process from state 1 to state 2, while the enclosed area Q is the total heat absorbed in the cycle. (b) Representation of a cyclic process in a pV -diagram. The area W_{12} under the curve (the enclosed area W) is the work delivered into the environment during the process from state 1 to state 2 (during the cycle).

It is useful to represent the process in both a (T, S) and a (p, V) diagrams, since the area TdS and pdV under the process curves have a clear physical meaning (see Fig. 1). The area under the $T = T(S)$ curve represents the heat

$$Q_{12} = \int_1^2 T dS \quad (7.1)$$

absorbed by the system. In a cyclic process the heat absorbed in a cycle is

$$Q = \oint T dS \quad (7.2)$$

The area under the $p = p(V)$ curve is the work done by the system, i.e., the work

$$W_{12} = \int_1^2 p dV \quad (7.3)$$

delivered by the system into the environment. In a cycle we have

$$W = \oint p dV \quad (7.4)$$

Since the energy is a state function and $dE = TdS - pdV$ we have

$$\oint dE = E \quad (7.5)$$

and

$$Q - W = 0. \quad (7.6)$$

Thus, with the appropriate scale, the areas enclosed by the cycles in the (T, S) and (p, V) diagrams are the same. See Eqs. (7.2), (7.4) and (7.6). Moreover, since the signs of Q and W are the same, the cycles are run in the same sense in both diagrams. Counterclockwise processes are also known as left processes and clockwise processes as right processes. Clockwise processes have $Q = W > 0$. A net amount of heat is absorbed and work is delivered. These are working machines. Counterclockwise processes have $Q = W < 0$. Work is absorbed and heat is delivered. These are refrigerators or heat pumps from lower to higher temperatures.

The cyclic processes can be usually subdivided in simpler subprocesses, in which a given thermodynamic property is kept constant. One distinguishes in particular isothermal, isobaric, isochoric and isentropic processes. The latter are also known as adiabatic. These can be illustrated as follows:

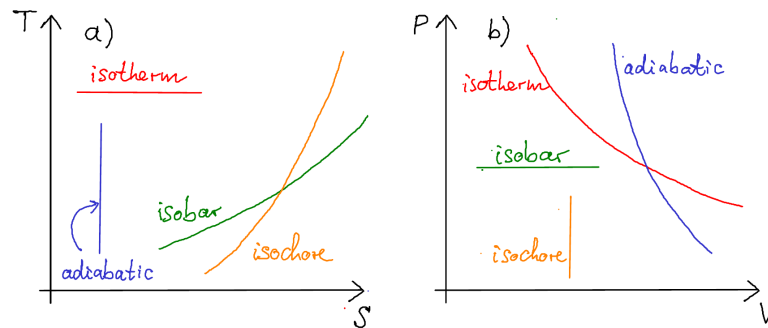


Figure 2: Adiabatic ($\delta Q = 0$), isothermal (constant temperature T), isobaric (constant pressure p) and isochoric (constant volume V) subprocesses illustrated in the a) TS -diagram and b) pV -diagram.

The different slopes of the isothermal and isentropic curves in the (p, V) diagram, and of the isochoric and isobaric curves in the (T, S) diagram can be inferred from the relations $\kappa_T \geq \kappa_S \geq 0$ and $C_p \geq C_V \geq 0$ [see Eqs. (6.25) and (6.26)].

7.1 Carnot cycle

The most important cyclic process is Carnot's cycle, which is obtained by alternating isothermal and adiabatic subprocesses:

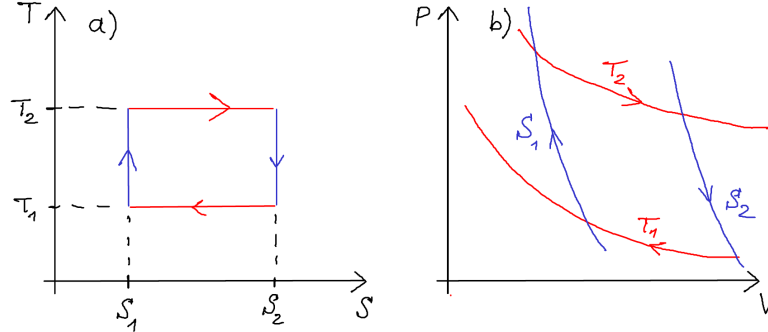


Figure 3: Illustration of the Carnot's cycle in the a) TS -diagram and b) pV -diagram. The cycle consists of alternating isothermal and adiabatic subprocesses. During the isothermal subprocesses, the system is in contact with heat baths at temperatures T_1 and $T_2 > T_1$. During the adiabatic subprocesses, S_1 and $S_2 > S_1$ denote the constant-kept entropies.

All the heat exchanges occur at constant temperature. As clockwise process the system absorbs the heat

$$Q_2 = T_2 (S_2 - S_1) > 0 \quad (7.7)$$

from the higher temperature T_2 , and rejects a smaller amount of heat

$$Q_1 = T_1 (S_2 - S_1) > 0 \quad (7.8)$$

to the cooler reservoir at T_1 . The total heat exchanged in a cycle is

$$Q = Q_2 - Q_1 = W > 0. \quad (7.9)$$

The work efficiency or energy-conversion efficiency of a machine is defined by

$$\eta^W = \frac{W}{Q_2} = \frac{\text{Work done}}{\text{Heat absorbed}}. \quad (7.10)$$

In a Carnot cycle this takes the value

$$\eta_{\text{Carnot}}^W = \frac{T_2 - T_1}{T_2} < 1, \quad (7.11)$$

where $T_2 > T_1$.

It is easy to see, using that $\oint \frac{\delta Q}{T} \leq 0$, that no machine involving heat baths in the temperature range $[T_1, T_2]$ can exceed the work efficiency of the Carnot cycle. We consider

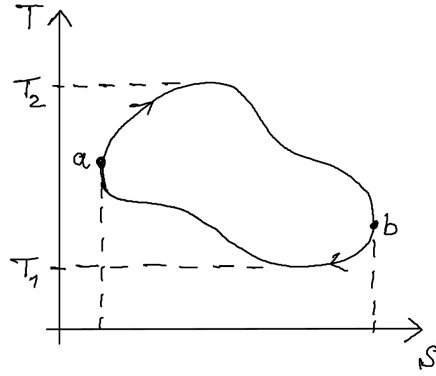


Figure 4: Illustration of a cyclic process involving a succession of heat baths in the temperature range $[T_1, T_2]$. During the process from state a to state b , the system absorbs heat from the baths ($\delta Q = TdS \geq 0$), while during the process from b to a , heat is delivered to the environment ($\delta Q = TdS \leq 0$).

an arbitrary cycle in which the system exchanges heat δQ with thermal baths at various temperatures $T(S)$. Let T_2 be the largest value of T along the cycle and T_1 the smallest one (see Fig. 4). We intend to compare η^W for such an arbitrary cycle with the efficiency of the Carnot cycle running between the largest and the smallest temperatures. The total work and heat are

$$W = Q = \oint \delta Q = \underbrace{\int_a^b \delta Q}_{Q_2 > 0} + \underbrace{\int_b^a \delta Q}_{-Q_1 < 0} = Q_2 - Q_1. \quad (7.12)$$

We consider Clausius inequality

$$\oint \frac{\delta Q}{T} \leq 0, \quad (7.13)$$

where T refers to the temperature of the thermal bath with which the heat exchange δQ occurs:

$$0 \geq \oint \frac{\delta Q}{T} = \underbrace{\int_a^b \frac{\delta Q}{T}}_{\delta Q > 0} + \underbrace{\int_b^a \frac{\delta Q}{T}}_{\delta Q < 0} \quad (7.14)$$

Since $T_2 \geq T \geq T_1$ we replace T by T_2 for all the $\delta Q > 0$ and T by T_1 for all the $\delta Q < 0$. We may thus write

$$0 \geq \underbrace{\int \frac{\delta Q}{T}}_{\delta Q > 0} + \underbrace{\int \frac{\delta Q}{T}}_{\delta Q < 0} \geq \int \frac{\delta Q}{T_2} + \int \frac{\delta Q}{T_1} = \frac{Q_2}{T_2} - \frac{Q_1}{T_1},$$

which implies

$$\frac{Q_1}{T_1} \geq \frac{Q_2}{T_2} \quad \Rightarrow \quad \frac{Q_1}{Q_2} \geq \frac{T_1}{T_2}. \quad (7.15)$$

The work efficiency η^W of the arbitrary cycle satisfies

$$\eta^W = \frac{W}{Q_2} = \frac{Q_2 - Q_1}{Q_2} = 1 - \frac{Q_1}{Q_2} \leq 1 - \frac{T_1}{T_2} = \eta_{\text{Carnot}}^W. \quad (7.16)$$

The equal sign holds only when the process is reversible (i.e., $\oint \frac{\delta Q}{T} = 0$) and at the same time all the heat absorbed ($\delta Q > 0$) comes from the highest temperature bath (T_2) and all the rejected heat ($\delta Q < 0$) goes to the coolest bath (T_1). In other words, only for the Carnot cycle running between the temperatures T_1 and T_2 .

As a counterclockwise process the Carnot cycle receives the work $-W > 0$ done by external forces, it absorbs the heat $Q_1 > 0$ from the reservoir at the lower temperature T_1 , and transfers the absorbed energy $\Delta E = -Q_2 = -W - Q_1 < 0$ in the form of heat to the hottest reservoir at T_2 . One defines the heat-transfer efficiency of an arbitrary heat pump as

$$\eta^H = \frac{\text{Rejected heat}}{\text{Absorbed work}} = \frac{-Q_2}{-W} = \frac{1}{\eta^W}. \quad (7.17)$$

In the Carnot cycle we have

$$\eta_{\text{Carnot}}^H = \frac{T_2}{T_2 - T_1} = \frac{1}{\eta_{\text{Carnot}}^W} > 1. \quad (7.18)$$

Notice that the efficiency of a Carnot cycle is much better than the plain dissipation of work, as one would have in a Joule experiment, for which $Q_2 = W$ and thus $\eta_{\text{Joule}}^H = 1$.

It is also interesting to consider the cooling efficiency of refrigerators, which is defined as

$$\eta^C = \frac{\text{Heat removed}}{\text{Absorbed Work}} = \frac{Q_1}{-W}. \quad (7.19)$$

For the Carnot cycle we have

$$\eta_{\text{Carnot}}^C = \frac{T_1}{T_2 - T_1}, \quad (7.20)$$

which shows that the efficiency improves when $\Delta T = T_2 - T_1$ is small. Therefore, cooling processes are usually split in a number of subprocesses with smaller ΔT 's.

Before closing this section we would like to mention the Stirling cycle, which alternates isothermal and isochoric processes, the Ericsson cycle, which alternates isobaric and isothermal processes, and the Rankine cycle (steam engine) which alternates adiabatic and isobaric processes.

7.2 Joule Thompson process

In the following sections we would like to discuss two important irreversible processes. The first one is the Joule-Thompson effect which consists in an adiabatic ($\delta Q = 0$) expansion at constant pressure. More precisely, we consider a gas or liquid at a pressure

p_1 in a volume V_1 which flows into a region of lower pressure p_2 through a small opening or porous material. During the process the pressures p_1 and p_2 are kept constant from the exterior and the volume change is so slow that the system is constantly at thermodynamic equilibrium. Initially, before the valve is open, we have $V_1 = V_1^0$ and $V_2 = 0$ and in the final state we have $V_1 = 0$ and $V_2 = V_2^\infty$.

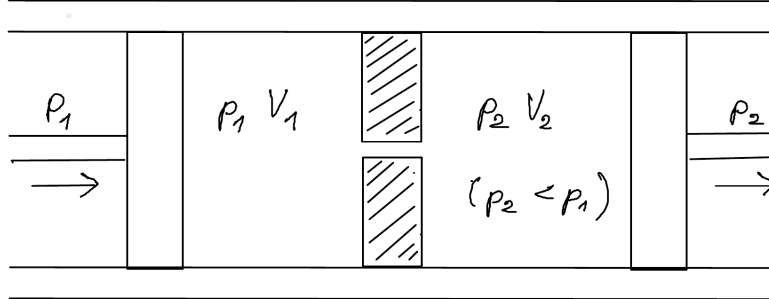


Figure 5: Scheme of the Joule-Thomson effect. A gas or liquid flows from a region at constant pressure p_1 into a region of lower pressure p_2 through a small opening.

It is easy to show that this is an isoenthalpic process ($dH = 0$). Since the system is thermally isolated, we have $\delta Q = 0$ and therefore

$$dE = \delta W = -p_1 dV_1 - p_2 dV_2. \quad (7.21)$$

The extensiveness of E implies $E = E_1 + E_2$ and thus

$$dE_1 + dE_2 + p_1 dV_1 + p_2 dV_2 = 0 \quad (7.22)$$

or

$$dH_1 + dH_2 = dH = 0. \quad (7.23)$$

The enthalpy is conserved throughout the process. Note that the change in the number of particles of each subsystem $dN_1 = -dN_2 < 0$ does not contribute to dE or dH , since the system is in equilibrium. Therefore, $\mu_1 = \mu_2 = \mu$ and $\mu dN_1 + \mu dN_2 = \mu dN = 0$. The total number of particles is conserved.

The change of temperature associated with a change of pressure at constant enthalpy H is given by the Joule-Thomson coefficient

$$\mu_{JT} = \left. \frac{\partial T}{\partial p} \right|_H = \frac{V}{C_p} (\alpha T - 1).$$

The total temperature change is then

$$T_2 - T_1 = \int_{p_1}^{p_2} \left. \frac{\partial T}{\partial p} \right|_H dp.$$

The observed temperature change is known as Joule-Thompson effect. In an ideal gas we have $\alpha = 1/T$, since the enthalpy depends only on the temperature. No temperature change can occur because $\Delta H = 0$. In real gases the Joule-Thompson coefficient changes sign as a function of T , becoming positive only at low temperatures, below the so-called inversion point. For $\mu_{JT} > 0$ we have a temperature drop upon pressure drop. Therefore, in order that the Joule-Thompson expansion becomes effective as cooler, the gas must already be below the inversion temperature where $\mu_{JT} = 0$.

Since $dH = 0$ and $dN = 0$, we have

$$dH = TdS + Vdp = 0, \quad (7.24)$$

which implies

$$\left. \frac{\partial S}{\partial p} \right|_{H,N} = -\frac{V}{T} < 0. \quad (7.25)$$

The process is obviously irreversible, since

$$\Delta S = -\frac{V}{T}\Delta p > 0 \quad \text{for} \quad \Delta p = p_2 - p_1 < 0. \quad (7.26)$$

Finally, we observe that any change of pressure occurring spontaneously is negative, as expected.

7.3 Adiabatic expansion in vacuum

A further important irreversible process is the adiabatic ($\delta Q = 0$) expansion into an evacuated chamber having pressure $p_2 = 0$. In order that the process takes place quasi-statically, and that the system remains always in equilibrium, we can device the system illustrated above, which involves a small valve or a porous media. Before opening the valve, we have $V = V_1$ and $p = p_1$. In the final state we have $V = V_2$. The external pressure is $p_2 \rightarrow 0$. In this case we have both $\delta Q = 0$ and $\delta W = 0$, which implies $dE = 0$. The energy of the gas is conserved. The temperature change can be calculated from

$$T_2 - T_1 = \int_{V_1}^{V_2} \left. \frac{\partial T}{\partial V} \right|_E dV.$$

The rate of temperature change at constant energy can be expressed in terms of known properties. Our starting point is

$$\left. \frac{\partial T}{\partial V} \right|_E = \frac{\partial(T, E)}{\partial(V, E)} = \frac{\partial(T, E)}{\partial(V, T)} \frac{\partial(V, T)}{\partial(V, E)} = -\frac{1}{C_V} \left. \frac{\partial E}{\partial V} \right|_T.$$

Moreover

$$dE = TdS - pdV \quad \Rightarrow \quad \left. \frac{\partial E}{\partial V} \right|_T = T \left. \frac{\partial S}{\partial V} \right|_T - p$$

and

$$dF = -SdT - pdV \quad \Rightarrow \quad \left. \frac{\partial S}{\partial V} \right|_T = \left. \frac{\partial p}{\partial T} \right|_V.$$

Combining these relations we obtain

$$\left. \frac{\partial E}{\partial V} \right|_T = T \left. \frac{\partial p}{\partial T} \right|_V - p$$

and finally

$$\left. \frac{\partial T}{\partial V} \right|_E = \frac{1}{C_V} \left[p - T \left. \frac{\partial p}{\partial T} \right|_V \right].$$

Again, for an ideal gas there is no effect, since at constant V , p depends linearly on T and thus $\left. \frac{\partial T}{\partial V} \right|_E = 0$. In principle, one can find both signs for $\left. \frac{\partial T}{\partial V} \right|_E$. However, in diluted real gases we always see a cooling. In fact, the van der Waals interactions between gas molecules are attractive at long distances. Consequently, increasing the volume implies an increase of the average interparticle distance. This raises the potential energy and by energy conservation lowers the kinetic energy. The equipartition theorem implies that the temperature must decrease.

Before closing this section, we may calculate the entropy change following an energy conserving expansion:

$$dE = TdS - pdV = 0 \quad \Rightarrow \quad \left. \frac{\partial S}{\partial V} \right|_{E,N} = \frac{p}{T} > 0$$

and

$$\Delta S = \frac{p}{T} \Delta V.$$

As expected, any spontaneous volume change into vacuum must be positive. The process is irreversible, since $\Delta S > 0$.

8 The Nernst theorem: The Third law of thermodynamics

In 1905 Walther Nernst postulated that the entropy of any system at absolute zero temperature is a universal constant, which may be taken to be zero. This postulate, which he actually formulated in other terms, was inferred from experiments at temperatures of the order of 1 K. For example, one observed that the changes in enthalpy H and in free enthalpy $G = H - TS$ in isothermal and isobaric processes become equal as $T \rightarrow 0$. Indeed, since $\Delta G = \Delta H - T\Delta S$ for constant T , $\Delta H = \Delta G$ implies $\Delta S = 0$. Moreover, since $\left.\frac{\partial G}{\partial T}\right|_p = -S$, the fact that $\left.\frac{\partial G}{\partial T}\right|_p \rightarrow 0$ for $T \rightarrow 0$ was used to infer that $S \rightarrow 0$ for $T \rightarrow 0$. This important fundamental property is general in two respects:

- i) It holds for any system and any substance, and
- ii) it holds irrespectively of the values of any other additive constants of motion $f_i = E, V$ or N or intensive properties $\alpha_i = p$ or μ , on which S usually depends.

Therefore, it is easy to imagine that the third law of thermodynamics has many far reaching consequences. In the following, we shall demonstrate this important result. Furthermore, we discuss the conditions for its validity from the point of view of statistical mechanics, which provides an unambiguous microscopic definition of S .

Let us consider the canonical density operator

$$\hat{\rho}_c = \frac{e^{-\beta\hat{H}}}{\text{Tr}\{e^{-\beta\hat{H}}\}} = \frac{\sum_{n=0}^{+\infty} e^{-\beta E_n} |n\rangle\langle n|}{\sum_{n=0}^{+\infty} e^{-\beta E_n}} = \frac{\hat{P}_0 + \sum_{\substack{n \\ E_n > E_0}} e^{-\beta(E_n - E_0)} |n\rangle\langle n|}{g + \sum_{\substack{n \\ E_n > E_0}} e^{-\beta(E_n - E_0)}},$$

where

$$\hat{P}_0 = \sum_{i=1}^g |i\rangle\langle i|$$

is the projector operator on the g -dimensional subspace spanned by the states having the lowest energy E_0 ($\hat{P}_0^2 = \hat{P}_0$). The degeneracy of the ground state is denoted by g , i.e., $E_i = E_0$ for $i = 1, \dots, g$. In the limit of $T \rightarrow 0$ (i.e., $\beta \rightarrow \infty$) we have

$$\hat{\rho}_c(T = 0) = \frac{1}{g} \hat{P}_0 = \frac{1}{g} \sum_{i=1}^g |i\rangle\langle i|,$$

which implies that the system is found in the ground-state subspace. In this case the entropy is given by

$$S_0 = -k_B \langle \ln \hat{\rho} \rangle = k_B \ln g.$$

Very often the ground state is non-degenerate (i.e., $g = 1$), in which case $S_0 = 0$ irrespectively of the other thermodynamic variables (e.g., V , N , p or μ). However, there are also many situations where $g > 1$. They deserve a more detailed consideration. For example, if a system has an odd number of electrons, its ground state has a non-vanishing half-integer total spin $s \geq 1/2$ which, combined with the orbital angular momentum l , yields

a half-integer total angular momentum $\vec{j} = \vec{s} + \vec{l}$. This implies a finite $2j + 1$ ground-state degeneracy as a consequence of rotational invariance. In general, time-inversion symmetry requires that the ground state is at least twofold degenerate when the number of electrons is odd. Although in these cases the ground-state entropy $S_0 = k_B \ln(2j + 1)$ or $S_0 = k_B \ln 2$ is finite, its value is negligible in comparison with the entropy of a macroscopic system at any finite T since the latter is an extensive property proportional to the system size. Apart from this, there are special situations where a higher ground state degeneracy $g \propto N$ is obtained. This is the case, for example, in a spin-rotational invariant electronic system having a ferromagnetic ground state, where the total spin angular momentum s and the degeneracy $g = 2s + 1$ grow linearly with the number of atoms N (e.g., $s \propto N$ in a saturated ferromagnetic solid). But even in these cases the ground-state entropy per particle $S_0/N \propto \ln N/N$ and the proportion $S_0/S(T)$ of S_0 relative to the entropy at any finite T tend to zero when $N \rightarrow \infty$. In addition, it should be noted that these cases are not truly relevant in practice. In fact, spin-orbit interactions introduce magnetic anisotropy which removes the $2s + 1$ spin-rotational degeneracy of ferromagnetic materials. Moreover, the total orbital momentum $\vec{j} = \vec{s} + \vec{l}$ is not conserved, since the solid is usually fixed to the laboratory frame. Only the time inversion symmetry and possibly some finite point-group symmetries are left, which yield a finite N -independent g . Therefore, we conclude that the additive or extensive part of the entropy is always zero as $T \rightarrow 0$, irrespectively of the values of all other thermodynamic quantities (V , N , p , μ , etc.).

8.1 Some experimental consequences

The third principle of thermodynamics has far reaching repercussions, which can be directly observed in experiments. From an historical perspective, the conclusions to be derived below are some of the observations that guided Nernst to infer and postulate his principle.

An immediate consequence of the third principle is that *any heat capacity vanishes at $T = 0$* . Consider some reversible path R connecting the $T = 0$ equilibrium state with some finite-temperature macrostate A whose entropy $S(A)$ we want to calculate. Since

$$\left. \frac{\partial S}{\partial T} \right|_R = \frac{C_R}{T}$$

we have

$$S(A) = \int_0^{T_A} \frac{C_R}{T} dT.$$

Since $S(A)$ is well defined and finite, $\frac{C_R}{T}$ needs to be integrable in $[0, T_A]$. This implies that

$$C_R \xrightarrow{T \rightarrow 0} 0.$$

In particular $C_V \rightarrow 0$ and $C_p \rightarrow 0$ for $T \rightarrow 0$, a result verified by all experiments done so far.

A further consequence of the third principle is that the thermal expansion coefficients α of all substances vanish for $T \rightarrow 0$. Since $S(T=0) = 0$ we can write

$$S(T) = \int_0^T \frac{C_p}{T'} dT',$$

where the integration path is isobaric ($\left. \frac{\partial S}{\partial T} \right|_p = \frac{C_p}{T}$). Moreover, we know that

$$V\alpha = \left. \frac{\partial V}{\partial T} \right|_p = - \left. \frac{\partial S}{\partial p} \right|_T = - \frac{\partial}{\partial p} \left(\int_0^T \frac{C_p}{T'} dT' \right) \Big|_T = - \int_0^T \frac{\partial C_p}{\partial p} \Big|_T \frac{dT'}{T'}. \quad (8.1)$$

Taking into account that

$$\frac{C_p}{T} = \left. \frac{\partial S}{\partial T} \right|_p \Rightarrow \frac{1}{T} \left. \frac{\partial C_p}{\partial p} \right|_T = \frac{\partial}{\partial T} \left(\left. \frac{\partial S}{\partial p} \right|_T \right) \Big|_p = - \frac{\partial}{\partial T} \left(\left. \frac{\partial V}{\partial T} \right|_p \right) = - \left. \frac{\partial^2 V}{\partial T^2} \right|_p.$$

Substituting in Eq. (8.1), we finally obtain

$$V\alpha = \int_0^T \left. \frac{\partial^2 V}{\partial T^2} \right|_p dT' = \left. \frac{\partial V}{\partial T} \right|_p(T) - \left. \frac{\partial V}{\partial T} \right|_p(0) \xrightarrow{T \rightarrow 0} 0.$$

This proves the statement. Moreover, from the previous relations (6.6), (6.7) and (6.22) we have

$$\frac{\alpha}{\kappa_T} = \left. \frac{\partial p}{\partial T} \right|_V = \left. \frac{\partial S}{\partial V} \right|_T \xrightarrow{T \rightarrow 0} 0.$$

In the above derivation it is irrelevant whether N or μ are kept constant. Thus, S is independent of V for $T \rightarrow 0$. A similar argument involving the charge compressibility κ_c shows that $\left. \frac{\partial S}{\partial N} \right|_T \xrightarrow{T \rightarrow 0} 0$ (see exercise).

On these thermodynamic grounds, we conclude that $S(T, V, N)$ is independent of V and N at $T = 0$. This result alone already allows us to choose the value of this universal constant equal to zero, even without any further microscopic explanation.

Exercise 8.15: Show that $\left. \frac{\partial S}{\partial N} \right|_T = -\alpha/\kappa_c$. Consider $\left. \frac{\partial \mu}{\partial T} \right|_{V,N}$ and work it out using the Jacobi determinant method as in Eq. (6.22). Conclude that S is independent of N for $T \rightarrow 0$.

8.2 Approaching $T = 0$ in experiment

A further important consequence of the third principle of thermodynamics—in particular, of the fact that the $T = 0$ entropy is independent of all other thermodynamic

variables— is the impossibility to reach the absolute zero of temperature in experiment. It is easy to show that an infinite number of intermediate thermodynamic processes would be necessary in order to succeed. Only if the number of intermediate steps tends to infinite one is able to come arbitrarily close to the $T = 0$ state, without ever reaching it.

The basic reasoning goes as follows. Since $S = 0$ independently of any other thermodynamic variables and of the details of the system, we must have that all curves in the (S, T) diagram corresponding to any constant value of a thermodynamic variable P (e.g., pressure or volume) merge at $T = 0$ in order to yield $S = 0$. This is illustrated in Fig. 6(a). In contrast, the situations illustrated in Figs. 6(b) and 6(c) are not possible, since they would contradict the third law.

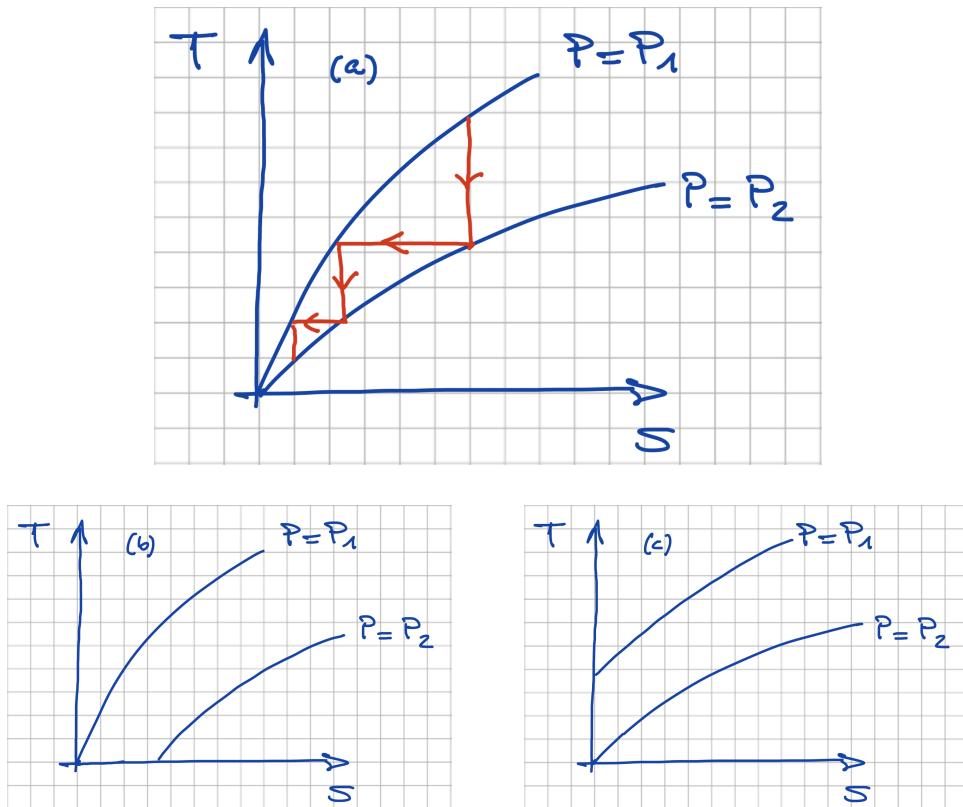


Figure 6: (a) Schematic curves corresponding to different constant values of some thermodynamic variable P (e.g., $P = p$ or $P = V$) in an (S, T) diagram. Notice the convergence of all constant- P curves to the same entropy $S = 0$ for $T = 0$. (b) Unphysical (S, T) diagram since it implies that S depends on the thermodynamic variable P at $T = 0$. (c) Unphysical (S, T) diagram since it implies that $\partial S / \partial T = 0$ for $T \geq 0$.

One can imagine, for example, an isoentropic cooling from $p = p_1$ to $p = p_2$ (vertical

line) but then, in order to be able to achieve further cooling, one should reduce the entropy. This can be done at best via an isothermal process, since (after some point) no thermal bath cooler than the system can be found (horizontal line). Since the values of P must be finite, the isothermal process leaves the system at some finite value of $p = p_1$. From there, a new isoentropic cooling can be made. In order to reach $T = 0$ (or even to approach $T = 0$ arbitrarily close) one would need an infinite number of constant T , constant S processes. Notice that this would not be necessary if all the constant- P curves in the (S, T) diagram would not converge to the same value of S at $T = 0$, as illustrated in Fig. 6(b).

9 The principle of indistinguishability of identical particles

Before discussing the properties of macroscopic systems at low temperatures and assessing the limitations of classical statistical mechanics, it is absolutely necessary to recall the principle of indistinguishability of identical particles and its consequences on the symmetry of many-body wave functions. In classical mechanics the particles preserve their identity despite having the same physical properties. A numbering is possible without altering the mechanical behavior in any significant respect. This allows us in principle to follow the trajectory of each particle along its path. This reasoning applies to any number of particles and in particular to particles that are identical in all respects. For example, from a classical perspective a state in which particle 1 has a position and momentum (q, p) and particle 2 has the coordinates (q', p') is fundamentally distinct for the case in which particle 1 is in state (q', p') and particle 2 in state (q, p) .

In quantum mechanics the situation is entirely different, since the motion of deterministic path ceases to exist as a consequence of the Heisenberg uncertainty principle. Even if a labeling of the particles is attempted by measuring all the positions of each particle with arbitrary accuracy at some time t , it is not possible to track the position of the particles at any future time $t' > t$. In fact, the coordinates have no definite values even at times t' arbitrarily close to t . Suppose that we localize (measure the position) of a particle at a given time $t' > t$. It would then be impossible to say which particle among the initial N ones has arrived at this point. This is clearly illustrated by the following scattering experiment:

[Picture]

In quantum mechanics identical particles entirely lose their individuality. They are completely indistinguishable. No experimental measurement can ever remove this indistinguishability. This fundamental principle of indistinguishability of identical particles has many far reaching consequences.

Consider two observers O and O' who attempt to defy the principle of indistinguishability by preparing two experiments, adopting different conventions for labeling the electronic coordinates. For example, the scattering between two electronic wave packets is prepared. According to O electron 1, in the wave packet $|a\rangle$, interacts with electron 2 in the wave packet $|b\rangle$. The corresponding state considered by O is $|\psi\rangle$ with coordinate wave function $\langle x_1, x_2 | \psi \rangle = \psi(x_1, x_2)$. According to observer O' , electron 2 is the one in state $|a\rangle$, while electron 1 is in state $|b\rangle$. The corresponding two particle state is $|\psi'\rangle$ with wave function $\langle x_1, x_2 | \psi' \rangle = \psi'(x_1, x_2)$. The principle of indistinguishability of identical particles states that $|\psi\rangle$ and $|\psi'\rangle$ are two equivalent representations of the same physical state. No measurement can ever make the difference between $|\psi\rangle$ and $|\psi'\rangle$. This means no more and no less that for any state $|\beta\rangle$ the probability of finding $|\psi\rangle$ in $|\beta\rangle$ is the same as the probability of finding $|\psi'\rangle$ in $|\beta\rangle$:

$$|\langle \beta | \psi \rangle|^2 = |\langle \beta | \psi' \rangle|^2$$

for any $|\beta\rangle$. Setting

$$|\beta\rangle = |\psi\rangle$$

we have

$$|\langle\psi|\psi\rangle|^2 = 1 = |\langle\psi|\psi'\rangle|^2$$

and setting $|\beta\rangle = |\psi'\rangle$ we have

$$|\langle\psi|\psi'\rangle|^2 = 1 = |\langle\psi|\psi\rangle|^2.$$

It is easy to see that two vectors having a norm equal to 1 and an overlap with absolute value equal to 1 must be collinear. This implies

$$|\psi'\rangle = e^{i\alpha}|\psi\rangle.$$

Consequently,

$$\langle x_1, x_2|\psi'\rangle = \psi'(x_1, x_2) = e^{i\alpha}\psi(x_1, x_2) = \langle x_1, x_2|\psi\rangle.$$

Repeating the interchange, we obtain

$$\begin{aligned} \psi(x_1, x_2) &= e^{i\alpha}\psi(x_2, x_1) = e^{2i\alpha}\psi(x_1, x_2) \quad \forall x_1, x_2 \\ \Rightarrow e^{2i\alpha} &= 1 \quad \Rightarrow e^{i\alpha} = \pm 1 \end{aligned}$$

Consequently

$$\psi(x_1, x_2) = \pm\psi(x_2, x_1).$$

Obviously, the same sign must hold for any two particles of the same kind.

Applying the previous argument to any two particles belonging to a larger N -particle system we have

$$\psi(x_1, \dots, x_i, \dots, x_j, \dots, x_N) = \pm\psi(x_1, \dots, x_j, \dots, x_i, \dots, x_N)$$

for all i and j . The particles in nature are thus divided in two disjoint groups: the particles having fully symmetric wave functions (+ sign), which are called *bosons*, and the particles having fully antisymmetric wave functions (− sign), which are called *fermions*. The property of being a boson or a fermion is an intrinsic property of any particle. It can be shown experimentally that there is a one-to-one correspondence between the bosonic or fermionic character and integer or half integer nature of the particles' intrinsic spin. Bosons have integer spin, while fermions have half-integer spin. Most elementary particles are fermions (e^- , p^+ , n , e^+). Photons and most elementary excitations in condensed matter are bosons (phonons, magnons, etc.).

Complex particles, for example, atoms, molecules or nuclei have fermionic character if the number of fermions constituting them (elementary particles) is odd. Otherwise, they show bosonic character. For instance, ^3He atoms are fermions while ^4He atoms are bosons. As we shall see, this has crucial consequences for the thermodynamic properties at low temperatures.

The exchange or transposition of the coordinates of two particles i and j is a particular case of the most general permutation of N coordinates x_1, \dots, x_N . A permutation $P : [1, N] \rightarrow [1, N]$ in the natural interval $[1, N]$ is a bijective function within $[1, N]$. We may denote it by

$$\begin{pmatrix} 1 & 2 & \dots & N \\ P(1) & P(2) & \dots & P(N) \end{pmatrix}$$

or simply by $P = [P(1), P(2), \dots, P(N)]$. For example, a transposition corresponds to

$$\begin{pmatrix} 1 & \dots & i & \dots & j & \dots & N \\ 1 & \dots & j & \dots & i & \dots & N \end{pmatrix}.$$

It is very useful to define the order of a permutation P , which we denote by $O(P) = p$, as the number of transpositions required to bring the sequence $[P(1), P(2), \dots, P(N)]$ to the normal order $[1, 2, \dots, N]$. For example

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \quad \text{has} \quad p = 1$$

and

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \quad \text{has} \quad p = 2.$$

A simple transposition has always $p = 1$. One may easily convince oneself that p corresponds to the number line crossing in the diagrams of the form

$$\begin{pmatrix} 1 & 2 & \dots & N \\ P(1) & P(2) & \dots & P(N) \end{pmatrix}, \quad (9.1)$$

where the lines connect the numbers i on the upper row with $P(i)$ on the lower row. Knowing that $\psi(x_1, \dots, x_i, \dots, x_j, \dots, x_N) = \pm \psi(x_1, \dots, x_j, \dots, x_i, \dots, x_N)$ we conclude that

$$\psi(x_{P(1)}, x_{P(2)}, \dots, x_{P(N)}) = \psi(x_1, x_2, \dots, x_N)$$

for bosons and

$$\psi(x_{P(1)}, \dots, x_{P(N)}) = (-1)^p \psi(x_1, \dots, x_N)$$

for fermions. Functions with these properties are said to be fully symmetric (bosons) or fully antisymmetric (fermions).

9.1 Many-particle wave functions

The superposition principle states that any linear combination of two quantum mechanical states $|\psi_1\rangle$ and $|\psi_2\rangle$ is a possible state. Therefore, all the wave functions of a system of N identical particles must have the same fermionic or bosonic character. Otherwise the linear combination would have no defined symmetry. In particular, the basis states must be fully antisymmetric or symmetric. We seek for a simple complete basis set

for fermions and bosons, in terms of which we can compute the partition functions and statistical averages of systems of identical particles.

Consider an arbitrary complete set of orthonormal single-particle wave functions or spin-orbitals

$$\{\varphi_1(x), \varphi_2(x), \dots\} = \{\varphi_\alpha \text{ with } \alpha = 1, 2, \dots\},$$

where

$$\sum_{\alpha} \varphi_{\alpha}^*(x) \varphi_{\alpha}(x') = \delta(x - x')$$

and

$$\int \varphi_{\alpha}^*(x) \varphi_{\beta}(x) dx = \delta_{\alpha\beta}.$$

For simplicity, we introduce the notation $x_i \equiv \vec{r}_i$ for the spatial coordinates of spinless particles, and also $x_i \equiv (\vec{r}_i, \sigma_i)$ for particles having a spin variable σ_i (e.g., electrons). Consequently, the delta functions and integrals are to be interpreted as

$$\delta(x - x') = \delta^{(3)}(\vec{r} - \vec{r}') \delta_{\sigma\sigma'}$$

and

$$\int \dots dx = \sum_{\sigma} \int \dots d^3r.$$

Since the set $\{\phi_{\alpha}(x)\}$ is a complete set for single-variable functions, any N variable function can be expanded as a linear combination of simple products of the form

$$\varphi_{k_1}(x_1), \varphi_{k_2}(x_2), \dots, \varphi_{k_N}(x_N).$$

However, simple products are not appropriate N -particle basis wave functions, since they are neither symmetric nor antisymmetric. In other words, simple products are not physically acceptable states for identical particles. It is, however, simple to obtain a complete basis set appropriate for fermions or bosons by symmetrizing the products $\varphi_{k_1}, \dots, \varphi_{k_N}$.

Exercise 9.16: Show that $\psi(x_1, \dots, x_N)$ as given by Eq. (9.2) fulfills the required anti-symmetry properties of fermion wave functions.

9.2 Fermions

In the case of fermions, the simplest fully antisymmetric wave functions that can be constructed from a single product reads

$$\begin{aligned}
 \psi(x_1, \dots, x_N) &= \frac{1}{\sqrt{N!}} \sum_P (-1)^p \varphi_{k_{P(1)}}(x_1) \varphi_{k_{P(2)}}(x_2) \cdots \varphi_{k_{P(N)}}(x_N) \\
 &= \frac{1}{\sqrt{N!}} \sum_P (-1)^p \varphi_{k_1}(x_{P(1)}) \varphi_{k_2}(x_{P(2)}) \cdots \varphi_{k_N}(x_{P(N)}) \\
 &= \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_{k_1}(x_1) & \cdots & \varphi_{k_1}(x_N) \\ \vdots & & \vdots \\ \varphi_{k_N}(x_1) & \cdots & \varphi_{k_N}(x_N) \end{vmatrix}. \tag{9.2}
 \end{aligned}$$

These functions are known as Slater determinants. Each Slater determinant is univocally defined by specifying the occupation numbers n_α of each single-particle spin-orbital φ_α , i.e., by specifying whether α appears in the list k_1, k_2, \dots, k_N , in which case $n_\alpha = 1$, or not, in which case $n_\alpha = 0$. The phase ambiguity is removed by requiring $k_1 < k_2 < \cdots < k_N$. Notice that $\Psi = 0$ if any of the single-particle orbitals is occupied twice. The Pauli exclusion principle imposes $n_\alpha = 0$ or 1 for fermions. Let us recall that the single particle states $\varphi_\alpha(x)$ concern both spatial coordinates \vec{r}_i and the spin variables σ_i . We may thus write

$$\Psi_{k_1, \dots, k_N}(x_1, \dots, x_N) = \langle x_1, \dots, x_N | n_1, n_2, \dots \rangle,$$

where $n_\alpha = 1$ for $\alpha = k_1, k_2, \dots, k_N$, and $n_\alpha = 0$ otherwise. From the occupation-number perspective, the following notation for the Slater-determinant wave function seems more appropriate:

$$\Psi_{\{n_\alpha\}}(x_1, \dots, x_N) = \langle x_1, \dots, x_N | n_1, n_2, \dots \rangle,$$

where $|n_1, n_2, \dots\rangle$ is the N -fermion ket corresponding to the occupation numbers $n_\alpha = 0, 1$ and $\sum_\alpha n_\alpha = N$. It is easy to verify that the Slater determinants are properly normalized and that two Slater determinants are orthogonal unless all occupation numbers coincide:

$$\begin{aligned}
 \langle n_1, n_2, \dots | n'_1, n'_2, \dots \rangle &= \int dx_1 \cdots dx_N \Psi_{\{n_\alpha\}}^*(x_1, \dots, x_N) \Psi_{\{n'_\alpha\}}(x_1, \dots, x_N) \\
 &= \delta_{n_1 n'_1} \delta_{n_2 n'_2} \cdots \delta_{n_\infty n'_\infty}.
 \end{aligned}$$

Once we have the orthonormal many-body basis $|n_1, n_2, \dots\rangle$ or $\psi_{\{n_\alpha\}}(x_1, \dots, x_N)$, we can compute the trace of any operator \hat{O} in the Hilbert space of fermion states having an arbitrary number of particles:

$$\begin{aligned}
 \text{Tr}\{\hat{O}\} &= \sum_{\{n_\alpha\}} \langle n_1, n_2, \dots | \hat{O} | n_1, n_2, \dots \rangle \\
 &= \sum_{n_1=0}^1 \sum_{n_2=0}^1 \cdots \sum_{n_\alpha=0}^1 \cdots \langle n_1, n_2, \dots | \hat{O} | n_1, n_2, \dots \rangle,
 \end{aligned}$$

where

$$\langle n_1, n_2, \dots | \hat{O} | n_1, n_2, \dots \rangle = \int dx_1 \dots dx_N \Psi_{\{n_\alpha\}}^*(x_1, \dots, x_N) \hat{O} \Psi_{\{n_\alpha\}}(x_1, \dots, x_N).$$

For example, the grand canonical partition function is given by

$$Z_{gc} = \sum_{n_1=0}^1 \sum_{n_2=0}^1 \dots \sum_{n_\alpha=0}^1 \dots \langle n_1, n_2, \dots | e^{-\beta(\hat{H} - \mu \hat{N})} | n_1, n_2, \dots \rangle.$$

Of course, in the case of a canonical partition function the constraint $\sum_\alpha n_\alpha = N$ must be enforced.

An alternative way to compute the trace is to sum over all possible choices of the N occupied orbitals (i.e., the orbitals appearing in the product $\varphi_{k_1}, \dots, \varphi_{k_N}$) and to correct for the multiple appearance of the same set of orbitals. In the case of fermions all non-allowed products showing the same orbitals two or more times are automatically ignored, since the corresponding Slater determinant is zero. However, if one sums over each value of k_i independently, one obtains $N!$ times the same set of orbitals. Therefore we have

$$\text{Tr}\{\hat{O}\} = \frac{1}{N!} \sum_{k_1=1}^{\infty} \dots \sum_{k_N=1}^{\infty} \int dx_1 \dots dx_N \Psi_{k_1, \dots, k_N}(x_1, \dots, x_N)^* \hat{O} \Psi_{k_1, \dots, k_N}(x_1, \dots, x_N).$$

This form is more appropriate for the canonical ensemble, where the number of particles is fixed.

9.3 Bosons

In the case of bosons, the fully symmetric wave function obtained from a simple product of single-particle orbitals looks even simpler than the Slater determinant. It can be written as

$$\begin{aligned} \tilde{\Phi}_{k_1, \dots, k_N}^s(x_1, \dots, x_N) &= \frac{1}{\sqrt{N!}} \sum_P \varphi_{k_{P(1)}}(x_1) \dots \varphi_{k_{P(N)}}(x_N) \\ &= \frac{1}{\sqrt{N!}} \sum_P \varphi_{k_1}(x_{P(1)}) \dots \varphi_{k_N}(x_{P(N)}) \\ &= \tilde{\Phi}_{\{n_\alpha\}}^s(x_1, \dots, x_N). \end{aligned}$$

It is easy to see that $\tilde{\Phi}^s(x_{Q(1)}, \dots, x_{Q(N)}) = \tilde{\Phi}^s(x_1, \dots, x_N)$ for any permutation Q . Moreover,

$$\int \left(\tilde{\Phi}_{\{n_\alpha\}}^s \right)^* \tilde{\Phi}_{\{n'_\alpha\}}^s dx_1 \dots dx_N = 0$$

unless the occupation numbers $n_\alpha = n'_\alpha$ are the same $\forall \alpha$. Indeed, if this is not the case one of the integrals $\int \varphi_i^*(x) \varphi_j(x) dx = 0$. The form of $\tilde{\Phi}^s$ is appealing, since

it is the same as the Slater determinant except for the phase factor $(-1)^p$. However, it has the inconvenience of not being normalized to 1. In fact in the \sum_P there are $n_1!n_2!\dots n_\alpha!\dots = \prod_\alpha n_\alpha!$ terms which are exactly the same. They correspond to the permutations of the indices $k_{P(i)}$ among the k_i corresponding to the same single-particle orbital α . The properly normalized fully symmetric states are therefore given by

$$\Phi_{\{n_\alpha\}}^s(x_1, \dots, x_N) = \frac{\tilde{\Phi}^s(x_1, \dots, x_N)}{\sqrt{\prod_\alpha n_\alpha!}}.$$

This can also be written as

$$\Phi_{\{n_\alpha\}}^s(x_1, \dots, x_N) = \sqrt{\frac{\prod_\alpha n_\alpha!}{N!}} \sum_{P \neq k} \phi_{k_{P(1)}}(x_1) \dots \phi_{k_{P(N)}}(x_N)$$

where $\sum_{P \neq k}$ is the sum over all possible permutations of the indices $i = 1, \dots, N$ such that $k_{P(i)} \neq k_i$ for some i . In other words, in $\sum_{P \neq k}$ only permutations which actually change the product are taken into account. There are $N!/(\prod_\alpha n_\alpha!)$ such permutations, since the number of permutations among indices corresponding to the same orbitals is $\prod_\alpha n_\alpha!$.

As in the case of fermions, indistinguishability implies that the many-particle state or wave function is univocally defined by the occupation numbers $\{n_\alpha\} = n_1, n_2, \dots$. We therefore introduce the bosons kets $|n_1, n_2, \dots\rangle$ corresponding to the occupation numbers $\{n_\alpha\}$, which are defined by

$$\langle x_1, \dots, x_N | n_1, n_2, \dots \rangle = \Phi_{\{n_\alpha\}}^s(x_1, \dots, x_N).$$

It is easy to see that

$$\langle n_1, n_2, \dots | n'_1, n'_2, \dots \rangle = \delta_{n_1 n'_1} \delta_{n_2 n'_2} \dots \delta_{n_\alpha n'_\alpha} \dots$$

for all $\{n_\alpha\}$ and all $N = \sum_\alpha n_\alpha$, since $\Phi_{\{n_\alpha\}}^s$ is normalized to 1.

Once we have the orthonormal many-body basis $|n_1, n_2, \dots\rangle$ or $\Phi_{\{n_\alpha\}}^s(x_1, \dots, x_N)$ we can compute the trace of any operator \hat{O} as

$$\text{Tr}\{\hat{O}\} = \sum_{\{n_\alpha\}} \langle n_1, n_2, \dots | \hat{O} | n_1, n_2, \dots \rangle = \sum_{n_1=0}^{\infty} \dots \sum_{n_\alpha=0}^{\infty} \dots \langle n_1, n_2, \dots | \hat{O} | n_1, n_2, \dots \rangle,$$

where

$$\langle n_1, n_2, \dots | \hat{O} | n_1, n_2, \dots \rangle = \int dx_1 \dots dx_N \Phi_{\{n_\alpha\}}^s(x_1, \dots, x_N)^* \hat{O} \Phi_{\{n_\alpha\}}^s(x_1, \dots, x_N).$$

Notice that each term in the sum corresponds to a distinct many-body state. It is therefore important that $|n_1, n_2, \dots\rangle$ and $\Phi_{\{n_\alpha\}}^s(x_1, \dots, x_N)$ are normalized to 1. For example, the grand canonical partition function is given by

$$Z_{gc} = \sum_{n_1=0}^{\infty} \dots \sum_{n_\alpha=0}^{\infty} \dots \langle n_1, n_2, \dots | e^{-\beta(\hat{H} - \mu\hat{N})} | n_1, n_2, \dots \rangle.$$

In this case there are no restrictions on the total number of bosons $N = \sum_{\alpha} n_{\alpha}$. Calculating the canonical partition function Z_c would be more complicated, since one would need to impose the constraint $\sum_{\alpha} n_{\alpha} = N$ on the occupation numbers.

An alternative way of computing the trace is to sum over all orbitals indices k_1, k_2, \dots, k_N independently (not occupation numbers!) and to correct for the multiple appearances. In the case of bosons the number of times that the same product of orbitals $\varphi_{k_1}, \dots, \varphi_{k_N}$ appears is given by $\frac{N!}{n_1!n_2!\dots n_{\infty}!} = \frac{N!}{\prod_{\alpha} n_{\alpha}!}$, which is the number of permutations between the indices k_i having different values, i.e., $k_i \neq k_{P(i)}$ for some i . We therefore have

$$\text{Tr}\{\hat{O}\} = \sum_{k_1=1}^{\infty} \cdots \sum_{k_N=1}^{\infty} \frac{\prod_{\alpha} n_{\alpha}!}{N!} \int dx_1 \dots dx_N \Phi_{k_1, \dots, k_N}^s(x_1, \dots, x_N)^* \hat{O} \Phi_{k_1, \dots, k_N}^s(x_1, \dots, x_N).$$

At first sight this seems quite inconvenient, since the factor $\prod_{\alpha} n_{\alpha}!$ depends on the values of k_1, k_2, \dots, k_N in a complicated way. However, the expression simplifies remarkably if we recall that

$$\tilde{\Phi}_{k_1, \dots, k_N}^s(x_1, \dots, x_N) = \sqrt{\prod_{\alpha} n_{\alpha}!} \Phi_{k_1, \dots, k_N}^s(x_1, \dots, x_N) \quad (9.3)$$

$$= \frac{1}{\sqrt{N!}} \sum_P \varphi_{k_1}(x_{P(1)}) \cdots \varphi_{k_N}(x_{P(N)}). \quad (9.4)$$

We finally conclude that

$$\text{Tr}\{\hat{O}\} = \frac{1}{N!} \sum_{k_1=1}^{\infty} \cdots \sum_{k_N=1}^{\infty} \int dx_1 \dots dx_N \tilde{\Phi}_{k_1, \dots, k_N}^s(x_1, \dots, x_N)^* \hat{O} \tilde{\Phi}_{k_1, \dots, k_N}^s(x_1, \dots, x_N), \quad (9.5)$$

which has the same form as in the case of fermions. Only the phase factor $(-1)^P$ is replaced by $(+1)^P = 1$ in the sum over all permutations P .

These two ways of computing $\text{Tr}\{\hat{O}\}$ should not be confused. In one case, $\sum_{\{n_{\alpha}\}}$, each term of the sum corresponds to a different many-body state, and the kets or wave functions involved must be normalized to 1. In the other case, $\sum_{k_1} \cdots \sum_{k_N}$, the boson wave-functions are not necessarily normalized to one. There are states which appear more than once in the sum, which is corrected by the factor $1/N!$ and the square norm $\prod_{\alpha} n_{\alpha}!$ of the wave functions $\tilde{\Phi}_{k_1, \dots, k_N}^s(x_1, \dots, x_N)$ used to calculate the averages. Thus, the prefactor $1/N!$ does not come exclusively from the normalization of the N -particle wave functions.

Exercise 9.17: Consider the Hamilton operator

$$\hat{H} = \sum_{i=1}^N \hat{h}(x_i)$$

of N non-interacting identical particles. For example, $\hat{H} = \sum_{i=1}^N \frac{\hat{p}_i^2}{2m}$. Note that $\hat{h}(x)$ is the same for all particles, since they are identical. Let $\varphi_k(x)$ with $k = 1, 2, \dots$ be the eigenfunctions of \hat{h} : $\hat{h}\varphi_k = \varepsilon_k\varphi_k$. Show that any simple product

$$\pi(x_1, \dots, x_N) = \varphi_{k_1}(x_1) \varphi_{k_2}(x_2) \dots \varphi_{k_N}(x_N)$$

is an eigenfunction of \hat{H} with eigenenergy $E = \sum_{i=1}^N \varepsilon_{k_i}$. Is $\pi(x_1, \dots, x_N)$ an appropriate wave function in some particular case? Show that the fully symmetrized and antisymmetrized functions $\psi_{\pm}(x_1, \dots, x_N)$ obtained from π are either zero or are eigenfunctions of \hat{H} with the same energy E . Conclude that $\hat{H}|n_1, n_2, \dots, n_{\infty}\rangle = \left(\sum_{\alpha=1}^{\infty} n_{\alpha}\varepsilon_{\alpha}\right)|n_1, n_2, \dots, n_{\infty}\rangle$, where $|n_1, n_2, \dots, n_{\infty}\rangle$ is the state with definite occupation numbers n_{α} of all orbitals φ_{α} [$\langle x_1, x_2, \dots, x_N | n_1, n_2, \dots, n_{\infty}\rangle = \Psi_{\pm}(x_1, \dots, x_N)$].

Exercise 9.18: Consider a degenerate energy level with degeneracy g , which is occupied by n particles ($0 \leq n \leq g$). Calculate the number of possible quantum states $\Omega_g(n)$ for n identical fermions, n identical bosons and n distinguishable particles. What would be the result for "correct Boltzmann counting"? Are there any cases in which $\Omega_g(n)$ is the same for all the three situations? When? Interpret physically.

Repeat the exercise by calculating the number of quantum states $\Omega(\{n_i\})$ for the case in which several different groups of levels i , each with degeneracy g_i , are occupied by n_i particles each ($\sum_i n_i = N$). Express $S = \ln(\Omega)$ in terms of the g_i and the average occupation numbers $\nu_i = n_i/g_i$. Analyze interesting limiting cases for Fermions, Bosons and Boltzmann particles. When do the results for different statistics coincide? Interpret the result physically.

Note: Correct Boltzmann counting corresponds to assuming that the particles are distinguishable, as in classical mechanics, dividing a posteriori the number of states simply by $N!$, where N is the number of particles, in order to mimic indistinguishability.

Exercise 9.19:: Identity of particles:

- i) Given a complete orthonormal single-variable basis set $\{\varphi_\alpha(x), \alpha = 1, 2, \dots\}$ satisfying

$$\sum_{\alpha} \varphi_{\alpha}(x)^* \varphi_{\alpha}(x') = \delta(x - x')$$

and

$$\int \varphi_{\alpha}(x)^* \varphi_{\beta}(x) dx = \delta_{\alpha\beta},$$

it is always possible to expand an N variable function Ψ in the form

$$\Psi(x_1 \dots x_N) = \sum_{k_1=1}^{\infty} \dots \sum_{k_N=1}^{\infty} c(k_1, \dots, k_N) \varphi_{k_1}(x_1) \dots \varphi_{k_N}(x_N).$$

- a) Consider now the case where $\Psi(x_1, \dots, x_N)$ is an N -particle fermionic wave function. Show that

$$c(k_1, \dots, k_i \dots k_j \dots k_N) = -c(k_1, \dots, k_j \dots k_i \dots k_N).$$

for all i and j . Conclude that two fermions can never occupy the same single-particle state or spin-orbital (Pauli exclusion principle). Remember that the index k_i (or α) defines the single-particle state completely, i.e., it includes the spin variable.

- b) Show that $c(k_{P(1)}, \dots, k_{P(N)}) = (-1)^p c(k_1, \dots, k_N)$ for an arbitrary permutation P , where p is the order of P . Conclude that for each choice of occupations n_1, n_2, \dots with $n_\alpha = 0, 1$ and $\sum_{\alpha} n_{\alpha} = N$, there is only one independent coefficient, for instance,

$$c(k_1 < k_2 < \dots < k_N) = c(n_1, \dots, n_{\alpha}, \dots).$$

- c) Consider now the case where $\Psi(x_1, \dots, x_N)$ is an N -particle bosonic wave function. Show that

$$c(k_1, \dots, k_i \dots k_j \dots k_N) = c(k_1, \dots, k_j \dots k_i \dots k_N)$$

for all i, j . Generalize the statement to an arbitrary permutation of the indices:

$$c(k_{P(1)} \dots k_{P(N)}) = c(k_1, \dots, k_N) \quad \forall P.$$

Conclude that for a given choice of orbital occupations n_1, n_2, \dots with $\sum_{\alpha} n_{\alpha} = N$, all the coefficient are the same. In other words, for each set of n_{α} there is only one independent coefficient $c(n_1, n_2, \dots, n_{\alpha}, \dots)$.

ii) Consider the Slater determinant or fully antisymmetric single-product state

$$\Phi_{\{n_\alpha\}}(x_1, \dots, x_N) = \frac{1}{\sqrt{N!}} \sum_P (-1)^p \varphi_{k_{P(1)}}(x_1) \dots \varphi_{k_{P(N)}}(x_N),$$

where k_1, \dots, k_N are the occupied orbitals (i.e., $n_\alpha = 1$ for $\alpha = k_1, \dots, k_N$) and the sum runs over all permutations P having order p . Show that $\Phi_{\{n_\alpha\}}(x_1, \dots, x_N)$ is normalized to 1 provided that the single-particle orbital $\varphi_\alpha(x)$ are orthonormal and $k_i \neq k_j$ for all i, j . Show that $\Phi_{\{n_\alpha\}} = 0$ if $k_i = k_j$ for some i and j .

iii) Consider the symmetrized bosonic single product state

$$\tilde{\Phi}_{\{n_\alpha\}}^s = \frac{1}{\sqrt{N!}} \sum_P \varphi_{k_{P(1)}}(x_1) \dots \varphi_{k_{P(N)}}(x_N)$$

for certain occupation numbers n_α satisfying $\sum_\alpha n_\alpha = N$.

10 The classical limit

In the previous section we have found closed expressions for the trace of any operator \hat{O} according to the laws of quantum mechanics and the principle of indistinguishable identical particles. In principle every real system should be described in terms of the corresponding symmetrized or antisymmetrized many-body quantum states. However, in many cases of interest the quantum mechanical description is superfluous. We would like to derive the classical approach to statistical mechanics, in which, according to the classical picture, the state of the system is defined by specifying the positions $\vec{r}_1, \dots, \vec{r}_N$ and the momenta $\vec{p}_1, \dots, \vec{p}_N$ of each particle. This is formally achieved by starting from the quantum expression for the partition function Z for fermions and bosons and taking the limit $\hbar \rightarrow 0$. An expansion in power of \hbar is thus obtained, in which the zeroth order term gives Gibbs classical expression for Z , and the higher order terms quantify the quantum corrections. More important, the derivation allows us to understand how the classical description emerges from the quantum principles and to delimit the domain of applicability of classical statistics.

We will show that if the temperature T is sufficiently high or if the density of particles N/V is sufficiently low (large volume per particle $v = V/N$) we can approximate the canonical partition function by

$$Z_c(T, V, N) = \frac{1}{N!} \frac{1}{(2\pi\hbar)^{3N}} \int dp^{3N} dq^{3N} e^{-\beta H(p,q)}, \quad (10.1)$$

where $H(p, q)$ is the classical Hamilton function. In Cartesian coordinates this is given by

$$H = \sum_{i=1}^N \frac{p_i^2}{2m} + W(\vec{r}_1 \dots \vec{r}_N),$$

where $W(\vec{r}_1, \dots, \vec{r}_N) = (1/2) \sum_{i,j} w(|\vec{r}_i - \vec{r}_j|) + \sum_i v(\vec{r}_i)$ is the interaction energy between the particles and with any external fields.

Before the derivation, it is interesting to discuss the range of validity and consequences of Eq.(10.1). The classical limit is valid when the average interparticle distance $v^{1/3} = (V/N)^{1/3}$ satisfies

$$v^{1/3} = \left(\frac{V}{N}\right)^{1/3} \gg \lambda, \quad (10.2)$$

where

$$\lambda = \frac{2\pi\hbar}{\sqrt{2\pi m k_B T}} = \sqrt{\frac{2\pi\hbar^2}{m k_B T}} \quad (10.3)$$

is the thermal de Broglie wavelength. It represents the quantum mechanical uncertainty in the position of a particle having an uncertainty in the momentum corresponding to the kinetic energy $p^2/2m = k_B T$. In fact, $\Delta p = \sqrt{\langle p^2 \rangle} = \sqrt{2m k_B T} \Rightarrow \Delta x \sim \hbar/\Delta p = \hbar/\sqrt{2m k_B T} = \lambda/2\sqrt{\pi}$. Thus, λ , which is proportional to \hbar , can be regarded as

a physically more appropriate expansion parameter than \hbar itself. In fact, \hbar never goes to zero. Still, the classical limit is reached for sufficiently large T , since $\lambda \propto 1/\sqrt{T}$. If $\lambda \ll v^{1/3} = (V/N)^{1/3}$ the particles can be regarded as occupying wave packets, whose width is of the order of λ . In this limit the overlap between the different wave packets is nearly zero and the differences between fermionic and bosonic statistics vanish. Indeed, if the occupied orbitals do not overlap at all [i.e., $\varphi_\alpha(\vec{r})\varphi_\beta(\vec{r}) \cong 0, \forall \vec{r} \in \mathbb{R}^3$] the probability of finding a particle at point \vec{r} and another particle at point \vec{r}' , usually given by

$$\begin{aligned} P(\vec{r}, \vec{r}') &= \langle \delta(\vec{r} - \vec{r}_1) \delta(\vec{r}' - \vec{r}_2) \rangle = \int dr_1 \dots dr_N |\Phi(\vec{r}_1 \dots \vec{r}_N)|^2 \delta(\vec{r} - \vec{r}_1) \delta(\vec{r}' - \vec{r}_2) \\ &= \int |\Phi(\vec{r}, \vec{r}', \vec{r}_3 \dots \vec{r}_N)|^2 dr_3 \dots dr_N, \end{aligned}$$

takes the simpler form

$$P(\vec{r}, \vec{r}') = \frac{1}{N(N-1)} \sum_{\alpha \neq \beta} |\varphi_\alpha(\vec{r})|^2 |\varphi_\beta(\vec{r}')|^2.$$

This holds independently of the symmetry of the many-body wave function provided that $\varphi_\alpha(\vec{r})\varphi_\beta(\vec{r}) = 0$ for all \vec{r} , α and β , as already discussed for the case of two particles.

On the opposite limit, $T \rightarrow 0$, $\lambda \propto 1/\sqrt{T}$ diverges, since the momenta of the particles become well defined ($\langle p^2 \rangle \sim k_B T \rightarrow 0$). In this case the overlap between the occupied single-particle states is always important ($\lambda \gtrsim v^{1/3}$). Thus, the quantum mechanical symmetry-correlations resulting from indistinguishability play a central role. One concludes that the classical approximation always breaks down for $T \rightarrow 0$, regardless of the tolerated accuracy by which the classical model may describe the microstates of the system.

In order to get a feeling of the order of magnitude of the temperature below which classical statistics fails, we may compute v and λ for electrons in solids and for a typical atomic gas. Using that $2\pi\hbar = 6,6 \times 10^{-34}$ Js, $m_e = 9,1 \times 10^{-31}$ kg, and $k_B = 1,4 \times 10^{-23}$ J/K, we have

$$T[\text{K}] = \frac{(2\pi\hbar)^2}{2\pi k_B} \frac{1}{\lambda^2 m} = \frac{5 \times 10^{-45}}{\lambda^2 [\text{m}^2] m [\text{kg}]} = \frac{5 \times 10^{-38}}{\lambda^2 [\text{cm}^2] m [\text{g}]} = \frac{5,44 \times 10^5}{\lambda^2 [\text{\AA}^2] m [m_e]}. \quad (10.4)$$

In solids we have typically one electron per cubic \AA , i.e., $v = V/N \simeq 1 \text{\AA}^3$. If we set $\lambda = v^{1/3}$ in Eq. (10.4), we obtain that the temperature should be larger than $5,5 \times 10^5$ K in order that the classical approximation starts to be applicable. Therefore, electrons in solids can never be treated classically. The symmetry correlations are always important in condensed matter. The situation is quite different in not too dense gases with, for example, $(V/N)^{1/3} = 10 \text{\AA}$ and $m = Am_p$, where A is the atomic weight and $m_p = 1,76 \times 10^3 m_e$ the proton mass. In this case, the symmetry of the wave function ceases to be important for temperatures above $T \simeq 3/A$ K. Lighter atoms are more affected by quantum effects, since for the same average kinetic energy or temperature the uncertainty

in momentum is smaller, for instance, $A = 4$ for He while $A = 131$ for Xe. If the density of the gas is lower, one may apply the classical approximation to lower temperatures but never for $T \rightarrow 0$, since in this case λ diverges. The important point, in order that the symmetry of the many-body wave function (fermionic or bosonic) becomes irrelevant, is that the probability of finding two particles in the same state should be extremely low. As we shall see in the exercises, the entropies of Fermi, Bose and Boltzmann gases coincide, if and only if all the occupation numbers are extremely low. This is only possible for sufficiently high temperatures or sufficiently low densities.

The classical partition function (10.1) shows important signs of its quantum mechanical origin, despite the fact that the microstates are characterized by (p, q) . The first one is the factor $1/2\pi\hbar$ per degree of freedom and the second one is the factor $1/N!$, where N is the number of identical particles. They are both important consequences of the right (quantum mechanical) counting of the number of (linearly independent) states having the same energy. The first one tells us that in phase space only one state fits in a volume element $2\pi\hbar$. A factor proportional to $1/\hbar$ correcting each component of the phase-space volume element, could have been expected, since we know that each state must have $\Delta p_i \Delta q_i \geq \hbar/2$. It is not possible to find states occupying a smaller volume in phase space. Note, moreover, that this factor renders Z dimensionless, as it should.

The second factor tells us that N identical particles having the set of momenta and coordinates $\{(p_1, q_1), (p_2, q_2), \dots, (p_N, q_N)\}$ correspond to one and the same state, irrespectively of which particle has which (p_i, q_i) . Notice that all $N!$ permutations of the coordinates and momenta appear in the integral. Dividing by $1/N!$ simply ensures that these are counted only once. Furthermore, as we shall see, this factor is necessary in order that the thermodynamic potentials derived from Z are extensive.

10.1 Boltzmann counting

Eq.(10.1) allows us to introduce the so-called correct Boltzmann counting, which corresponds to the correct counting of states in the limit where the symmetry correlations for fermions or bosons play no role. According to classical mechanics identical particles are distinguishable. Therefore, from a classical perspective, the states $p_1, \dots, p_N, q_1, \dots, q_N$ and all the permutations of them are distinguishable and should be counted separately when computing Z . Correct Boltzmann counting means that we consider the particles as distinguishable for the purpose of computing Z , or the number of accessible states $\Omega(E)$, and that we then divide the result blindly by $N!$, when the particles are identical. This does not correspond to Bosons or Fermions. And there is no real particle following this artificial statistics. Nevertheless, correct Boltzmann counting, or simply Boltzmann statistics, provides a shortcut for the high temperature and low density limits of quantum statistics. It is useful for analyzing them.

10.2 Configurational Integrals

The integration with respect to the momenta in Eq.(10.1) can be done straightforwardly.

Let us recall that $\int_{-\infty}^{\infty} dx e^{-\frac{x^2}{2\sigma^2}} = \sqrt{2\pi\sigma^2}$. Using that

$$H = \sum_i \frac{p_i^2}{2m} + W(\vec{r}_1 \dots \vec{r}_N)$$

we have

$$Z_c = \frac{1}{N!} \int d^3r_1 \dots d^3r_N e^{-\beta W(\vec{r}_1 \dots \vec{r}_N)} \frac{1}{(2\pi\hbar)^{3N}} \left(\int_{-\infty}^{\infty} e^{-\beta \frac{p^2}{2m}} dp \right)^{3N},$$

where

$$\int_{-\infty}^{\infty} e^{-\beta \frac{p^2}{2m}} dp = \sqrt{2\pi m k_B T} = \frac{2\pi\hbar}{\lambda}.$$

Consequently,

$$Z_c(V, T, N) = \frac{1}{N!} \int \frac{d^3r_1 \dots d^3r_N}{\lambda^{3N}} e^{-\beta W(\vec{r}_1 \dots \vec{r}_N)}, \quad (10.5)$$

where $\lambda = 2\pi\hbar/\sqrt{2\pi m k_B T}$. The classical partition function is therefore given by a configurational integral over all possible distributions of the particles having $e^{-\beta W}$ as weighting factor. Performing such integrals is, however, non trivial.

Using Eq. (10.5) one may show that

$$C_V = \frac{3}{2} N k_B + \frac{\langle \Delta W^2 \rangle}{k_B T^2},$$

where $\langle \Delta W^2 \rangle = \langle W^2 \rangle - \langle W \rangle^2$ is the square mean deviation of the interaction energy. Notice that all types of interactions (attractive or repulsive) lead to an enhancement of C_V . Since the second term cannot be negative, it is clear that C_V does not tend to zero for $T \rightarrow 0$. The reasons for the breakdown of the classical approximation have been discussed.

10.3 Virial and equipartition theorems

Let y be any coordinate or momentum on which H depends. Then we have

$$-\frac{1}{\beta} \frac{\partial}{\partial y} e^{-\beta H} = \frac{\partial H}{\partial y} e^{-\beta H}.$$

This implies

$$\left\langle y \frac{\partial H}{\partial y} \right\rangle = \frac{\int y \frac{\partial H}{\partial y} e^{-\beta H} dp dx}{\int e^{-\beta H} dp dx} = -\frac{1}{\beta} \frac{\int y \frac{\partial}{\partial y} e^{-\beta H} dp dx}{\int e^{-\beta H} dp dx}.$$

Integration by parts gives

$$\int y \frac{\partial}{\partial y} e^{-\beta H} dp dx = y e^{-\beta H} \Big|_{y=-\infty}^{y=\infty} - \int e^{-\beta H} dp dx, \quad (10.6)$$

which implies

$$\left\langle y \frac{\partial H}{\partial y} \right\rangle = \frac{1}{\beta} = k_B T, \quad (10.7)$$

provided that $e^{-\beta H} \rightarrow 0$ more rapidly than y for $y \rightarrow \pm\infty$ or at the boundaries of the volume. This is obviously the case for $y = p_i$. For $y = x_i$ we need to incorporate the finite volume restriction in the form of a single particle potential which constraints the motion of the particles within V . This cannot be done in the case of free particles. Therefore Eq.(10.7) is not applicable in the absence of a potential for $y = x_i$.

We take $y = p_i$ and obtain $y \frac{\partial H}{\partial y} = p_i \frac{\partial H}{\partial p_i} = \frac{p_i^2}{m}$. Consequently,

$$\left\langle \frac{p_i^2}{2m} \right\rangle = \frac{1}{2} k_B T. \quad (10.8)$$

This important relation is known as *equipartition theorem*. It tells us that in classical systems the average of *each momentum component* is equal to $\frac{k_B T}{2}$, independently of the particle mass, the nature of the interparticle interactions, or any other parameters characterizing the equilibrium state, such as system size, pressure, volume or energy. Of course, the temperature and the kinetic energy per particle depend on all these parameters. In the classical limit, we may therefore associate temperature to the average kinetic energy of the system.

Example: Consider a close-shell nanoparticle or droplet in which an atom can be promoted from the closed surface shell to above the surface, creating an adatom and a vacancy. Which configuration is hot? Which one is cold?

Applying (10.7) to $y = x_i$ we obtain

$$\left\langle x_i \frac{\partial H}{\partial x_i} \right\rangle = -\langle x_i \dot{p}_i \rangle = k_B T,$$

provided that the boundary contribution to the partial integration can be neglected. Summing over all degrees of freedom we have

$$-\sum_{i=1}^{3N} \left\langle x_i \frac{\partial H}{\partial x_i} \right\rangle = \left\langle \sum_{i=1}^{3N} x_i \dot{p}_i \right\rangle = -3N k_B T = -2 \left\langle \sum_i \frac{p_i^2}{2m} \right\rangle$$

and using Eq.(10.8)

$$\left\langle \sum_{i=1}^{3N} x_i \dot{p}_i \right\rangle = -2 \langle E_{kin} \rangle.$$

This relation is known as *virial theorem*.

Example: Consider a system of N non-interacting particles in a 3D harmonic potential well:

$$W = \sum_{i=1}^N k \frac{r_i^2}{2} \quad \Rightarrow \quad \langle W \rangle = -\frac{1}{2} \sum_{i=1}^{3N} \langle x_i \dot{p}_i \rangle = \frac{3N}{2} k_B T = \langle E_{kin} \rangle. \quad (10.9)$$

10.4 The ideal classical gas

We consider a classical system in the absence of interactions. Setting $W = 0$ in Eq. (10.5), the canonical partition function is given by

$$Z_c(T, V; N) = \frac{1}{N!} \left(\frac{V}{\lambda^3} \right)^N = \frac{1}{N!} [Z_c(T, V, 1)]^N.$$

Using Stirling's formula, $\ln N! = N \ln \left(\frac{N}{e} \right) + O(\ln N)$, we obtain

$$\begin{aligned} \ln Z_c &= N \ln \left(\frac{V}{\lambda^3} \right) - N \ln \left(\frac{N}{e} \right) \\ &= N \ln \left(\frac{V}{N} \frac{e}{\lambda^3} \right). \end{aligned}$$

The free energy is thus given by

$$F(T, V, N) = -N k_B T \ln \left(\frac{e}{N} \frac{V}{\lambda^3} \right).$$

It is important to remark that the factor $1/N!$ in Z_c , which is a consequence of the indistinguishability of identical particles, is absolutely necessary in order that F is extensive, i.e., $F(T, \alpha V, \alpha N) = \alpha F(T, V, N)$. Keeping in mind that $\lambda = 2\pi\hbar/\sqrt{2\pi m k_B T} \sim 1/\sqrt{T}$,

one easily obtains all thermodynamic properties:

$$\begin{aligned}
p &= -\left.\frac{\partial F}{\partial V}\right|_T = \frac{Nk_B T}{V}, \\
S &= -\left.\frac{\partial F}{\partial T}\right|_V = Nk_B \left[\left(\ln \frac{eV}{N\lambda^3} \right) + \frac{3}{2} \right] = -\frac{F}{T} + \frac{3}{2}Nk_B, \\
E &= F + TS = \frac{3}{2}Nk_B T, \\
H &= E + pV = \frac{5}{2}Nk_B T, \\
C_V &= \left.\frac{\partial E}{\partial T}\right|_V = \frac{3}{2}Nk_B, \\
C_p &= \left.\frac{\partial H}{\partial T}\right|_p = \frac{5}{2}Nk_B, \\
\alpha &= \frac{1}{V} \left.\frac{\partial V}{\partial T}\right|_p = \frac{1}{T}, \\
\kappa_T &= -\frac{1}{V} \left.\frac{\partial V}{\partial p}\right|_T = \frac{1}{p}, \text{ and} \\
\kappa_S &= \kappa_T \frac{C_V}{C_p} = \frac{3}{5p}.
\end{aligned}$$

The grand canonical partition function is given by

$$\begin{aligned}
Z_{gc} &= \sum_{N=0}^{\infty} e^{\beta\mu N} Z_c(N) = \sum_{N=0}^{\infty} z^N \frac{1}{N!} \left(\frac{V}{\lambda^3} \right)^N \\
&= e^{zV/\lambda^3},
\end{aligned} \tag{10.10}$$

where we have introduced the fugacity

$$z = e^{\beta\mu},$$

which satisfies the following useful relations:

$$\frac{\partial z}{\partial \mu} = \beta z \quad \text{and} \quad \frac{\partial z}{\partial \beta} = \mu z.$$

The grand canonical potential reads

$$\Phi(T, V, z) = -k_B T \frac{zV}{\lambda^3}$$

or

$$\Phi(T, V, \mu) = -k_B T e^{\beta\mu} \frac{V}{\lambda^3}.$$

The average number of particles is given by

$$N = -\frac{\partial\Phi}{\partial\mu} = e^{\beta\mu} \frac{V}{\lambda^3} = \frac{zV}{\lambda^3}. \quad (10.11)$$

Using this expression for N we can write

$$\Phi = -Nk_B T. \quad (10.12)$$

Moreover, from the definition of Φ (Laplace transformation) we know that $\Phi = -pV$ in general. In the present case Eq. (10.12) implies

$$pV = Nk_B T,$$

which is the known equation of state of the ideal classical gas.

Finally, as a side comment, one may revisit the ideal classical gas from the perspective of the grand canonical pressure ensemble and derive the equation of state by imposing the condition $Z_{gcp} = 1$ to the intensive variables β , μ and p that define the ensemble. From (5.37) and using (10.10) we know that $Z_{gcp} = 1$ implies

$$\beta p = \frac{z}{\lambda^3}, \quad (10.13)$$

which involves only intensive quantities and represents the equation of state of the ideal classical gas, although in an unfamiliar way. In order to rewrite it in the more usual form we need to calculate the average density $\langle\hat{N}\rangle/\langle V\rangle$ taking into account volume fluctuations. To this aim use Eq. (10.11) which holds for any fixed volume V . Therefore, we can take averages and write

$$\langle\hat{N}\rangle = \frac{z}{\lambda^3} \langle V\rangle.$$

Replacing this result in Eq. (10.13) we obtain $p\langle V\rangle = \langle\hat{N}\rangle k_B T$.

10.5 The ideal classical gas in the pressure ensemble

We consider an ideal classical gas which is open with respect to all constants of motion E , N and V . Its equilibrium density operator

$$\hat{\rho}_{gcp} = e^{-\beta(\hat{H} - \mu\hat{N} + pV)},$$

which must satisfy the normalization condition $Z_{gcp} = \text{Tr}\{\rho_{gcp}\} = 1$. The variables defining ρ_{gcp} are here β , μ and p . They define the average energy per particle $E = \langle\hat{H}\rangle/\langle\hat{N}\rangle$ and average volume per particle $\langle\hat{V}\rangle/\langle\hat{N}\rangle$. The intensive variables β , μ and p cannot define the actual size of the system. They are linked by the condition $Z_{gcp} = 1$ which represents the equation of state of the system. Our goal here is to derive the known equation of state of the ideal gas from the grand canonical pressure perspective.

In order to calculate the trace over all possible volumes of the system we introduce a maximal attainable volume V_m and then remove this constraint by letting V_m tend to infinity. Thus, $Z_{gcp} = \text{Tr}\{e^{-\beta(\hat{H}-\mu\hat{N}+pV)}\}$ can be expressed in terms of the already calculated grand canonical partition function $Z_{gc}(V) = e^{zV/\lambda^3}$ for a fixed volume V as

$$\begin{aligned} Z_{gcp} &= \lim_{V_m \rightarrow \infty} \frac{1}{V_m} \int_0^{V_m} e^{-\beta pV} Z_{gc}(V) dV \\ &= \lim_{V_m \rightarrow \infty} \frac{1}{V_m} \int_0^{V_m} e^{-\beta pV} e^{zV/\lambda^3} dV \\ &= \lim_{V_m \rightarrow \infty} \frac{1}{V_m} \int_0^{V_m} e^{(z/\lambda^3 - \beta p)V} dV. \end{aligned} \quad (10.14)$$

The normalization of the density matrix ($Z_{gcp} = 1$) implies that the intensive variables β , z and p must satisfy the relation

$$\frac{z}{\lambda^3} - \beta p = 0, \quad (10.15)$$

since otherwise Z_{gcp} would be zero or diverge. It remains to be shown that this relation is equivalent to the usual equation of state of the ideal classical gas.

For this purpose we calculate the average number of particles, which is given by

$$\begin{aligned} \langle N \rangle &= \text{Tr}\{\hat{N} \hat{\rho}_{gcp}\} \\ &= \lim_{V_m \rightarrow \infty} \frac{1}{V_m} \int_0^{V_m} e^{-\beta pV} \text{Tr}\{\hat{N} e^{-\beta(\hat{H}-\mu\hat{N})}\} dV \\ &= \lim_{V_m \rightarrow \infty} \frac{1}{V_m} \int_0^{V_m} e^{-\beta pV} Z_{gc}(V) \frac{\text{Tr}\{\hat{N} e^{-\beta(\hat{H}-\mu\hat{N})}\}}{Z_{gc}(V)} dV \\ &= \lim_{V_m \rightarrow \infty} \frac{1}{V_m} \int_0^{V_m} e^{-\beta pV} Z_{gc}(V) \frac{1}{\beta} \frac{\partial \ln Z_{gc}(V)}{\partial \mu} dV \\ &= \lim_{V_m \rightarrow \infty} \frac{1}{V_m} \int_0^{V_m} e^{-\beta pV} Z_{gc}(V) \frac{zV}{\lambda^3} dV \\ &= \frac{z}{\lambda^3} \lim_{V_m \rightarrow \infty} \frac{1}{V_m} \int_0^{V_m} V e^{-\beta pV} Z_{gc}(V) dV \\ &= \frac{z}{\lambda^3} \langle V \rangle, \end{aligned} \quad (10.16)$$

where we have used that $Z_{gc}(V) = e^{zV/\lambda^3}$ and therefore $\partial \ln Z_{gc}(V)/\partial \mu = \frac{V}{\lambda^3} \beta z$. Notice that $\langle N \rangle$ is, as expected, proportional to $\langle V \rangle$ since the intensive variables β , z and p can only define densities, not the size of the system. Finally, replacing Eq. (10.15) in Eq. (10.16) we obtain the equation of state in the usual form

$$p \langle V \rangle = \langle N \rangle k_B T.$$

10.6 Derivation of the classical limit

Far too often we hear and read, from most respected people, the statement that classical mechanics follows from quantum mechanics by taking the limit of Planck's constant $\hbar \rightarrow 0$. And each time I do I am dissatisfied and puzzled. Classical mechanics exists. We observe it in experiment. We know that it works to an extremely high level of accuracy in countless everyday occasions. And still, \hbar is nowhere going to zero. Its value has been the same as far as far back as we can look into the universe's past. I am willing to admit that $\hbar \rightarrow 0$ is a practical shortcut. Indeed, if one wants to recover a classical picture of the universe one can do that by setting $\hbar \rightarrow 0$ in some quantum mechanical relations, basically in order to make Heisenberg's uncertainty principle, the quintessence of QM, disappear. In this way a number of classical features, though not all, are recovered. But still, from a scientific perspective, the argumentation cannot be satisfying.

The purpose of this section is to discuss how to recover the classical limit of statistical mechanics, starting from the quantum theory that we have developed so far, by pointing out what assumptions need to be made, when they are justified, and when they are not. A detailed derivation with strict bounds and error estimates is beyond the scope of this lecture, and is probably even difficult to formulate in rigorous mathematical terms. Nevertheless, discussing the subject is physically worthwhile, since it forces us to deal with the fundamentals of quantum mechanics. Indeed, we shall see that in order to be able to apply the classical limit of statistical mechanics, one needs to circumvent two of the most basic principles of quantum mechanics, namely, the uncertainty principle and the principle of indistinguishability of identical particles. Needless to say that these are also the quantum phenomena which contradict classical determinism in the most open and profound way. The following discussion is intended to show why the classical approximation is valid in the limit of high temperatures and low densities. Moreover, we should be able to understand how these two principles survive in some way, despite our most resolute efforts of classical simplification.

The quantum expression for the canonical partition function is

$$Z_c(N) = \frac{1}{N!} \sum_{k_1=1}^{\infty} \cdots \sum_{k_N=1}^{\infty} \langle k_1, \dots, k_N | e^{-\beta \hat{H}} | k_1, \dots, k_N \rangle, \quad (10.17)$$

where the sums run over a complete set of single-particle orbitals $\varphi_k(x)$ and the symmetrized states $|k_1, \dots, k_N\rangle$ are given by

$$\langle x_1, \dots, x_N | k_1, \dots, k_N \rangle = \Phi^s(x_1 \dots x_N) = \frac{1}{\sqrt{N!}} \sum_P (\pm 1)^P \varphi_{k_1}(x_{P(1)}) \cdots \varphi_{k_N}(x_{P(N)}).$$

The corresponding average values are

$$\langle k_1, \dots, k_N | e^{-\beta \hat{H}} | k_1, \dots, k_N \rangle = \int dx_1 \dots dx_N \Phi^s(x_1, \dots, x_N)^* e^{-\beta \hat{H}} \Phi^s(x_1, \dots, x_N).$$

In the case of bosons the kets $|k_1, \dots, k_N\rangle$ and many-body wave functions $\Phi^s(x_1, \dots, x_N)$ are not normalized to one. They have a square norm equal to $\prod_{\alpha} n_{\alpha}!$ which ensures that

Eq. (10.17) is valid [see the derivation leading to Eq. (9.5) in Sec. 9.3]. Moreover, notice that only $Z_c(N)$ needs to be considered, since $Z_{gc} = \sum_N z^N Z_c(N)$.

The first hurdle is to deal with the impossibility of defining the position and momentum of a particle at the same time or, in more mathematical terms, the consequences of the non-commutativity of position and momentum operators. In order to develop our reasoning we focus first on just one particle in a volume V . The single-particle kets $|\vec{p}\rangle$, having the wave functions

$$\langle \vec{r} | \vec{p} \rangle = \frac{1}{\sqrt{V}} e^{i\vec{p}\cdot\vec{r}}, \quad (10.18)$$

form a complete orthonormal set in the volume V . Thus, the one-particle partition function is given by

$$Z_1 = \sum_{\vec{p}} \langle \vec{p} | e^{-\beta(\frac{\vec{p}^2}{2m} + U(\vec{r}))} | \vec{p} \rangle,$$

where the sum runs over the complete set of linearly independent momenta within the cubic volume $V = L^3$ with side L . According to the theory of Fourier series, the plane waves (10.18) satisfying periodic boundary conditions at the boundaries of the cubic volume form a complete orthonormal basis in the volume V . Any function, including plane waves not satisfying the periodic boundary condition, can be obtained as linear combination of the periodic ones *within the volume* V . This implies that the so-called allowed values of \vec{p} to be taken into account in the sum are given by

$$\frac{p_\alpha}{\hbar} L = 2\pi\nu_\alpha$$

where $\alpha = x, y, z$ and $\nu_\alpha \in \mathbb{Z}$. In a more compact form we may write

$$\vec{p} = \frac{2\pi\hbar}{L} \vec{\nu}, \quad (10.19)$$

where $\vec{\nu} \in \mathbb{Z}^3$ is a vector with integer components. The sum of any well-behaved function $f(\vec{p})$ over all possible \vec{p} is in fact a sum over all integers $\vec{\nu} \in \mathbb{Z}^3$ according to Eq. (10.19):

$$\sum_{\vec{p}} f(\vec{p}) = \sum_{\vec{\nu}} f(\vec{\nu}).$$

In the limit of $V \rightarrow \infty$ the allowed values of \vec{p} form nearly a continuum since the separation between two consecutive values decreases as $1/L$. It is then possible to replace the sum over \vec{p} by an integral over all momenta \vec{p} since the spectrum of the momentum operator tends to a continuum. In order to do that, we need to determine the number of linearly independent momentum eigenstates contained in a volume element d^3p . This is very important in general since the number of allowed \vec{p} increases linearly with V thus ensuring the extensivity of $\ln Z_c$. From Eq. (10.19) one obtains that the number of the summands $d^3\nu$ contained in a volume element d^3p is

$$d^3\nu = d\nu_x d\nu_y d\nu_z = \frac{V}{(2\pi\hbar)^3} d^3p.$$

Assuming that $f(\vec{p})$ remains finite for all \vec{p} when of $V \rightarrow \infty$, which is certainly the case for the Z_c as we approach the classical limit, we may replace the sum over \vec{p} by an integral as follows:

$$\sum_{\vec{p}} f(\vec{p}) = \sum_{\vec{v}} f(\vec{v}) \quad \rightarrow \quad \int f(\vec{v}) d^3\nu = \frac{V}{(2\pi\hbar)^3} \int f(\vec{p}) dp^3.$$

The factor $V/(2\pi\hbar)^3$ represents the number of allowed values of \vec{p} in a volume element d^3p . Should there be terms $f(\vec{p})$ in the sum which diverge in the limit of $V \rightarrow \infty$, one would need to identify them and add them up separately, since isolated points give no contribution to an integral. Later on we shall see, in the context of the Bose-Einstein condensation, that this situation appears when computing the average occupation of the $\vec{p} = 0$ state of bosons at low temperatures.

In order to calculate the classical limit of Z_1 for large V we thus replace the sum by an integral and obtain

$$Z_1 = \frac{V}{(2\pi\hbar)^3} \int d^3p \langle \vec{p} | e^{-\beta(\frac{\hat{p}^2}{2m} + U(\vec{r}))} | \vec{p} \rangle.$$

We may now split the exponential of the sum of the non-commuting kinetic and potential-energy operators in a product of exponentials by using the Zassenhaus formula

$$e^{t(X+Y)} = e^{tX} e^{tY} e^{-\frac{t^2}{2}[X,Y]} e^{-\frac{t^3}{6}(2[Y,[X,Y]] + [X,[X,Y]])} \dots$$

which to lowest order in t reads

$$e^{t(X+Y)} = e^{tX} e^{tY} \left(1 - \frac{t^2}{2}[X, Y] + \dots \right). \quad (10.20)$$

This relation follows from the more widespread Baker-Campbell-Hausdorff equation $e^X e^Y = e^Z$ with $Z = X + Y + [X, Y]/2 + \dots$. Before we can use Eq. (10.20) for $t = \beta$ in the high temperature limit we need the commutator

$$[\hat{p}^2, U(\vec{r})] = -i2\hbar\vec{\nabla}U \cdot \hat{\vec{p}} - \hbar^2\nabla^2U.$$

We may then approximate the partition function to the lowest (second) order in β as

$$Z_1 = \frac{V}{(2\pi\hbar)^3} \int d^3p \langle \vec{p} | e^{-\beta\frac{\hat{p}^2}{2m}} e^{-\beta U(\vec{r})} \left(1 - i\frac{\hbar\beta^2}{2m} \vec{\nabla}U \cdot \hat{\vec{p}} - \frac{\hbar^2\beta^2}{4m} \nabla^2U + \dots \right) | \vec{p} \rangle.$$

Introducing the Dirac's completeness relation $\int d^3r |\vec{r}\rangle\langle\vec{r}| = 1$ between the exponentials we obtain

$$Z_1 = \frac{V}{(2\pi\hbar)^3} \int d^3r \int d^3p \langle \vec{p} | e^{-\beta\frac{\hat{p}^2}{2m}} |\vec{r}\rangle\langle\vec{r}| e^{-\beta U(\vec{r})} \left(1 - i\frac{\hbar\beta^2}{2m} \vec{\nabla}U \cdot \hat{\vec{p}} - \frac{\hbar^2\beta^2}{4m} \nabla^2U + \dots \right) | \vec{p} \rangle.$$

Acting with the momentum and position operators on the bras or kets having well-defined \vec{p} , respectively \vec{r} , as appropriate yields

$$Z_1 = \frac{V}{(2\pi\hbar)^3} \int d^3r \int d^3p e^{-\beta \frac{p^2}{2m}} e^{-\beta U(\vec{r})} \left(1 - i \frac{\hbar\beta^2}{2m} \vec{\nabla}U \cdot \vec{p} - \frac{\hbar^2\beta^2}{4m} \nabla^2 U + \dots \right) |\langle \vec{r} | \vec{p} \rangle|^2.$$

Notice that the second term in brackets vanishes because it is linear in \vec{p} , an expected result since it would not have been real. Taking advantage of the fact that single-electron plane waves satisfy $|\langle \vec{r} | \vec{p} \rangle|^2 = 1/V$ for all \vec{r} and \vec{p} , we obtain Z_1 and its first nonvanishing correction in the high-temperature limit:

$$Z_1 = Z_1^{\text{cl}} \left(1 - \frac{\hbar^2\beta^2}{4m} \langle \nabla^2 U \rangle + \dots \right), \quad (10.21)$$

where

$$Z_1^{\text{cl}} = \frac{1}{(2\pi\hbar)^3} \int d^3r \int d^3p e^{-\beta[\frac{p^2}{2m} + U(\vec{r})]} \quad (10.22)$$

is the classical partition function, and

$$\langle \nabla^2 U \rangle = \frac{\int d^3r e^{-\beta U(\vec{r})} \nabla^2 U}{\int d^3r e^{-\beta U(\vec{r})}} \quad (10.23)$$

is the classical average of the local curvature of the potential U . Notice that the momentum distribution is irrelevant for calculating the classical average of any function of the coordinates. Since $\langle \nabla^2 U \rangle$ is finite and tends to a constant in the high temperature limit, where all positions in the volume V are equally probable, we conclude that the quantum partition function converges to the classical one as $1/T^2$ for large T . The quantum correction is also proportional to \hbar^2/m , which is understandable since in this case it results exclusively from Heisenberg's uncertainty. One may also write $\hbar^2\beta^2/4m = \beta\lambda^2/8\pi$. The particle density v does not appear in the expansion as part of the convergence criterion, since only a single particle has been considered so far.

In the case of N identical particles we follow an analogous procedure. We consider the complete basis of symmetrized (bosons) or antisymmetrized (fermions) states with defined linear momenta $\vec{p}_1, \dots, \vec{p}_N$, which we denote in a compact way by $|p\rangle \equiv |\vec{p}_1, \dots, \vec{p}_N\rangle$, as well the basis of states with defined positions $\vec{r}_1, \dots, \vec{r}_N$, which we denote by $|q\rangle \equiv |\vec{r}_1, \dots, \vec{r}_N\rangle$. Accordingly, $\int dp$ and $\int dq$ stand for the multiple integrals $\prod_i d^3p_i$, respectively $\prod_i d^3r_i$. If present, for example for fermions, the spin projection is also taken into account among the quantum numbers \vec{p}_i and \vec{r}_i defining the kets involved in calculating the trace.

In the limit of large volume V Eq. (10.17) can then be written as

$$Z_c(N) = \frac{1}{N!} \frac{V^N}{(2\pi\hbar)^{3N}} \int dp \langle p | e^{-\beta \hat{H}} | p \rangle,$$

where the Hamiltonian governing the system is given by

$$\hat{H} = \hat{T} + \hat{W} = \sum_i \frac{\hat{p}_i^2}{2m} + W(\vec{r}_1, \dots, \vec{r}_N).$$

The relevant commutator between the kinetic energy \hat{T} and the potential energy \hat{W} is

$$[\hat{T}, \hat{W}] = -i \frac{\hbar}{m} \sum_i \vec{\nabla}_i W \cdot \hat{\vec{p}}_i - \frac{\hbar^2}{2m} \sum_i \nabla_i^2 W,$$

where $\nabla_i^2 W$ stands for the Laplacian of the interaction energy W with respect to the coordinates \vec{r}_i of particle i . Using the Zassenhaus formula we find that to lowest order in β the partition function is given by

$$Z_c(N) = \frac{1}{N!} \frac{V^N}{(2\pi\hbar)^{3N}} \int dp \langle p | e^{-\frac{\beta}{2m} \sum_i \hat{p}_i^2} e^{-\beta \hat{W}(\vec{r}_1 \dots \vec{r}_N)} \times \\ \times \left(1 - i \frac{\hbar\beta^2}{2m} \sum_i \vec{\nabla}_i W \cdot \hat{\vec{p}}_i - \frac{\hbar^2\beta^2}{4m} \sum_i \nabla_i^2 W + \dots \right) | p \rangle.$$

Introducing the Dirac's completeness relation $\int dq |\vec{q}\rangle \langle \vec{q}| = 1$ between the exponentials and acting with the momentum (position) operators $\hat{\vec{p}}_i$ ($\hat{\vec{r}}_i$) either on the bras or on the kets having well-defined \vec{p}_i (\vec{r}_i) we obtain

$$Z_c(N) = \frac{1}{N!} \frac{V^N}{(2\pi\hbar)^{3N}} \int dq \int dp \langle p | e^{-\frac{\beta}{2m} \sum_i \hat{p}_i^2} | q \rangle \langle q | e^{-\beta \hat{W}(\vec{r}_1 \dots \vec{r}_N)} \times \\ \times \left(1 - i \frac{\hbar\beta^2}{2m} \sum_i \vec{\nabla}_i W \cdot \hat{\vec{p}}_i - \frac{\hbar^2\beta^2}{4m} \sum_i \nabla_i^2 W + \dots \right) | p \rangle$$

and

$$Z_c(N) = \frac{1}{N!} \frac{V^N}{(2\pi\hbar)^N} \int dq \int dp e^{-\beta[\sum_i \frac{p_i^2}{2m} + W(\vec{r}_1 \dots \vec{r}_N)]} \left(1 - \frac{\hbar^2\beta^2}{4m} \sum_i \nabla_i^2 W + \dots \right) |\langle q|p\rangle|^2,$$

where we have used that the terms that are linear in the momentum operators vanish, since inverting all \vec{p}_i yields the same kinetic energy and probability distribution $|\langle q|p\rangle|^2$.

This expression can be further simplified by noting that the amplitude of the normalized plane waves (10.18) scales as $1/\sqrt{V}$. Therefore $|\langle q|p\rangle|^2 \propto 1/V^N$ which cancels the factor V^N . It therefore meaningful to change the normalization of the single-particle wave functions entering the symmetrized N -particle kets $|p\rangle$ by using the plane waves

$$\langle \vec{r} | \vec{p} \rangle = e^{\frac{i}{\hbar} \vec{p} \cdot \vec{r}}, \quad (10.24)$$

whose amplitude is independent of the volume V and which satisfy the simpler relation $|\langle \vec{r} | \vec{p} \rangle|^2 = 1$. Denoting the corresponding symmetrized states by $|\tilde{p}\rangle$ we have

$$|\tilde{p}\rangle = V^{N/2} |p\rangle$$

and therefore

$$Z_c(N) = \frac{1}{N!} \frac{1}{(2\pi\hbar)^N} \int dq \int dp e^{-\beta[\sum_i \frac{p_i^2}{2m} + W(\vec{r}_1 \dots \vec{r}_N)]} \left(1 - \frac{\hbar^2\beta^2}{4m} \sum_i \nabla_i^2 W + \dots \right) |\langle q|\tilde{p}\rangle|^2. \quad (10.25)$$

The above semiclassical approximation of $Z_c(N)$ starts to closely resemble the classical partition function (10.1). It allows us to obtain the classical limit and to identify the origin of the leading quantum corrections. The first one is given by

$$-\frac{\hbar^2\beta^2}{4m}N\langle\nabla_i^2W\rangle = -\frac{\hbar^2\beta^2}{4m}N\frac{\int dq e^{-\beta W(q)}\nabla_i^2W}{\int dq e^{-\beta W(q)}}, \quad (10.26)$$

which has its origin in Heisenberg's uncertainty principle or, in other words, in the impossibility that the momentum and position of any particle take arbitrarily sharp values. Notice that $\langle\nabla_i^2W\rangle$ represents the classical statistical average of the curvature of the interaction energy felt by each particle. It is independent of i since all particles are identical. This contribution is proportional to \hbar^2 and vanishes with increasing temperature as $1/T^2$, thus showing the expected convergence to the classical expression at high temperatures. An analogous correction was found for a single particle in an external potential [see Eqs. (10.21)–(10.23)]. This quantum correction can be regarded as the consequence of the motion of each individual particle in the field created by the others. Consequently, the volume per particle $v = V/N$ and the average interparticle distance $v^{1/3}$ also play an important role. Indeed, larger v implies weaker average interactions $\langle W\rangle$, as well as a longer length scale for its spatial variations, which both contribute to reducing the average local curvature $\langle\nabla_i^2W\rangle$ until, in the limit of $v \rightarrow \infty$, $\langle\nabla_i^2W\rangle$ vanishes altogether [see Eq (10.26)]. Therefore, the classical limit is not only a high-temperature approximation, but also a low-density one. For any finite T , whatever high, the quantum effects cease to be negligible at a sufficiently high density $1/v$. Conversely, the validity of the classical approximation can be pushed to lower temperatures (but never for $T \rightarrow 0!$) provided that the density is sufficiently low.

The second quantum correction is a more subtle consequence of the fact that in a fully symmetrized or antisymmetrized quantum state, the joint probability distribution $|\langle q|p\rangle|^2 = |\langle\vec{r}_1, \dots, \vec{r}_N|\vec{p}_1, \dots, \vec{p}_N\rangle|^2$ of finding the particles at the positions $\vec{r}_1, \dots, \vec{r}_N$ depends on both the point $q = (\vec{r}_1, \dots, \vec{r}_N)$ as well as on the occupied momenta $p = (\vec{p}_1, \dots, \vec{p}_N)$. One says that the particles are correlated by their symmetry with respect to interchange (bosonic or fermionic). This phenomenon is a clear manifestation of the indistinguishability of identical particles, a truly many-body quantum behavior which lies well beyond the scope of a classical description. It is of course absent in the single particle case, where $|\langle\vec{r}|\vec{p}\rangle|^2 = 1/V$ or 1 depending on the chosen normalization [see Eqs. (10.18) and (10.24)]. Moreover, it is important to note that these symmetry correlations are also present in the absence of interactions W . Therefore, one may determine their effect analytically by setting $W = 0$, in which case it is possible to calculate the exact quantum partition function and thermodynamic properties for all temperatures and densities. We shall develop this approach in the following chapters.

In order to understand the classical limit of $|\langle q|p\rangle|^2$ and the consequences of the symmetry with respect to interchange in the high-temperature low-density limit, we consider the joint particle distribution in a symmetrized product state $\Phi^s(x_1, \dots, x_N)$, in which the occupied orbitals are not necessarily plane waves but arbitrary single particle orbitals $\varphi_1(x), \dots, \varphi_N(x)$. Thus, we replace the ket $|p\rangle = |\vec{p}_1, \dots, \vec{p}_N\rangle$, in which the oc-

cupied single-particle states have defined momenta $\vec{p}_1, \dots, \vec{p}_N$, by the more general ket $|\varphi\rangle = |\varphi_1, \dots, \varphi_N\rangle$. As in chapter 9 we denote by $x_i \equiv (\vec{r}_i, \sigma_i)$ the position and spin quantum numbers of particle i . Thus, the N -particle ket $|x\rangle \equiv |x_1, \dots, x_N\rangle$ with well defined coordinates x_1, \dots, x_N takes the role of $|q\rangle \equiv |\vec{r}_1, \dots, \vec{r}_N\rangle$ and the wave function $\langle x|\varphi\rangle$ corresponds to $\langle q|p\rangle$. The fully symmetrized wave functions are given by

$$\begin{aligned}\Phi^s(x_1, \dots, x_N) &= \langle x_1, \dots, x_N | \varphi_1, \dots, \varphi_N \rangle \equiv \langle x | \varphi \rangle \\ &= \frac{1}{\sqrt{N!}} \sum_P (\pm 1)^p \varphi_1(x_{P(1)}), \dots, \varphi_N(x_{P(N)}) \\ &= \frac{1}{\sqrt{N!}} \sum_P (\pm 1)^p \varphi_{P(1)}(x_1), \dots, \varphi_{P(N)}(x_N),\end{aligned}$$

where the sum runs over all permutations P in the natural interval $[1, N]$, p is the order of the permutation P , and the $+$ ($-$) sign applies to bosons (fermions). The corresponding joint probability distribution of finding the particle at the coordinates x_1, \dots, x_N is

$$\begin{aligned}|\langle x | \varphi \rangle|^2 &= |\langle x_1, \dots, x_N | \varphi_1, \dots, \varphi_N \rangle|^2 \\ &= \frac{1}{N!} \sum_{P, Q} (\pm 1)^{p+q} \varphi_{P(1)}(x_1)^* \dots \varphi_{P(N)}(x_N)^* \varphi_{Q(1)}(x_1) \dots \varphi_{Q(N)}(x_N) \\ &= \frac{1}{N!} \sum_P |\varphi_{P(1)}(x_1)|^2 \dots |\varphi_{P(N)}(x_N)|^2 +\end{aligned}\tag{10.27}$$

$$+ \frac{1}{N!} \sum_{P \neq Q} (\pm 1)^{p+q} \varphi_{P(1)}(x_1)^* \varphi_{Q(1)}(x_1) \dots \varphi_{P(N)}(x_N)^* \varphi_{Q(N)}(x_N).\tag{10.28}$$

One recognizes two types of terms. In (10.27) the permutations P and Q are identical, the phase factor $(\pm 1)^{p+q}$ disappears and with it any differences between fermionic and bosonic behavior. This contribution is simply the product of the probability densities $|\varphi_\alpha(x)|^2$ of finding a particle in the orbital α , yet symmetrized in order that the same result is obtained regardless of which particle has which coordinates x_i . In contrast, in (10.28) the nature of the particles is crucial. Here we find at least one i for which $P(i) \neq Q(i)$ and therefore the overlap $\varphi_{P(i)}(x_i)^* \varphi_{Q(i)}(x_i)$ and the phase factor $(\pm 1)^{p+q}$ matter. This expression is quite complicated in general. However, it simplifies enormously in the particular case where the orbitals $\varphi_\alpha(x)$ that the particles occupy do not overlap at all, i.e., if the particles always (or almost always) occupy completely different regions in space [i.e., $\varphi_\alpha(x) \varphi_\beta(x) = 0, \forall \alpha \neq \beta$ and $\forall x$] which corresponds to the classical notion of distinguishability. Notice that this presupposes that also for bosons no orbital is occupied more than once. In this case the contribution (10.28) where $P \neq Q$ vanishes altogether, since $P(i) \neq Q(i)$ for some i and therefore $\varphi_{P(i)}(x)^* \varphi_{Q(i)}(x) = 0$ for all x . In this case the joint probability distribution reads

$$|\langle x_1 \dots x_N | \varphi_1 \dots \varphi_N \rangle|^2 = \frac{1}{N!} \sum_P |\varphi_{P(1)}(x_1)|^2 \dots |\varphi_{P(N)}(x_N)|^2.$$

If the particles are localized in wave packets occupying different regions in space, the only consequence of the indistinguishability of identical particles and of the associated symmetry of the many-body wave functions is that only one distinct many-body state (and not $N!$) can be constructed when a set of N orbitals $\varphi_1, \dots, \varphi_N(x)$ are occupied. This provides a justification for the Boltzmann counting introduced *ad hoc* in Sec. 10.1 as a shortcut to the classical limit. Furthermore, the above analysis shows that in the classical limit, when the particles states can be approximated by non-overlapping orbitals, the bosonic or fermionic nature of the wave-function symmetry plays no role. Moreover, $|\langle x_1, \dots, x_N | \varphi_1, \dots, \varphi_N \rangle|^2$ is simply given by the symmetrized product of the probabilities of finding the particles at the points x_1, \dots, x_N . In the context of Eq. (10.25) this means that the classical limit of the probability density distribution $|\langle q | \vec{p} \rangle|^2 = 1$, where we have used that $|\langle \vec{r} | \vec{p} \rangle|^2 = 1$ [see also Eq. (10.24)]. In this way the classical partition function (10.1) is finally obtained.

Before closing it is important to stress that the accuracy of the assumption of non-overlapping single particle states, and thus of the classical approximation to $Z_c(N)$, depends both on the density $1/v$ or average interparticle distance $v^{1/3}$ as well as on the temperature T . Clearly, the higher the particle density is, the stronger is the localization which has to be imposed to the single-particle orbitals in order to avoid any overlaps. As the average interparticle distance decreases, the spatial extension Δr of the orbitals φ_α must also decrease roughly as $v^{1/3}$. Consequently, the kinetic energy of these orbitals increases roughly as $\Delta E = \frac{\Delta p^2}{2m} \sim \frac{\hbar^2}{2m\Delta r^2} \sim \frac{\hbar^2}{2m} v^{-2/3}$. This localization-induced energy enhancement ΔE has little effect on the calculation of $Z_c(N)$ when the temperature T is high enough, since in this case large values of the kinetic energy $p^2/2m$ dominate and any energy differences that are much smaller than $k_B T$ turn out to be irrelevant. Quantitatively, this condition is satisfied when $k_B T \gg \Delta E$ or equivalently $\lambda^2/v^{2/3} \ll 1$, which corresponds to (10.2). Conversely, the assumption of localized orbital, which as we have seen is inherent to the classical approximation and to the very notion of classical distinguishability, becomes unacceptable when the temperature is low and λ is comparable to $v^{1/3}$. In this case, small momenta having a large spatial extension and a small kinetic energy are the most probable ones. In the low temperature regime the indistinguishability of identical particles remains crucial in every respect.

11 The ideal quantum gases

The simplest system of N identical particles is found in the absence of interactions. In this case the Hamiltonian is given by

$$\hat{H} = \sum_{i=1}^N \frac{\hat{p}_i^2}{2m},$$

where m is the particles' rest mass. The case of photons will be discussed separately. In order to compute the partition function we need to find the eigenstates of the system, or at least we need to classify them by finding the appropriate quantum numbers, the corresponding eigenvalues and their degeneracies. In the following we assume for simplicity that the particles are spinless. In other words, we ignore the particles' spin degree of freedom and the degeneracies associated with it, even in the case of fermions.

Since there are no interactions among the particles, the Hamiltonian is simply the sum of the kinetic energy operators $\hat{p}_i^2/2m$, each acting on a different variable \vec{r}_i . Under these circumstances the eigenfunctions of \hat{H} are products of single-particle wave functions $u_\alpha(\vec{r})$, each one depending on the coordinates \vec{r}_i of the i th electron and being eigenfunctions of $\hat{p}_i^2/2m$. Any simple product, however, needs to be symmetrized (antisymmetrized) in order to comply with the symmetry requirements on the wave functions of indistinguishable bosons (fermions). The general form of the N -particle eigenfunctions is

$$\Psi_{\{n_{\vec{p}}\}}(x_1, \dots, x_N) = \frac{1}{\sqrt{N!}} \sum_P (\pm 1)^P u_{\vec{p}_1}(x_{P(1)}) u_{\vec{p}_2}(x_{P(2)}) \dots u_{\vec{p}_N}(x_{P(N)}), \quad (11.1)$$

where the $+$ ($-$) sign corresponds to bosons (fermions), the single-particle wave functions are the known plane waves

$$u_{\vec{p}}(\vec{r}) = \frac{1}{\sqrt{V}} e^{i\vec{p}\cdot\vec{r}}, \quad (11.2)$$

and $\vec{p}_1, \vec{p}_2, \dots, \vec{p}_N$ are the momenta of the particles. One usually says that $\vec{p}_1, \vec{p}_2, \dots, \vec{p}_N$ are the momenta which are occupied in the state $\Psi_{\{n_{\vec{p}}\}}$ since $n_{\vec{p}} \neq 0$ only for $\vec{p} = \vec{p}_j$. Notice that the norm of $\Psi_{\{n_{\vec{p}}\}}$ is not equal to 1 in the case of bosons having $n_{\vec{p}} \geq 2$ for some \vec{p} , even though the single-particle states $u_{\vec{p}}(\vec{r})$ are normalized. As already discussed, $\Psi_{\{n_{\vec{p}}\}}$ depends only on the set of occupation numbers $\{n_{\vec{p}}\}$ which count how often the single-particle state $u_{\vec{p}}(\vec{r})$ appears on the right hand side of (11.1). For bosons all positive or zero values of $n_{\vec{p}}$ are allowed ($n_{\vec{p}} = 0, 1, 2, \dots$), while for fermions $n_{\vec{p}}$ can only take the values 0 or 1 (Pauli exclusion principle). In all cases the sum of all occupation numbers must be equal to the number of particles:

$$N = \sum_{\vec{p}} n_{\vec{p}}, \quad (11.3)$$

where the sum runs over all possible values of the momentum of a single particle in the volume V . The fact that the eigenstates of \hat{H} depend only on $\{n_{\vec{p}}\}$ can be stressed by

introducing the occupation number representation of the many-body states. Thus, the kets

$$|\{n_{\vec{p}}\}\rangle = |n_1, n_2, \dots, n_\infty\rangle$$

have definite occupation numbers of all single-particle eigenstates of the momentum operator \hat{p} . The normalized wave function corresponding to $|\{n_{\vec{p}}\}\rangle$ is given by

$$\langle x_1, \dots, x_N | \{n_{\vec{p}}\} \rangle = \frac{1}{\sqrt{\prod_{\vec{p}} n_{\vec{p}}!}} \Psi_{\{n_{\vec{p}}\}}(x_1, \dots, x_N). \quad (11.4)$$

The states $|n_1, n_2, \dots, n_\infty\rangle$ constitute a complete orthonormal set satisfying

$$\langle n_1, n_2, \dots, n_\infty | n'_1, n'_2, \dots, n'_\infty \rangle = \delta_{n_1 n'_1} \dots \delta_{n_\infty n'_\infty}.$$

Since $u_{\vec{p}}(r)$ is an eigenfunction of \hat{p} we have

$$\frac{\hat{p}^2}{2m} u_{\vec{p}'} = \frac{p'^2}{2m} u_{\vec{p}'} = \varepsilon_{\vec{p}'} u_{\vec{p}'}. \quad (11.5)$$

This implies that the N -particle wave function satisfies

$$\hat{H} \Psi_{\{n_{\vec{p}}\}} = E_{\{n_{\vec{p}}\}} \Psi_{\{n_{\vec{p}}\}} \quad (11.6)$$

where the eigenenergy

$$E_{\{n_{\vec{p}}\}} = \sum_{\vec{p}} n_{\vec{p}} \varepsilon_{\vec{p}} \quad (11.7)$$

is the sum of the single-particle eigenenergies $\varepsilon_{\vec{p}} = p^2/2m$ of the occupied states $u_{\vec{p}}$. In the occupation number representation we have

$$\hat{H} |n_1, n_2, \dots, n_\infty\rangle = \left(\sum_i n_i \varepsilon_i \right) |n_1, n_2, \dots, n_\infty\rangle \quad (11.8)$$

where $i = 1, 2, 3, \dots$ is an index that enumerates the different possible values of the momentum \vec{p} . Alternatively, we can also write

$$\hat{H} |\{n_{\vec{p}}\}\rangle = \left(\sum_{\vec{p}} n_{\vec{p}} \varepsilon_{\vec{p}} \right) |\{n_{\vec{p}}\}\rangle. \quad (11.9)$$

In order to determine the allowed values of \vec{p} , we consider a cubic volume $V = L^3$ with side L . A complete orthonormal basis of eigenfunctions of the momentum operator \hat{p} is given by

$$u_{\vec{p}} = \frac{1}{\sqrt{V}} e^{i\vec{p}\cdot\vec{r}},$$

where the values of \vec{p} satisfy periodic boundary condition

$$\frac{p_\alpha}{\hbar}L = 2\pi\nu_\alpha$$

where $\alpha = x, y, z$ and $\nu_\alpha \in \mathbb{Z}$. This constitutes a complete set of single-particle wave functions in the volume V . In a more compact form we may write

$$\vec{p} = \frac{2\pi\hbar}{L}\vec{\nu} \quad (11.10)$$

where $\vec{\nu} \in \mathbb{Z}^3$ is a vector with integer components. In the limit of $V \rightarrow \infty$ the allowed values of \vec{p} form nearly a continuum. This will allow us to replace a sum of a function $f(\vec{p})$ over all possible \vec{p} by the corresponding integral in momentum space:

$$\sum_{\vec{p}} f(\vec{p}) \rightarrow \frac{V}{(2\pi\hbar)^3} \int f(\vec{p}) d^3p, \quad (11.11)$$

provided that all the terms in the sum over \vec{p} are finite. Later on we shall see, in the context of the Bose-Einstein condensation, that this condition is not always satisfied. In such cases the diverging term or terms (e.g., $\vec{p} = 0$) need to be identified and added up separately. The factor $V/(2\pi\hbar)^3$ gives the number of allowed values of \vec{p} in a volume element d^3p . Formally, we may write

$$\sum_{\vec{p}} f(\vec{p}) = \sum_{\vec{\nu}} f(\vec{\nu}) \rightarrow \int f(\vec{\nu}) d^3\nu = \frac{V}{(2\pi\hbar)^3} \int f(\vec{p}) dp^3$$

since

$$\vec{\nu} = \frac{L}{2\pi\hbar} \vec{p} \Rightarrow d^3\nu = d\nu_x d\nu_y d\nu_z = \frac{V}{(2\pi\hbar)^3} d^3p.$$

Computing the canonical partition function $Z_c(N)$ for a fixed number of particles is quite complicated, since we would need to impose the constraint

$$N = \sum_{\vec{p}} n_{\vec{p}}$$

on the occupation numbers $n_{\vec{p}}$. However, this restriction is absent in the grand canonical ensemble, where all possible values of $n_{\vec{p}}$ and all total numbers of particles N are *a priori* possible. The grand canonical partition function is given by

$$\begin{aligned} Z_{gc} &= \text{Tr}\{e^{-\beta(\hat{H}-\mu\hat{N})}\} \\ &= \sum_{\{n_{\vec{p}}\}} e^{-\beta \sum_{\vec{p}} (\epsilon_{\vec{p}} - \mu) n_{\vec{p}}}, \end{aligned} \quad (11.12)$$

where we have used that in $|\{n_{\vec{p}}\}\rangle$ the number of particles is

$$N = \sum_{\vec{p}} n_{\vec{p}} \quad (11.13)$$

and the eigenenergy is

$$E = \sum_{\vec{p}} \varepsilon_{\vec{p}} n_{\vec{p}}. \quad (11.14)$$

The sum in Eq. (11.12) runs over all states $|\{n_{\vec{p}}\}\rangle$ with defined occupations $\{n_{\vec{p}}\}$. Notice that each set of occupations $\{n_{\vec{p}}\}$ defines one and only one distinct state. Therefore, we must consider all possible values of $n_{\vec{p}}$ for each \vec{p} . To this aim we enumerate the different \vec{p} in some arbitrary way according to the index i , so that $\vec{p} \leftrightarrow i$ and $\{n_{\vec{p}}\} = \{n_1, n_2, \dots\}$. The partition function is then given by

$$\begin{aligned} Z_{gc} &= \sum_{n_1} \sum_{n_2} \dots e^{-\beta \sum_i (\varepsilon_i - \mu) n_i} \\ &= \left(\sum_{n_1} e^{-\beta(\varepsilon_1 - \mu) n_1} \right) \left(\sum_{n_2} e^{-\beta(\varepsilon_2 - \mu) n_2} \right) \dots \\ &= \prod_{\vec{p}} \left[\sum_n e^{-\beta(\varepsilon_{\vec{p}} - \mu) n} \right] \\ &= \prod_{\vec{p}} \left[\sum_n \left(e^{-\beta(\varepsilon_{\vec{p}} - \mu)} \right)^n \right]. \end{aligned}$$

The sum over the occupation number n is of course different for fermions ($n = 0, 1$) and bosons ($n \in \mathbb{N}$). For fermions we have simply

$$Z_{gc} = \prod_{\vec{p}} \left(1 + e^{-\beta(\varepsilon_{\vec{p}} - \mu)} \right). \quad (11.15)$$

For bosons we obtain

$$Z_{gc} = \prod_{\vec{p}} \frac{1}{(1 - e^{-\beta(\varepsilon_{\vec{p}} - \mu)})} \quad (11.16)$$

after summation of a geometric series, whose convergence requires $e^{-\beta(\varepsilon_{\vec{p}} - \mu)} < 1$, or equivalently, $\mu \leq 0$ since $\varepsilon_{\vec{p}} \geq 0$.

In order to facilitate the mathematical manipulations it is useful to replace the chemical potential μ by the fugacity

$$z = e^{\beta\mu}$$

as one of the independent intensive variables defining Z_{gc} . In terms of z we have

$$Z_{gc} = \begin{cases} \prod_{\vec{p}} (1 + z e^{-\beta\varepsilon_{\vec{p}}}) & \text{for fermions} \\ \prod_{\vec{p}} (1 - z e^{-\beta\varepsilon_{\vec{p}}})^{-1} & \text{for bosons.} \end{cases} \quad (11.17)$$

The grand canonical potential is thus given by

$$\Phi = \begin{cases} -k_B T \sum_{\vec{p}} \ln [1 + e^{-\beta(\varepsilon_{\vec{p}} - \mu)}] & \text{for fermions} \\ k_B T \sum_{\vec{p}} \ln [1 - e^{-\beta(\varepsilon_{\vec{p}} - \mu)}] & \text{for bosons.} \end{cases} \quad (11.18)$$

Moreover, since $\Phi = -pV$, the pressure-volume relation reads

$$\frac{pV}{k_B T} = \begin{cases} \sum_{\vec{p}} \ln [1 + e^{-\beta(\varepsilon_{\vec{p}} - \mu)}] & \text{for fermions} \\ -\sum_{\vec{p}} \ln [1 - e^{-\beta(\varepsilon_{\vec{p}} - \mu)}] & \text{for bosons.} \end{cases} \quad (11.19)$$

The chemical potential μ , or the fugacity z , are defined by the average number of particles. In the case of fermions we have

$$\begin{aligned} N &= -\frac{\partial \Phi}{\partial \mu} \\ &= \sum_{\vec{p}} \frac{e^{-\beta(\varepsilon_{\vec{p}} - \mu)}}{1 + e^{-\beta(\varepsilon_{\vec{p}} - \mu)}} \\ &= \sum_{\vec{p}} \frac{1}{e^{\beta(\varepsilon_{\vec{p}} - \mu)} + 1} \\ &= \sum_{\vec{p}} \frac{1}{z^{-1} e^{\beta \varepsilon_{\vec{p}}} + 1}. \end{aligned} \quad (11.20)$$

Analogously for bosons we have

$$\begin{aligned} N &= \sum_{\vec{p}} \frac{e^{-\beta(\varepsilon_{\vec{p}} - \mu)}}{1 - e^{-\beta(\varepsilon_{\vec{p}} - \mu)}} \\ &= \sum_{\vec{p}} \frac{1}{e^{\beta(\varepsilon_{\vec{p}} - \mu)} - 1} \\ &= \sum_{\vec{p}} \frac{1}{z^{-1} e^{\beta \varepsilon_{\vec{p}}} - 1}. \end{aligned} \quad (11.21)$$

These equations need to be solved to find $\mu(N)$ or $z(N)$, which can then be replaced in (11.19) in order to obtain the equations of state relating p , T and N/V .

The average occupation of each single-particle state \vec{p} can be calculated as

$$\begin{aligned} \langle n_{\vec{p}} \rangle &= \text{Tr} \{ \hat{n}_{\vec{p}} \hat{\rho}_{gc} \} \\ &= \frac{\text{Tr} \{ \hat{n}_{\vec{p}} e^{-\beta(\hat{H} - \hat{N})} \}}{Z_{gc}} \\ &= \frac{\text{Tr} \left\{ \hat{n}_{\vec{p}} e^{-\beta \sum_{\vec{p}'} (\varepsilon_{\vec{p}'} - \mu) \hat{n}_{\vec{p}'}} \right\}}{Z_{gc}} \\ &= -\frac{1}{\beta} \frac{\partial}{\partial \varepsilon_{\vec{p}}} \ln Z_{gc} \\ &= \frac{\partial \Phi}{\partial \varepsilon_{\vec{p}}}, \end{aligned}$$

which yields

$$\langle n_{\vec{p}} \rangle = \begin{cases} \frac{1}{e^{\beta(\varepsilon_{\vec{p}} - \mu)} + 1} & \text{for fermions} \\ \frac{1}{e^{\beta(\varepsilon_{\vec{p}} - \mu)} - 1} & \text{for bosons.} \end{cases}$$

Comparison with Eqs. (11.20) and (11.21) shows, as expected, that $\sum_{\vec{p}} \langle n_{\vec{p}} \rangle = N$. Here we may also confirm that $e^{-\beta(\varepsilon_{\vec{p}} - \mu)} = z e^{-\beta\varepsilon_{\vec{p}}} < 1$ for bosons, as requested in order that the geometric series leading to (11.16) converges, since $\hat{n}_{\vec{p}}$ has only positive or zero eigenvalues. In other terms, $\langle n_{\vec{p}} \rangle \geq 0 \Leftrightarrow e^{\beta(\varepsilon_{\vec{p}} - \mu)} \geq 1 \Leftrightarrow \mu \leq \varepsilon_{\vec{p}}$ for all \vec{p} , which implies that $\mu \leq 0$, or equivalently, $z \leq 1$. Notice that if $\mu \rightarrow 0$, $\langle n_{\vec{p}} \rangle$ diverges for $\vec{p} = 0$. However, the density N/V is necessarily finite. Therefore $\langle n_{\vec{p}} \rangle/V$ is always finite for all \vec{p} . For fermions $\langle n_{\vec{p}} \rangle$ itself is obviously always finite for all \vec{p} .

In order to pursue the discussion we take the limit $V \rightarrow \infty$ and replace

$$\sum_{\vec{p}} \rightarrow \frac{V}{(2\pi\hbar)^3} \int d^3p$$

whenever this is possible. In fact, such a replacement is valid as long as the terms in the sum remain finite in the limit of $V \rightarrow \infty$. If this is not the case one must single out the terms that risk to diverge and sum them separately. In the following chapters the Fermi and Bose gases are discussed specifically in some detail.

12 Fermi systems: The ideal Fermi gas

In order to analyze the properties of the ideal Fermi gas in more detail, it is convenient to consider the fugacity $z = e^{\beta\mu}$, T and V as the thermodynamic variables. The fugacity $z = e^{\beta\mu}$ is an intensive variable which, like μ , defines the particle number N at the given temperature T . Alternatively, one may regard z as a function of T and N .

Let us examine the equation of state

$$\frac{pV}{k_B T} = \sum_{\vec{p}} \ln \left[1 + z e^{-\beta \varepsilon_{\vec{p}}} \right]. \quad (12.1)$$

Since $z = e^{\beta\mu} > 0$ and $e^{-\beta \varepsilon_{\vec{p}}} > 0$, all the terms in the sum are finite. We may then write

$$\frac{pV}{k_B T} = \frac{V}{(2\pi\hbar)^3} \int \ln \left[1 + z e^{-\beta p^2/2m} \right] d^3 p \quad (12.2)$$

$$= \frac{V}{(2\pi\hbar)^3} 4\pi \int_0^\infty p^2 \ln \left[1 + z e^{-\beta p^2/2m} \right] dp. \quad (12.3)$$

Introducing the dimensionless integration variable $x = \frac{p}{\sqrt{2mk_B T}}$ in Eq. (12.3) we have

$$\frac{p}{k_B T} = \frac{4\pi}{(2\pi\hbar)^3} \left(\sqrt{2mk_B T} \right)^3 \int_0^\infty x^2 \ln \left[1 + z e^{-x^2} \right] dx. \quad (12.4)$$

Recalling that $\lambda = \frac{2\pi\hbar}{\sqrt{2\pi mk_B T}}$ is the thermal wave length, the equation of state takes the form

$$\frac{p}{k_B T} = \frac{1}{\lambda^3} f_{5/2}(z), \quad (12.5)$$

where

$$f_{5/2}(z) = \frac{4}{\sqrt{\pi}} \int_0^\infty x^2 \ln \left[1 + z e^{-x^2} \right] dx = \sum_{l=1}^\infty (-1)^{l-1} \frac{z^l}{l^{5/2}} \quad (12.6)$$

is one of Riemann's z -functions. Notice that the integral is well defined for all $z \geq 0$ and that $f_{5/2}(0) = 0$. However, the Taylor expansion converges only for $|z| \leq 1$.

In order to obtain the Taylor expansion of $f_{5/2}(z)$ for small z we develop $\ln(1 + \alpha z)$ around $z = 0$. The derivatives are

$$\ln(1 + \alpha z) \xrightarrow{\frac{d}{dz}} \frac{\alpha}{1 + \alpha z} \xrightarrow{(2)} -\frac{\alpha^2}{(1 + \alpha z)^2} \xrightarrow{(3)} \frac{2\alpha^3}{(1 + \alpha z)^3} \rightarrow \dots \quad (12.7)$$

In general we have for $l \geq 1$

$$\frac{d^l}{dz^l} \ln(1 + \alpha z) = \frac{(-1)^{l-1} (l-1)! \alpha^l}{(1 + \alpha z)^l}$$

and for $z = 0$

$$\left. \frac{d^l}{dz^l} \ln(1 + \alpha z) \right|_{z=0} = (-1)^{l-1} (l-1)! \alpha^l.$$

Since we are interested in $f_{5/2}(z)$ we set $\alpha = e^{-x^2}$. Therefore, we need the integrals

$$\frac{4}{\sqrt{\pi}} \int_0^\infty x^2 e^{-lx^2} dx = \frac{4}{\sqrt{\pi}} \frac{1}{l^{3/2}} \underbrace{\int_0^\infty y^2 e^{-y^2} dy}_{\sqrt{\pi}/4} = \frac{1}{l^{3/2}},$$

where we have introduced $y = \sqrt{l}x$ and $dx = dy/\sqrt{l}$. Since

$$f_{5/2}(z) = \frac{4}{\sqrt{\pi}} \int_0^\infty dx x^2 \ln(1 + ze^{-x^2})$$

we have

$$\left. \frac{d^l f_{5/2}(z)}{dz^l} \right|_{z=0} = (-1)^{l-1} (l-1)! \underbrace{\frac{4}{\sqrt{\pi}} \int_0^\infty x^2 e^{-lx^2} dx}_{l^{-3/2}} = (-1)^{l-1} (l-1)! \frac{1}{l^{3/2}} = (-1)^{l-1} \frac{l!}{l^{5/2}}$$

for $l \geq 1$. Recalling that $f_{5/2}(0) = 0$, we finally have

$$f_{5/2}(z) = \sum_{l=1}^{\infty} (-1)^{l-1} \frac{z^l}{l^{5/2}} = z - \frac{z^2}{4\sqrt{2}} + \frac{z^3}{9\sqrt{3}} - \dots$$

The equation for the number of particles reads

$$N = \sum_{\vec{p}} \frac{1}{e^{\beta(\varepsilon_{\vec{p}} - \mu)} + 1} = \sum_{\vec{p}} \frac{1}{z^{-1} e^{\beta\varepsilon_{\vec{p}}} + 1}.$$

Since $z = e^{\beta\mu} > 0$ all terms in the sum are finite. Thus, for $V \rightarrow \infty$ we may write

$$N = \frac{V}{(2\pi\hbar)^3} 4\pi \int_0^\infty p^2 \frac{1}{z^{-1} e^{\beta\varepsilon_p} + 1} dp$$

and

$$\frac{N}{V} = \frac{1}{v} = \frac{1}{\lambda^3} f_{3/2}(z), \quad (12.8)$$

where

$$\begin{aligned} f_{3/2}(z) &= \frac{4}{\sqrt{\pi}} \int_0^\infty \frac{x^2}{z^{-1} e^{x^2} + 1} dx \\ &= \frac{4}{\sqrt{\pi}} \int_0^\infty \frac{zx^2}{e^{x^2} + z} dx. \end{aligned} \quad (12.9)$$

Notice that

$$\begin{aligned} z \frac{df_{5/2}}{dz} &= z \frac{4}{\sqrt{\pi}} \int_0^{\infty} x^2 \frac{e^{-x^2}}{1 + ze^{-x^2}} dx \\ &= \frac{4}{\sqrt{\pi}} \int_0^{\infty} x^2 \frac{1}{z^{-1}e^{x^2} + 1} dx = f_{3/2}(z). \end{aligned}$$

Consequently,

$$f_{3/2}(z) = z \frac{df_{5/2}}{dz} = \sum_{l=1}^{\infty} (-1)^{l-1} \frac{z^l}{l^{3/2}} \quad (12.10)$$

Both $f_{5/2}$ and $f_{3/2}$ have convergent Taylor expansions for $z \rightarrow 0$:

$$f_{3/2}(z) \approx z - \frac{z^2}{2\sqrt{2}} + \dots$$

and

$$f_{5/2}(z) \approx z - \frac{z^2}{4\sqrt{2}} + \dots$$

with the convergence radius $|z| = 1$.

12.1 Energy-pressure relation

An interesting relation may be derived between the average energy $E = \langle \hat{H} \rangle$ and the pressure p of an ideal fermionic gas. On the one hand we know that

$$pV = -\Phi = k_B T \ln Z_{gc}.$$

Therefore, using Eq. (12.5) we have

$$\ln Z_{gc} = \frac{pV}{k_B T} = \frac{V}{\lambda^3} f_{5/2}(z). \quad (12.11)$$

On the other hand, we know that

$$Z_{gc} = \text{Tr} \left\{ \sum_N z^N e^{-\beta \hat{H}} \right\},$$

which implies that

$$-\frac{\partial \ln Z_{gc}}{\partial \beta} \Big|_{z,V} = \frac{1}{Z_{gc}} \text{Tr} \left\{ \sum_N z^N e^{-\beta \hat{H}} \hat{H} \right\} = \langle E \rangle. \quad (12.12)$$

Notice that z and not μ is kept fixed in the partial derivation. Combining (12.11) and (12.12), and recalling that $\beta \frac{\partial}{\partial \beta} \left(\frac{1}{\lambda^3} \right) = -T \frac{\partial}{\partial T} \left(\frac{1}{\lambda^3} \right) = -\frac{3}{2} \frac{1}{\lambda^3}$, where $\lambda = 2\pi\hbar/\sqrt{2\pi m k_B T} \propto T^{-1/2}$, we have

$$\begin{aligned} \langle E \rangle &= -\frac{\partial \ln Z_{gc}}{\partial \beta} \\ &= -V f_{5/2}(z) \frac{1}{\beta} \left(-\frac{3}{2} \right) \frac{1}{\lambda^3} \\ &= \frac{3}{2} k_B T \frac{V}{\lambda^3} f_{5/2}(z) \\ &= \frac{3}{2} pV \end{aligned}$$

Remarkably, the relation

$$E = \frac{3}{2} pV \quad (12.13)$$

is the same as in the classical limit. However, pV is not simply proportional to T ! Using thermodynamic relations a number of other properties can be derived from Eq.(12.13). See the following exercises.

An alternative derivation of the energy-pressure relation can be provided, which helps revealing its more general validity. Given an arbitrary system (interacting or not) governed by the Hamiltonian \hat{H} we have

$$\begin{aligned} p &= -\left. \frac{\partial \Phi}{\partial V} \right|_{T, \mu} = -\left. \frac{\partial \Phi}{\partial V} \right|_{T, z} \\ &= k_B T \left. \frac{\partial \ln Z_{gc}}{\partial V} \right|_{T, z} \\ &= k_B T \frac{1}{Z_{gc}} \text{Tr} \left\{ z^{\hat{N}} e^{-\beta \hat{H}} \left(-\beta \frac{\partial \hat{H}}{\partial V} \right) \right\} \\ &= -\left\langle \frac{\partial \hat{H}}{\partial V} \right\rangle \end{aligned} \quad (12.14)$$

If the particles do not interact with each other or with an external field, we have

$$\hat{H} = \sum_{\vec{p}} \hat{n}_{\vec{p}} \varepsilon_{\vec{p}}$$

with $\varepsilon_{\vec{p}} \propto V^{-2/3}$. Thus,

$$\frac{\partial \varepsilon_{\vec{p}}}{\partial V} = -\frac{2}{3} \frac{\varepsilon_{\vec{p}}}{V}$$

and

$$\frac{\partial \hat{H}}{\partial V} = \sum_{\vec{p}} \hat{n}_{\vec{p}} \frac{\partial \varepsilon_{\vec{p}}}{\partial V} = -\frac{2}{3} \frac{\hat{H}}{V}.$$

Consequently, replacing in Eq. (12.14) we obtain

$$p = \frac{2}{3} \frac{E}{V}.$$

In this way one concludes that the relation holds for fermions, bosons and classical particles alike.

Exercise 12.20: The energy of the quantum Fermi gas is given by $E = 3/2pV$. Show that this implies that the adiabatic compressibility is also given by the classical expression $\kappa_S = \frac{3}{5p}$. Are there other classical-ideal-gas relations which survive in the quantum case?

Exercise 12.21: Calculate the entropy S of the non-interacting Fermi gas. Express the result in terms of $\langle n_{\vec{p}} \rangle$ and interpret the result physically. Hint: For the interpretation calculate the number of accessible states Ω for N_i Fermions occupying g_i states, and express Ω in terms of the probability $n_i = N_i/g_i$ for an individual state to be occupied.

Summarizing so far, the functions

$$f_{5/2}(z) = \frac{4}{\sqrt{\pi}} \int_0^{\infty} dx x^2 \ln(1 + ze^{-x^2})$$

and

$$f_{3/2}(z) = z f'_{5/2}(z) = \frac{4}{\sqrt{\pi}} \int_0^{\infty} dx \frac{x^2}{z^{-1}e^{x^2} + 1},$$

define, on the one hand, the grand canonical potential Φ per particle and the equation of state in terms of λ^3/v and the fugacity z as

$$\begin{aligned} \Phi &= -k_B T N \frac{v}{\lambda^3} f_{5/2}(z), \\ \frac{pv}{k_B T} &= \frac{v}{\lambda^3} f_{3/2}(z). \end{aligned}$$

On the other hand, they allow us to obtain z as a function of λ^3/v by solving

$$\frac{\lambda^3}{v} = f_{3/2}(z).$$

These Riemann functions can be expressed in the form of integrals, power series or asymptotic expansions. Otherwise, they can only be evaluated numerically. In particular, there are no closed expressions, neither for N as a function of z , nor for the much needed $z = z(N)$ in the general case. There are, however, two most important limits, which deserve to be worked out analytically.

12.2 High temperatures and low densities

The limit of low densities and/or high temperatures corresponds to $\frac{\lambda^3}{v} \rightarrow 0$. Since $f_{3/2}(z)$ is a monotonously increasing function of z , this corresponds to $z \rightarrow 0$. This limit is easy to discuss since

$$f_{3/2}(z) = \sum_{l=1}^{\infty} (-1)^{l-1} \frac{z^l}{l^{3/2}} = z - \frac{1}{2\sqrt{2}} z^2 + O(z^3)$$

and

$$f_{5/2}(z) = \sum_{l=1}^{\infty} (-1)^{l-1} \frac{z^l}{l^{5/2}} = z - \frac{1}{4\sqrt{2}} z^2 + O(z^3)$$

have a convergent Taylor expansion at $z = 0$. We obtain z from

$$\frac{\lambda^3}{v} = z - \frac{z^2}{2\sqrt{2}} + O(z^3),$$

which implies

$$z = \frac{\lambda^3}{v} + \frac{1}{2\sqrt{2}} z^2 + O(z^3)$$

and

$$\begin{aligned} z &= \frac{\lambda^3}{v} + \frac{1}{2\sqrt{2}} \left(\frac{\lambda^3}{v}\right)^2 + O\left(\frac{\lambda^3}{v}\right)^3 \\ &= \frac{\lambda^3}{v} \left[1 + \frac{1}{2\sqrt{2}} \left(\frac{\lambda^3}{v}\right) + O\left(\frac{\lambda^3}{v}\right)^2 \right]. \end{aligned} \quad (12.15)$$

Taylor developing $f_{5/2}(z)$ for small z and replacing z with the expansion (12.15) we obtain the equation of state

$$\begin{aligned} \frac{pv}{k_B T} &= \left(\frac{v}{\lambda^3}\right) \left[z - \frac{1}{4\sqrt{2}} z^2 + O(z^3) \right] \\ &= 1 + \frac{1}{2\sqrt{2}} \frac{\lambda^3}{v} - \frac{1}{4\sqrt{2}} \frac{\lambda^3}{v} + O\left(\frac{\lambda^3}{v}\right)^2 \\ &= 1 + \frac{1}{4\sqrt{2}} \frac{\lambda^3}{v} + O\left(\frac{\lambda^3}{v}\right)^2. \end{aligned}$$

We recover the classical limit and the first non-vanishing quantum correction, which is entirely due to the symmetry correlations in the fermion wave functions (no interactions). Notice that the leading correction is proportional to $\lambda^3 \sim \hbar^3/T^{3/2}$. In the Fermi gas the pressure is always higher than in the classical Boltzmann gas. This is the consequence of the Pauli exclusion principle, which precludes identical fermions from occupying the same region in space, thus resulting in an effective repulsion among the particles. This enhances p , despite the lack of any interaction. As we shall see, the pressure of the Fermi gas is finite even at $T = 0$.

The grand canonical potential in the limit of $\left(\frac{\lambda^3}{v}\right) \rightarrow 0$ reads

$$\Phi = -k_B T N \left[1 + \frac{1}{4\sqrt{2}} \frac{\lambda^3}{v} + O\left(\frac{\lambda^3}{v}\right)^2 \right]$$

and

$$E = \frac{3}{2} k_B T N \left[1 + \frac{1}{4\sqrt{2}} \frac{\lambda^3}{v} + O\left(\frac{\lambda^3}{v}\right)^2 \right]$$

from which the leading quantum corrections to all other thermodynamic properties can be obtained (C_V , C_p , κ_T , α , S , etc.).

The average occupation $\langle \hat{n}_{\vec{p}} \rangle$ for $z \rightarrow 0$ (i.e., $\beta\mu \rightarrow -\infty$) given by

$$\langle \hat{n}_{\vec{p}} \rangle = \frac{1}{z^{-1} e^{\beta\varepsilon_{\vec{p}}} + 1} \simeq z e^{-\beta\varepsilon_{\vec{p}}} \simeq \frac{\lambda^3}{v} e^{-\beta\varepsilon_{\vec{p}}} + O\left[\left(\frac{\lambda^3}{v}\right)^2\right],$$

takes the Maxwell-Boltzmann form.

Exercise 12.22: Generalize the expressions for Φ and N for Fermi and Bose gases to the case where the single-particle energy levels $\varepsilon_{\vec{p}\alpha}$ have a degeneracy g (i.e., $\varepsilon_{\vec{p},\alpha} = \varepsilon_{\vec{p}}$ for $\alpha = 1, \dots, g$). In this way one can take into account the intrinsic angular momentum S of the particles ($g = 2S + 1$). Derive the corresponding expressions for $\langle \hat{n}_{\vec{p},\alpha} \rangle$.

12.3 Low temperatures and high densities

A far more interesting and important limit is found at low temperatures and/or high densities, i.e., $\frac{\lambda^3}{v} \rightarrow +\infty$. Since $f_{3/2}(z)$ is monotonously increasing we also have $z = e^{\beta\mu} \rightarrow +\infty$. We therefore need a systematic asymptotic approximation of the diverging functions $f_{5/2}(z)$ and $f_{3/2}(z)$ in the limit of $z \rightarrow +\infty$. This development is known as the

Sommerfeld expansion after Arnold Sommerfeld (1868–1951). It is given by

$$f_{3/2}(z) = \frac{4}{3\sqrt{\pi}}(\ln z)^{3/2} \left(1 + \frac{\pi^2}{8}(\ln z)^{-2} + O[(\ln z)^{-4}] \right)$$

and

$$f_{5/2}(z) = \frac{8}{15\sqrt{\pi}}(\ln z)^{5/2} \left(1 + \frac{5\pi^2}{8}(\ln z)^{-2} + O[(\ln z)^{-4}] \right).$$

In the following a formal derivation of the Sommerfeld expansion is presented. The physical consequences are discussed in Sec. 12.3.2.

12.3.1 The Sommerfeld expansion

We start from

$$f_{5/2}(z) = \frac{4}{\sqrt{\pi}} \int_0^{\infty} x^2 \ln(1 + ze^{-x^2}) dx$$

and replace $y = x^2$, $dy = 2xdx$, $dx = dy/2\sqrt{y}$ to obtain

$$\begin{aligned} f_{5/2}(z) &= \frac{4}{\sqrt{\pi}} \int_0^{\infty} \frac{1}{2\sqrt{y}} y \ln(1 + ze^{-y}) dy \\ &= \frac{2}{\sqrt{\pi}} \int_0^{\infty} \underbrace{\sqrt{y}}_{u'} \underbrace{\ln(1 + ze^{-y})}_v dy. \end{aligned}$$

We integrate by parts as

$$\begin{aligned} f_{5/2}(z) &= \frac{2}{\sqrt{\pi}} \left\{ \underbrace{\frac{2}{3}y^{3/2} \ln(1 + ze^{-y})}_{\Big|_0^{\infty}} + \int_0^{\infty} \frac{2}{3}y^{3/2} \frac{ze^{-y}}{1 + ze^{-y}} dy \right\} \\ f_{5/2}(z) &= \frac{4}{3\sqrt{\pi}} \int_0^{\infty} \frac{y^{3/2}}{z^{-1}e^y + 1} dy. \end{aligned}$$

Before further manipulations, let us apply the same change of variables $x^2 = y$ to

$$f_{3/2}(z) = \frac{4}{\sqrt{\pi}} \int_0^{\infty} \frac{x^2}{z^{-1}e^{x^2} + 1} dx$$

and obtain

$$f_{3/2}(z) = \frac{2}{\sqrt{\pi}} \int_0^{\infty} \frac{y^{1/2}}{z^{-1}e^y + 1} dy.$$

We see that both functions of interest may be cast in the same form. We thus need to expand

$$\tilde{f}_m(z) = \int_0^\infty \frac{\overbrace{y^{m-1}}^{u'}}{\underbrace{z^{-1}e^y + 1}_v} dy.$$

for $m = 3/2$ and $5/2$. We integrate by parts and obtain

$$\underbrace{\frac{1}{m} y^m \frac{1}{z^{-1}e^y + 1}}_0 \Big|_0^\infty + \int_0^\infty \frac{1}{m} y^m \frac{z^{-1}e^y}{(z^{-1}e^y + 1)^2} dy.$$

At this point it is useful to introduce the variable

$$\nu = \beta\mu \quad \Rightarrow \quad z = e^\nu \quad \text{and} \quad \nu = \ln z$$

and write

$$\tilde{f}_m(z) = \frac{1}{m} \int_0^\infty y^m \frac{e^{y-\nu}}{(e^{y-\nu} + 1)^2} dy.$$

Replacing $t = y - \nu$ we have

$$\tilde{f}_m(z) = \frac{1}{m} \int_{-\nu}^\infty (\nu + t)^m \frac{e^t}{(e^t + 1)^2} dt.$$

As $z \rightarrow \infty$, $\nu \rightarrow \infty$ and the integral approaches $\int_{-\infty}^\infty$. The function

$$h(t) = \frac{e^t}{(e^t + 1)^2} = \frac{e^{-2t}}{e^{-2t}} \frac{e^t}{(e^t + 1)^2} = \frac{e^{-t}}{(1 + e^{-t})^2} = h(-t)$$

is even, has a bell-like form with a maximum at $t = 0$ and a width (mean square deviation) $\sigma = \pi/\sqrt{3}$. It decreases exponentially for $|t| \rightarrow \infty$.

$$\int_{-\infty}^\infty h(t) dt = 1.$$

$$h(t) = -\frac{d}{dt} \left(\frac{1}{e^t + 1} \right), \quad t = \beta(\varepsilon - \mu)$$

Since $h(t)$ has a peak form with a finite width of about 2–4 around $t = 0$, we can expand

$$(\nu + t)^m = \nu^m + m\nu^{m-1}t + \frac{m(m-1)}{2}\nu^{m-2}t^2 + \dots$$

for small t and obtain

$$\tilde{f}_m(z) = \frac{1}{m} \int_{-\nu}^{\infty} dt \frac{e^t}{(e^t + 1)^2} \left(\nu^m + m\nu^{m-1}t + \frac{m(m-1)}{2}\nu^{m-2}t^2 + \dots \right)$$

We may now replace $\int_{-\nu}^{\infty} \dots dt = \int_{-\infty}^{\infty} \dots dt - \int_{-\infty}^{-\nu} \dots dt$ knowing that

$$\int_{-\infty}^{-\nu} t^l \frac{e^t}{(e^t + 1)^2} dt \approx \int_{-\infty}^{-\nu} t^l e^{-t} dt = t^l e^t \Big|_{-\infty}^{-\nu} - l \int_{-\infty}^{-\nu} t^{l-1} e^t dt \propto \nu^l e^{-\nu} \rightarrow 0$$

which vanishes as $\nu = \ln z \rightarrow +\infty$. For further details one may note that

$$\begin{aligned} A_l &= (-\nu)^l e^{-\nu} - l A_{l-1} \\ A_l = P_l e^{-\nu} &\Rightarrow P_l e^{-\nu} = -(\nu)^l e^{-\nu} - l P_{l-1} e^{-\nu} \\ P_l &= (-\nu)^l - l P_{l-1}, \end{aligned}$$

where P_l is a polynomial of degree l . The integrals $\int_{-\infty}^{-\nu} t^l \frac{e^t}{(e^t+1)^2} dt$ are of the order $\nu^l e^{-\nu}$ and can therefore be neglected in the limit $\nu = \ln z \rightarrow +\infty$. We finally have

$$\tilde{f}_m(z) = \frac{1}{m} \left(I_0 \nu^m + m I_1 \nu^{m-1} + \frac{m(m-1)}{2} I_2 \nu^{m-2} + \dots \right)$$

where

$$I_l = \int_{-\infty}^{+\infty} t^l \frac{e^t}{(e^t + 1)^2} dt$$

Since $h(t) = \frac{e^t}{(e^t+1)^2}$ is even all odd- l integrals vanish. For $l = 0$ we have

$$I_0 = \int_{-\infty}^{+\infty} \frac{e^t}{(e^t + 1)^2} dt = 1$$

and for even $l > 0$

$$\begin{aligned} I_l &= -2 \left[\frac{\partial}{\partial \lambda} \int_0^{\infty} dt \frac{t^{l-1}}{e^{\lambda t} + 1} \right]_{\lambda=1} \\ &= (l-1)! (2l) (1 - 2^{l-1}) \zeta(l) \end{aligned}$$

where $\zeta(l)$ is the Riemann zeta function:

$$\zeta(2) = \frac{\pi^2}{6}, \quad \zeta(4) = \frac{\pi^4}{90}, \dots$$

We can now replace this information explicitly for $m = 3/2$ and $5/2$:

$$\tilde{f}_{3/2}(z) = \frac{2}{3} \left[\nu^{3/2} + \frac{\pi^2}{3} \frac{3}{8} \nu^{-1/2} + O(\nu^{-5/2}) \right]$$

and therefore

$$\begin{aligned} f_{3/2}(z) &= \frac{2}{\sqrt{\pi}} \tilde{f}_{3/2}(z) = \frac{4}{3\sqrt{\pi}} \left(\nu^{3/2} + \frac{\pi^2}{8} \nu^{-1/2} + \dots \right) = \frac{4}{3\sqrt{\pi}} \nu^{3/2} \left(1 + \frac{\pi^2}{8} \nu^{-2} + \dots \right) \\ &= \frac{4}{3\sqrt{\pi}} \left((\ln z)^{3/2} + \frac{\pi^2}{8} (\ln z)^{-1/2} + O[(\ln z)^{-5/2}] \right). \end{aligned}$$

In the same way we have

$$f_{5/2}(z) = \frac{4}{3\sqrt{\pi}} \tilde{f}_{5/2}(z) \quad (12.16)$$

where

$$\tilde{f}_{5/2}(z) = \frac{2}{5} \left(\nu^{5/2} + \frac{\pi^2}{3} \frac{15}{8} \nu^{1/2} + O(\nu^{-3/2}) \right).$$

Consequently,

$$\begin{aligned} f_{5/2}(z) &= \frac{8}{15\sqrt{\pi}} \left(\nu^{5/2} + \frac{5\pi^2}{8} \nu^{1/2} + \dots \right) \\ &= \frac{8}{15\sqrt{\pi}} \left\{ (\ln z)^{5/2} + \frac{5\pi^2}{8} (\ln z)^{1/2} + O[(\ln z)^{-3/2}] \right\}. \end{aligned}$$

It is probably clearer to write

$$f_{3/2}(z) = \frac{4}{3\sqrt{\pi}} (\ln z)^{3/2} \left(1 + \frac{\pi^2}{8} (\ln z)^{-2} + O[(\ln z)^{-4}] \right) \quad (12.17)$$

and

$$f_{5/2}(z) = \frac{8}{15\sqrt{\pi}} (\ln z)^{5/2} \left(1 + \frac{5\pi^2}{8} (\ln z)^{-2} + O[(\ln z)^{-4}] \right). \quad (12.18)$$

The above Sommerfeld expansion will allow us to calculate the properties of the ideal fermion gas in the most interesting quantum limit. The expressions become a lot more transparent once one replaces $\ln z$ by $\beta\mu$ on the right hand side. In the following section we use the expansion (12.17) of $f_{3/2}(z)$ in order to obtain $z = f_{3/2}^{-1}(\lambda^3/v)$ as a function of λ^3/v . Replacing then z in the the expansion (12.18) of $f_{5/2}(z)$ gives access to the equation of state and to all thermodynamic properties as a function of T , N and V .

12.3.2 Physics of the low temperature limit

Before using the Sommerfeld expansion and solving for z as a function of the density N/V or the volume per particle $v = \frac{V}{N}$ it is useful to consider the $T = 0$ limit explicitly.

In this case the occupation numbers are

$$\langle n_{\vec{p}} \rangle = \frac{1}{e^{\beta(\varepsilon_{\vec{p}} - \mu)} + 1} = \begin{cases} 1 & \text{if } \varepsilon_{\vec{p}} < \mu = \varepsilon_F \\ 0 & \text{if } \varepsilon_{\vec{p}} > \mu = \varepsilon_F \end{cases}$$

All states are occupied below the Fermi energy ε_F and all states above ε_F are empty, where

$$\varepsilon_F = \frac{p_F^2}{2m}$$

and

$$N = \frac{V}{(2\pi\hbar)^3} 4\pi \int_0^{p_F} p^2 dp = \frac{V}{(2\pi\hbar)^3} \frac{4}{3} \pi p_F^3 = \frac{V}{6\pi^2\hbar^3} p_F^3.$$

The Fermi wave vector is thus given by the density or volume per particle v as

$$p_F = \hbar \left(\frac{6\pi^2}{v} \right)^{1/3} = \hbar k_F.$$

It is an eminently quantum property proportional to \hbar divided by the average distance $v^{1/3}$ between the Fermions. Accordingly, the Fermi energy is given by

$$\varepsilon_F = \frac{\hbar^2}{2m} \left(\frac{6\pi^2}{v} \right)^{2/3}. \quad (12.19)$$

Both ε_F and p_F are independent of the size of the system. They are intensive quantum properties, which are defined entirely by the particle density $N/V = 1/v$ or volume per particle v , and which increase monotonously as a function of $1/v$. In order to quantify the temperature scale below which quantum effects dominate we introduce the Fermi temperature

$$T_F = \frac{\varepsilon_F}{k_B}.$$

In fact, the Sommerfeld expansion and the following discussion hold only as long as T is clearly smaller than T_F . Quantitative values of T_F in usual metals (of the order of 10^4 K) show that this is perfectly valid assumption.

The importance of ε_F and T_F becomes clearer when one considers the properties at finite low temperatures. To calculate μ to the lowest non-vanishing order in T we use the expansion of $f_{3/2}(z)$ for large z :

$$\frac{\lambda^3}{v} = f_{3/2}(z) = \frac{4}{3\sqrt{\pi}} (\ln z)^{3/2} \left(1 + \frac{\pi^2}{8} (\ln z)^{-2} + O[(\ln z)^{-4}] \right).$$

Recalling that

$$\lambda = \frac{2\pi\hbar}{\sqrt{2\pi m k_B T}} = \sqrt{\frac{2\pi\hbar^2}{m}} \beta^{1/2} \quad \rightarrow \quad \lambda^3 = \left(\frac{2\pi\hbar^2}{m} \right)^{3/2} \beta^{3/2}$$

and replacing $\ln z = \beta\mu$ we have

$$\frac{1}{v} \left(\sqrt{\frac{2\pi\hbar^2}{m}} \right)^3 \beta^{3/2} = \frac{4\mu^{3/2}}{3\sqrt{\pi}} \beta^{3/2} \left(1 + \frac{\pi^2}{8}(\beta\mu)^{-2} + \dots \right).$$

Rearranging the prefactor we obtain

$$\varepsilon_F^{3/2} = \frac{1}{v} 6\pi^2 \left(\frac{\hbar^2}{2m} \right)^{3/2} = \mu^{3/2} \left(1 + \frac{\pi^2}{8}(\beta\mu)^{-2} + O[(\beta\mu)^{-4}] \right)$$

which yields

$$\begin{aligned} \varepsilon_F &= \frac{\hbar^2}{2m} \left(\frac{6\pi^2}{v} \right)^{2/3} = \mu \left(1 + \frac{\pi^2}{8}(\beta\mu)^{-2} + \dots \right)^{2/3} \\ &= \mu \left(1 + \frac{\pi^2}{12}(\beta\mu)^{-2} + \dots \right), \end{aligned}$$

or equivalently,

$$\mu = \varepsilon_F - \frac{\pi^2}{12} \left(\frac{k_B T}{\varepsilon_F} \right)^2 \left(\frac{\varepsilon_F}{\mu} \right) + O \left(\frac{k_B T}{\mu} \right)^4.$$

Replacing μ on the right hand side and keeping the terms up to order $(k_B T/\varepsilon_F)^2$ we have

$$\begin{aligned} \frac{\mu}{\varepsilon_F} &= 1 - \frac{\pi^2}{12} \left(\frac{k_B T}{\varepsilon_F} \right)^2 \left(\frac{\varepsilon_F}{\mu} \right) + O \left(\frac{k_B T}{\mu} \right)^4 \\ \frac{\mu}{\varepsilon_F} &= 1 - \frac{\pi^2}{12} \left(\frac{k_B T}{\varepsilon_F} \right)^2 + O \left(\frac{k_B T}{\varepsilon_F} \right)^4 \\ \frac{\mu}{\varepsilon_F} &= 1 - \frac{\pi^2}{12} \left(\frac{T}{T_F} \right)^2 + O \left(\frac{T}{T_F} \right)^4 \end{aligned}$$

Notice that $\mu \approx \varepsilon_F$ up to temperatures of the order of T_F . To give an order of magnitude, in narrow band metals we find $\varepsilon_F \approx 5$ eV. This corresponds to $T_F \approx 5 \times 10^4$ K. Larger values are found in simple metals, which are all much higher than the melting temperature, for example. Consequently, ignoring the temperature dependence of μ in solid metals is usually a very good approximation.

The fact that μ decreases with increasing T is a consequence of the form of the single-particle energy spectrum of free particles in three dimensions. We may compute the single-particle density of states (DOS)

$$\rho(\varepsilon) = \sum_{\vec{p}} \delta(\varepsilon_{\vec{p}} - \varepsilon) = \frac{V}{(2\pi\hbar)^3} \int d^3 p \delta(\varepsilon_{\vec{p}} - \varepsilon),$$

which represents the number of single-particle states per unit energy at the energy ε . For non-interacting particles of mass m we have $\varepsilon_{\vec{p}} = p^2/2m$ and

$$\rho(\varepsilon) = \frac{V}{(2\pi\hbar)^3} 4\pi \int_0^\infty p^2 \delta\left(\frac{p^2}{2m} - \varepsilon\right) dp.$$

Introducing

$$\varepsilon' = \frac{p^2}{2m} \quad \Rightarrow \quad d\varepsilon' = \frac{p}{m} dp$$

and

$$dp = \frac{m}{\sqrt{2m\varepsilon'}} d\varepsilon'$$

we obtain

$$\begin{aligned} \rho(\varepsilon) &= \frac{V}{(2\pi\hbar)^3} 4\pi \sqrt{\frac{m}{2}} 2m \int_0^\infty \sqrt{\varepsilon'} \delta(\varepsilon' - \varepsilon) d\varepsilon' \\ &= \frac{Nv(2m)^{3/2}}{4\pi\hbar^3} \sqrt{\varepsilon} = \frac{3}{2} \frac{N}{\varepsilon_F^{3/2}} \sqrt{\varepsilon}. \end{aligned}$$

The number of particles can then be calculated from

$$N = \int_{-\infty}^{+\infty} \rho(\varepsilon) f(\varepsilon) d\varepsilon = \int_{-\infty}^{+\infty} \rho(\varepsilon) \frac{1}{e^{\beta(\varepsilon-\mu)} + 1} d\varepsilon$$

where we have replaced the average occupation of the states with energy ε by the Fermi function

$$f(\varepsilon) = \frac{1}{e^{\beta(\varepsilon-\mu)} + 1}.$$

In particular at $T = 0$ we have

$$N = \int_{-\infty}^{\varepsilon_F} \rho(\varepsilon) d\varepsilon$$

and in general for $T \geq 0$

$$N = \int_{-\infty}^{+\infty} \rho(\varepsilon) f(\varepsilon) d\varepsilon.$$

The equation of state is obtained from

$$\begin{aligned}
\frac{pv}{k_B T} &= \frac{v}{\lambda^3} f_{5/2}(z) = \frac{f_{5/2}(z)}{f_{3/2}(z)} \\
&= \frac{2}{5} \ln z \left(\frac{1 + \frac{5\pi^2}{8} (\ln z)^{-2} + \dots}{1 + \frac{\pi^2}{8} (\ln z)^{-2} + \dots} \right) \\
&= \frac{2}{5} \ln z \left(1 + \frac{5\pi^2}{8} (\ln z)^{-2} - \frac{\pi^2}{8} (\ln z)^{-2} + \dots \right) \\
&= \frac{2}{5} \ln z \left(1 + \frac{\pi^2}{2} (\ln z)^{-2} + O(\ln z)^{-4} \right) \\
&= \frac{2}{5} \beta \mu \left(1 + \frac{\pi^2}{2} (\beta \mu)^{-2} + \dots \right) \\
&= \frac{2}{5} \frac{\varepsilon_F}{k_B T} \left[1 - \frac{\pi^2}{12} \left(\frac{k_B T}{\varepsilon_F} \right)^2 + \frac{\pi^2}{2} \left(\frac{k_B T}{\varepsilon_F} \right)^2 + \dots \right] \\
&= \frac{2}{5} \frac{\varepsilon_F}{k_B T} \left[1 + \frac{5\pi^2}{12} \left(\frac{k_B T}{\varepsilon_F} \right)^2 + O\left(\frac{k_B T}{\varepsilon_F} \right)^4 \right].
\end{aligned}$$

This can be written in a more compact form as

$$pv = \frac{2}{5} \varepsilon_F \left[1 + \frac{5\pi^2}{12} \left(\frac{k_B T}{\varepsilon_F} \right)^2 + \dots \right].$$

Notice that the pressure does not vanish at $T = 0$. This is of course a consequence of Pauli's exclusion principle, which forces a macroscopic number of fermions to occupy non-vanishing momentum states —up to $p_F = \hbar(6\pi^2/v)^{1/3}$ — and the Heisenberg principle, which enhances the kinetic energy as the volume is reduced.

For non-interacting Fermi gases the pressure gives us directly the internal energy through

$$E = \frac{3}{2} pV = \frac{3}{5} N \varepsilon_F \left(1 + \frac{5\pi^2}{12} \left(\frac{k_B T}{\varepsilon_F} \right)^2 + \dots \right),$$

Notice the ground state energy $E_0 = \frac{3}{5} N \varepsilon_F$ is simply proportional to ε_F . This result could also have been obtained from

$$E_0 = \int_0^{\varepsilon_F} \rho(\varepsilon) \varepsilon d\varepsilon = \frac{3}{2} \frac{N}{\varepsilon_F^{3/2}} \int_0^{\varepsilon_F} \varepsilon^{3/2} d\varepsilon = \frac{3}{2} \frac{N}{\varepsilon_F^{3/2}} \frac{2}{5} \varepsilon_F^{5/2} = \frac{3}{5} N \varepsilon_F.$$

From the temperature dependence of E we straightforwardly obtain the physically important heat capacity at constant volume

$$C_V = \left. \frac{\partial E}{\partial T} \right|_{V,N} = \frac{3}{5} N \varepsilon_F \frac{5\pi^2}{6} \frac{k_B^2 T}{\varepsilon_F^2} = \overbrace{\frac{\pi^2}{2} \frac{k_B^2}{\varepsilon_F}}^{\gamma} N T + O\left(\frac{k_B T}{\varepsilon_F} \right)^3,$$

which is *linear in T* and thus satisfies the third principle of thermodynamics. Knowing that $C_V \rightarrow \frac{3}{2}Nk_B$ for $T \rightarrow +\infty$, we can infer qualitatively the full temperature dependence.

A few important properties of the Fermi energy

$$\varepsilon_F = \frac{\hbar^2}{2m} \left(\frac{6\pi^2}{v} \right)^{2/3}$$

deserve to be underlined:

- i) ε_F (or the Fermi temperature $T_F = \varepsilon_F/k_B T$) defines the temperature scale below which quantum effects and in particular wave-function-symmetry correlations dominate the behavior of the ideal Fermi gas.
- ii) ε_F is a quantum property proportional to \hbar^2 , since $\varepsilon_F = p_F^2/2m$ and $p_F \sim \hbar$.
- iii) ε_F increases when the mass of the particles decreases.
- iv) ε_F is an intensive property. It does not depend on the size of the system but on the volume per particle $v = V/N$, which defines the average interparticle distance $v^{1/3}$. Equivalently, we may express ε_F in terms of the density $n = \frac{N}{V}$.
- v) As the density increases, the Fermi energy ε_F and the Fermi momentum p_F increase. A larger density (e.g., due to compression) increases the size of the Fermi sphere. This behavior is understandable taking into account the symmetry correlations, which preclude two Fermions from occupying the same state, and the Heisenberg principle, which implies an increase of the kinetic energy of each single-particle state with decreasing volume. Notice that neither ε_F nor p_F change when both the system volume V and the particle number N are scaled keeping $v = V/N$ constant. If V is increased only the density of the allowed values of the momentum $p_\alpha = 2\pi\nu_\alpha/L$ increases.

After all these calculations it is interesting to qualitatively discuss how the properties change as a function of T within the quantum (also known as degenerate) regime ($T \ll T_F$). At $T = 0$ the occupation-number distribution $n_{\vec{p}}$ looks like an ice cube or ice block (Heaviside step function). As the temperature increases, at $T > 0$, the ice cube melts. The step in the Fermi function gives place to a smooth transition, from $f(\varepsilon) \simeq 1$ for ε well below the chemical potential [$(\varepsilon - \mu) \ll -k_B T$], to $f(\varepsilon) \simeq 0$ for ε well above μ [$(\varepsilon - \mu) \gg k_B T$]. At the same time as $f(\varepsilon)$ describes this smooth crossover, the chemical potential decreases slightly. This is a consequence of the shape of the density of single-particle states (DOS) $\rho(\varepsilon) \propto \sqrt{\varepsilon}$, which increases with energy. In fact, if one would increase T keeping μ constant, the number of particles N would increase, since the number of states with energy $\varepsilon > \varepsilon_F$ is larger than the number of states with $\varepsilon < \varepsilon_F$. Notice, that $f(\varepsilon)$ occupies as many states above μ as it produces holes below μ :

$$1 - \frac{1}{e^x + 1} = \frac{e^x + 1 - 1}{e^x + 1} = \frac{1}{e^{-x} + 1}.$$

A decrease in occupation below μ ($x < 0$) implies and increase in occupation above μ

($-x > 0$). Thus, $f(\varepsilon)$ alone does not change the number of particles as T increases. In other words, if $\rho(\varepsilon)$ would be independent of ε , N and thus μ would not change with increasing T :

$$\int_{-\infty}^0 \left(1 - \frac{1}{1 + e^x}\right) dx = \int_{-\infty}^0 \frac{dx}{e^{-x} + 1} = - \int_{+\infty}^0 \frac{dx}{e^x + 1} = \int_0^{+\infty} \frac{dx}{e^x + 1}.$$

Thus, the number of unoccupied states below μ equals the number of occupied states above μ if $\rho(\varepsilon)$ is independent of ε around ε_F .

The energy range where the Fermi function $f(\varepsilon)$ is significantly different from 0 or 1 (i.e., the crossover region) is of the order of $k_B T$. The physical reason for this is simply that the energy available for excitations is of the order of $k_B T$. With increasing T , the entropy increases and electron-hole excitations become increasingly probable. The number of these excitations is of the order of $k_B T \rho(\varepsilon_F)$, where $\rho(\varepsilon_F)$ is the single-particle density of states at ε_F . The energy involved in each excitation is in average of the order of $k_B T$. Thus, the internal energy E increases proportional to $(k_B T)^2$, which explains qualitatively why the specific heat $C_V = (\partial E / \partial T)_V$ is linear in T , with a coefficient γ that is proportional to $\rho(\varepsilon_F)$. More precisely, the DOS at ε_F is given by

$$\rho(\varepsilon_F) = \frac{3}{2} \frac{N}{\varepsilon_F^{3/2}} \sqrt{\varepsilon_F} = \frac{3}{2} \frac{N}{\varepsilon_F} \quad (12.20)$$

and the heat capacity is given by

$$C_V = \frac{\pi^2}{2} k_B^2 \frac{N}{\varepsilon_F} T = \frac{\pi^2 k_B^2}{3} \underbrace{\rho(\varepsilon_F)}_{\gamma} T. \quad (12.21)$$

The reduction of C_V with decreasing T reflects the reduction of the number of degrees of freedom which are available to excite the fermions as the temperature is lowered. In other words, the volume in momentum space, within which excitations are possible, becomes smaller and smaller as we approach zero temperature. This is, again, a direct consequence of the correlations imposed by the antisymmetry of fermionic wave functions. Indeed, the exclusion principle precludes all electrons except those occupying states close the Fermi energy —within an energy range $k_B T$ — to be excited, since the thermally available energy for these excitations is only $k_B T$. It is this reduction of accessible volume in momentum space what reduces C_V , in a way that it complies with the 3rd principle as $T \rightarrow 0$.

13 Bose systems: Photons and black body radiation

The most important application of Bose-Einstein statistics is the theory of electromagnetic radiation in thermal equilibrium. One considers the electromagnetic radiation enclosed on a volume V at temperature T . This cavity is known as black-body cavity, since it would absorb all radiation incident on it. Such a body can be realized by a cavity with highly absorbing internal walls and a small aperture. Any ray entering through the aperture could only leave the cavity after numerous reflections on the internal walls of the cavity. When the aperture is sufficiently small all radiation incident on the aperture will be practically absorbed and thus the surface of the aperture behaves like a black body.

The cavity can be stabilized at any desired temperature T . The atoms in the cavity surface constantly emit and absorb photons, so that in the equilibrium state, a certain amount of electromagnetic radiation will be enclosed in the cavity. The linearity of the equations of electrodynamics implies that the photons do not interact with each other. We can of course ignore the entirely negligible interaction resulting from virtual creation of electron-positron pairs (scattering of light by light). However, notice that, precisely because of the lack of interaction between the photons, the presence of the cavity walls and the resulting interaction between photons and matter is crucial in order that the photon gas reaches equilibrium.

The mechanism by which equilibrium is reached consists in the absorption and emission of photons by the cavity walls. The photon gas in the cavity differs therefore from conventional gases of massive particles in an essential point: the number of particles (photons) N is not conserved. It is not a constant of motion as in ordinary gases. The variable number N of photons must be determined from the equilibrium conditions, namely, temperature and volume, by minimizing the free energy or grand canonical potential without any restriction. In fact, since N is not conserved, it does not appear in the expansion of the logarithm of the density matrix $\ln \hat{\rho}$, as a linear combination of the conserved additive constants of motion. Consequently, the statistical description of the photon gas follows the grand canonical ensemble with

$$\mu = 0.$$

The partition function is obtained by summing over all possible occupations of the photon states without any Lagrange multiplier or chemical potential μ . This is consistent with the condition that the average equilibrium value of N is obtained when the free energy F (or the grand canonical potential Φ) of the gas is minimal for the given T and V , which also corresponds to

$$\mu = \left. \frac{\partial F}{\partial N} \right|_{T,V} = 0.$$

13.1 Basic properties of the quantized electromagnetic field

The quantum electromagnetic field is characterized by the following properties:

- i) A stationary state of the electromagnetic field in the absence of sources (free EMF) is characterized uniquely by the photon occupation numbers $\{n_{\vec{k}\alpha}\}$, where \vec{k} refers to the wave vector of the photon and $\alpha = +, -$ denote the polarization directions, which are perpendicular to each other and to \vec{k} ($\hat{\varepsilon}_+ \cdot \hat{\varepsilon}_- = 0$, and $\hat{\varepsilon}_\alpha \cdot \vec{k} = 0$). The occupation numbers $n_{\vec{k}\alpha}$ can take any integer value $n_{\vec{k}\alpha} = 0, 1, 2, \dots$. Such a state of the EMF has the total energy

$$E = \sum_{\vec{k}, \alpha} n_{\vec{k}\alpha} \hbar \omega_k = \sum_{\vec{k}, \alpha} n_{\vec{k}\alpha} \varepsilon_k,$$

with

$$\omega_k = ck, \quad \varepsilon_k = \hbar \omega_k = cp$$

and a total momentum

$$\vec{P} = \sum_{\vec{k}, \alpha} n_{\vec{k}\alpha} \hbar \vec{k}.$$

The values of \vec{k} are those allowed by the boundary conditions in the volume $V = L^3$, which form a complete orthonormal plane-wave basis set. The periodic boundary conditions read

$$k_i L = 2\pi m_i$$

with $m_i \in \mathbb{Z}$ and $i = 1, 2, 3$, or equivalently,

$$\vec{k} = \frac{2\pi}{L} \vec{m} = \frac{2\pi}{V^{1/3}} \vec{m} \quad (13.1)$$

with $\vec{m} \in \mathbb{Z}^3$. Since $k = \sqrt{k_x^2 + k_y^2 + k_z^2} \propto V^{-1/3}$ we have $\omega_k \propto V^{-1/3}$. In the limit of large V we may replace the sums over the allowed values of \vec{k} by integrals as

$$\sum_{\vec{k}} \dots \rightarrow \frac{V}{(2\pi)^3} \int \dots d^3k.$$

- ii) The lowest energy state of the EMF is obtained by setting all the photon occupation numbers $n_{\vec{k}\alpha} = 0$. This is known as the vacuum state and is denoted by $|0\rangle$. All other stationary states $|\{n_{\vec{k}\alpha}\}\rangle$ can be obtained from $|0\rangle$ by applying the photon creation operators $\hat{a}_{\vec{k}\alpha}^\dagger$ as in the quantum harmonic oscillator:

$$|\{n_{\vec{k}\alpha}\}\rangle = \prod_{\vec{k}\alpha} \frac{1}{\sqrt{n_{\vec{k}\alpha}!}} \left(\hat{a}_{\vec{k}\alpha}^\dagger\right)^{n_{\vec{k}\alpha}} |0\rangle$$

- iii) The creation operators $\hat{a}_{\vec{k}\alpha}^\dagger$ and annihilation operators $\hat{a}_{\vec{k}\alpha} = (\hat{a}_{\vec{k}\alpha}^\dagger)^\dagger$ satisfy the boson commutation relations

$$[\hat{a}_{\vec{k}\alpha}, \hat{a}_{\vec{k}'\alpha'}] = [\hat{a}_{\vec{k}\alpha}^\dagger, \hat{a}_{\vec{k}'\alpha'}^\dagger] = 0$$

and

$$[\hat{a}_{\vec{k}\alpha}, \hat{a}_{\vec{k}'\alpha'}^\dagger] = \delta_{kk'} \delta_{\alpha\alpha'}$$

for all $\vec{k}, \vec{k}', \alpha$ and α' .

- iv) The state $|\{n_{\vec{k}\alpha}\}\rangle$ is fully symmetric with respect to interchange of all the operators $\hat{a}_{\vec{k}\alpha}^\dagger$ that define it. Consequently, photons are bosons.
- v) The state $\hat{a}_{\vec{k}\alpha}^\dagger |0\rangle$ is a single-photon state having an energy $\varepsilon_k = \hbar\omega_k = \hbar ck$ and a momentum $\hbar\vec{k}$. The state $|\{n_{\vec{k}\alpha}\}\rangle$ contains $n_{\vec{k}\alpha}$ photons in each state $\vec{k}\alpha$. The total number of photons is thus

$$N = \sum_{\vec{k}\alpha} n_{\vec{k}\alpha}.$$

- vi) The operator $\hat{a}_{\vec{k}\alpha}^\dagger$ ($\hat{a}_{\vec{k}\alpha}$) creates (annihilates) a photon in the state $\vec{k}\alpha$ according to the usual boson algebraic rules

$$\begin{aligned} \hat{a}_{\vec{k}\alpha}^\dagger |\{n_{\vec{k}\alpha}\}\rangle &= \sqrt{n_{\vec{k}\alpha} + 1} |n_{\vec{k}\alpha} + 1\rangle \\ \hat{a}_{\vec{k}\alpha} |\{n_{\vec{k}\alpha}\}\rangle &= \sqrt{n_{\vec{k}\alpha}} |n_{\vec{k}\alpha} - 1\rangle. \end{aligned}$$

13.2 The Planck distribution

The Bose-Einstein statistics applies to the quantized EMF, since its stationary states are uniquely defined by the occupation numbers $n_{\vec{k}\alpha}$ of photons states $\vec{k}\alpha$ and all integer values of $n_{\vec{k}\alpha} = 0, 1, 2, \dots$ are allowed. The partition function is then given by

$$Z_{gc} = \prod_{\vec{k}, \alpha} \left(\sum_{n=0}^{+\infty} e^{-\beta \hbar \omega_k n} \right) = \prod_{\vec{k}, \alpha} \frac{1}{1 - e^{-\beta \hbar \omega_k}} = \prod_{\vec{k}} \left(\frac{1}{1 - e^{-\beta \hbar \omega_k}} \right)^2.$$

Since $\mu = 0$, the grand canonical potential $\Phi = -\frac{1}{\beta} \ln Z_{gc}$ and the free energy $F = \Phi + \mu N$ coincide. In the following we rather refer to it as free energy and write

$$\Phi = F = -\frac{1}{\beta} \ln Z_{gc} = 2k_B T \sum_{\vec{k}} \ln(1 - e^{-\beta \hbar \omega_k}).$$

The average occupation number for photons with wave vector \vec{k} and polarization α is given by

$$\langle n_{\vec{k}\alpha} \rangle = \frac{1}{2} \frac{\partial F}{\partial (\hbar \omega_k)} = \frac{1}{e^{\beta \hbar \omega_k} - 1},$$

which is known as *Planck's distribution*:

$$\langle n_{\vec{k}\alpha} \rangle = \frac{1}{e^{\beta \hbar \omega_k} - 1}.$$

It corresponds to the Bose-Einstein distribution for $\mu = 0$. Summing the contributions of both polarizations we have

$$\langle n_{\vec{k}} \rangle = \frac{2}{(e^{\beta \hbar \omega_k} - 1)}.$$

13.3 The energy-pressure relation

The internal energy is given by

$$\begin{aligned} E &= -\frac{\partial}{\partial\beta} \ln Z_{gc} = \sum_{\vec{k}} \frac{2\hbar\omega_k e^{-\beta\hbar\omega_k}}{1 - e^{-\beta\hbar\omega_k}} \\ &= \sum_{\vec{k}} \hbar\omega_k \langle n_{\vec{k}} \rangle \end{aligned}$$

We can obtain the pressure (i.e., the equation of state) from

$$p = -\frac{\partial F}{\partial V} = \frac{2}{\beta} \sum_{\vec{k}} \frac{e^{-\beta\hbar\omega_k}}{1 - e^{-\beta\hbar\omega_k}} \left(-\beta\hbar \frac{\partial\omega_k}{\partial V} \right) = -\sum_{\vec{k}} \langle \hat{n}_{\vec{k}} \rangle \hbar \frac{\partial\omega_k}{\partial V}.$$

Knowing that $\omega_k = ck = c \frac{2\pi}{V^{1/3}} m$, where $m = \sqrt{m_x^2 + m_y^2 + m_z^2}$ [see Eq. (13.1)], we have

$$\frac{\partial\omega_k}{\partial V} = -\frac{1}{3} c \frac{2\pi}{V^{4/3}} m = -\frac{1}{3} \frac{\omega_k}{V},$$

which allows us to write

$$p = \sum_{\vec{k}} \frac{1}{3} \langle n_{\vec{k}} \rangle \frac{\hbar\omega_k}{V}.$$

Therefore,

$$pV = \frac{1}{3} E.$$

The same relation would have obtained for fermions having a dispersion relation $\varepsilon_{\vec{p}} = cp$, since this follows from the dependence of $\varepsilon_{\vec{p}}$ on V . The reader may wish to prove this statement as an exercise.

13.4 Thermodynamic properties

We may now take the limit $V \rightarrow \infty$ and obtain the number of linearly independent single-particle states or photon modes in a volume element d^3k around \vec{k} as

$$2 \frac{V}{(2\pi)^3} d^3k.$$

The number of photon modes having a magnitude k in the interval $(k, k + dk)$ is thus

$$\frac{8\pi V}{(2\pi)^3} k^2 dk.$$

Since $\omega = ck$, the number of modes with a frequency ω in the interval $(\omega, \omega + d\omega)$ is

$$\frac{8\pi V}{(2\pi)^3 c^3} \omega^2 d\omega.$$

Therefore, the density of photon modes with frequency ω is given by

$$\rho(\omega) = \frac{V}{\pi^2 c^3} \omega^2.$$

Note that $\rho(\omega)$ includes the contribution of both polarizations.

Consequently, the internal energy is given by

$$E = \int_0^{+\infty} \frac{\hbar\omega \rho(\omega) d\omega}{e^{\beta\hbar\omega} - 1} = V \frac{\hbar}{\pi^2 c^3} \int_0^{+\infty} \frac{\omega^3 d\omega}{e^{\beta\hbar\omega} - 1}.$$

The corresponding energy density per unit volume and unit frequency is

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

in terms of which we have

$$\frac{E}{V} = \int_0^{+\infty} u(\omega, T) d\omega.$$

This is *Planck's radiation law*, which gives the energy density $u(\omega, T)$ due to photons of frequency ω at temperature T .

Introducing the variable $x = \beta\hbar\omega$ we have

$$E = \frac{V(k_B T)^4}{\pi^2 c^3 \hbar^3} \underbrace{\int_0^{\infty} \frac{x^3}{e^x - 1} dx}_{\pi^4/15}$$

$$\frac{E}{V} = \frac{\pi^2 (k_B T)^4}{15 c^3 \hbar^3}$$

Introducing the *Stefan-Boltzmann constant*

$$\sigma = \frac{\pi^2 k_B^4}{60 \hbar^3 c^2} = 5,67 \times 10^{-5} \frac{\text{erg}}{\text{cm}^2 \text{ s K}^4},$$

we can write

$$\frac{E}{V} = \frac{4\sigma}{c} T^4.$$

The free energy can be obtained in the same way:

$$\begin{aligned}
 F &= k_B T \int_0^{\infty} \rho(\omega) \ln(1 - e^{-\beta\hbar\omega}) d\omega \\
 &= k_B T \frac{V}{\pi^2 c^3} \int_0^{\infty} \omega^2 \ln(1 - e^{-\beta\hbar\omega}) d\omega \\
 &= \frac{V(k_B T)^4}{\pi^2 c^3 \hbar^3} \underbrace{\int_0^{\infty} x^2 \ln(1 - e^{-x}) dx}_{-\pi^4/45} \\
 &= -\frac{\pi^2}{45} \frac{V(k_B T)^4}{c^3 \hbar^3} = -\frac{4\sigma}{3c} V T^4
 \end{aligned}$$

The other thermodynamic properties follow straightforwardly:

$$\begin{aligned}
 S &= -\frac{\partial F}{\partial T} = \frac{16\sigma}{3c} V T^3 \\
 E &= TS + F = \frac{4\sigma}{c} V T^4 \\
 C_V &= \frac{16\sigma}{c} V T^3.
 \end{aligned}$$

Note that $C_V \rightarrow \infty$ for $T \rightarrow \infty$, since the energy and number of photons has no upper bound. We shall see that this is not the case for lattice vibrations (phonons) which do reach a proper classical limit. Finally, the equation of state can be written as

$$p = -\frac{\partial F}{\partial V} = \frac{1}{3} \frac{E}{V} = \frac{4\sigma}{3c} T^4.$$

13.5 Revisiting Planck's law

It is useful to return to the average occupation number

$$\langle n_{\vec{k}\alpha} \rangle = \frac{1}{e^{\beta\hbar\omega_k} - 1}$$

for photon modes with wave vector \vec{k} and polarization α and to analyze how the energy of the electromagnetic radiation within the cavity is distributed as a function of the frequency ω . First of all notice that for $T = 0$ there are no photons with finite frequency, i.e., $\langle n_{\vec{k}\alpha} \rangle = 0, \forall \vec{k} \neq 0$. For any finite temperature we consider the spectral energy-density per unit volume

$$\begin{aligned} u(\omega, T) &= \frac{1}{V} \sum_{\vec{k}, \alpha} \hbar\omega_k \langle n_{\vec{k}\alpha} \rangle \delta(\omega - \omega_k) \\ &= 2 \frac{4\pi}{(2\pi)^3} \int_0^\infty \frac{\hbar\omega_k \delta(\omega - \omega_k)}{e^{\beta\hbar\omega_k} - 1} k^2 dk. \end{aligned}$$

Noting that $\omega_k = ck$ we have

$$\begin{aligned} u(\omega, T) &= \frac{1}{\pi^2} \frac{\hbar}{c^3} \int_0^\infty \frac{\omega'^3 \delta(\omega - \omega')}{e^{\beta\hbar\omega'} - 1} d\omega' \\ &= \frac{\hbar}{\pi^2 c^3} \frac{\omega^3}{e^{\beta\hbar\omega} - 1}, \end{aligned}$$

which is known as *Planck's law*. We have already seen that $\int d\omega u(\omega, T) = E/V = \frac{4\sigma}{c} T^4$.

For frequencies that are low relative to the temperature ($\hbar\omega \ll k_B T$) we obtain

$$u(\omega, T) \approx \frac{\hbar}{\pi^2 c^3} \frac{k_B T}{\hbar\omega} \omega^3 = \frac{\omega^2}{\pi^2 c^3} k_B T = \frac{\rho(\omega)}{V} k_B T$$

This is the *Rayleigh-Jeans formula*, which does not involve the Planck constant \hbar and corresponds to the classical limit (high temperatures). This formula has been derived from classical electromagnetism. If applied beyond its domain of validity, it implies an ultraviolet catastrophe of the form $u(\omega, T) \propto \omega^2$, in which the high frequencies get so much energy that the total energy of the EMF in the cavity would diverge. It is interesting to note that the Rayleigh-Jeans limit ($\hbar\omega \ll k_B T$) corresponds to the equipartition theorem because $u(\omega, T) = \frac{\rho(\omega)}{V} k_B T$, where $\rho(\omega)$ is the density of states at the frequency ω and $k_B T$ is the energy per photon mode.

In the opposite limit of high frequencies relative to T , i.e., $\hbar\omega \gg k_B T$, we obtain the *Wien formula*

$$u(\omega, T) \approx \frac{\hbar}{\pi^2 c^3} \omega^3 e^{-\beta\hbar\omega},$$

which was empirically known to describe the high frequency spectrum quite accurately. Planck's law interpolates nicely between the two limits. We can write Planck's law in the form

$$u(\omega, T) = \frac{(k_B T)^3}{\pi^2 c^3 \hbar^2} \frac{x^3}{e^{-x} - 1} \quad (13.2)$$

with $x = \frac{\hbar\omega}{k_B T}$. The maximum is found at $x_m = 2.82$. The corresponding frequency of the maximum $\hbar\omega_m = 2.82 k_B T$ scales linearly with temperature. This is known as *Wien's displacement law* (Verschiebungsgesetz).

Exercise 13.23: Estimate the position of the maximum in $u(\omega, T)$ [Eq. (13.5)].

13.6 Phonons

Lattice vibrations in solids can be treated within the harmonic approximation in a very similar way as photons. The quantization of the translational degrees of freedom relative to the lattice equilibrium positions leads to the elementary excitations or normal modes known as phonons, in a very similar way as photons are derived from the quantization of the electromagnetic field. Assuming periodic boundary conditions in a volume $V = L^3$, the allowed phonon wave-vectors are

$$\vec{k} = \left(\frac{2\pi}{L}\right) \vec{m} \quad \text{with} \quad \vec{m} \in \mathbb{Z}.$$

For each \vec{k} there are three possible polarizations: a longitudinal mode and two transversal modes. The corresponding frequencies are denoted by $\omega_\alpha(\vec{k})$. The density of phonon states $\rho(\omega) = \sum_\alpha \rho_\alpha(\omega)$ per unit frequency is given by

$$\rho(\omega) = \sum_{\vec{k}, \alpha} \delta[\omega_\alpha(\vec{k}) - \omega] = \frac{V}{(2\pi)^3} \sum_\alpha \int \delta[\omega_\alpha(\vec{k}) - \omega] dk^3.$$

Assuming for simplicity a linear isotropic dispersion relation $\omega_k = ck$, with the same speed of sound c for longitudinal and transversal modes, we have

$$\rho(\omega) = 3 \frac{V}{(2\pi)^3} \frac{4\pi k^2}{c} = V \frac{3\omega^2}{2\pi^2 c^3}. \quad (13.3)$$

In contrast to photons, phonons frequencies are bounded by the constraint that the total number of eigenmodes must be equal to the number of degrees of freedom $3N$. This means that there is an upper bound ω_D for the phonon frequencies which satisfies

$$\int_0^{\omega_D} \rho(\omega) d\omega = 3N, \quad (13.4)$$

where ω_D is the largest attainable *Debye frequency*. Replacing Eq. (13.3) in (13.4) one obtains

$$\omega_D = c \left(\frac{6\pi^2}{v}\right)^{1/3}, \quad (13.5)$$

where $v = V/N$ is the volume per atom. That phonon frequencies are bounded corresponds physically to the fact that in a discrete lattice the wavelengths cannot be shorter than the lattice spacing $a = L/N^{1/3} = v^{1/3}$. In the Debye model the shortest wavelength is $\lambda_D = 2\pi/k_D = 2\pi c/\omega_D = (4\pi/3)^{1/3} v^{1/3} \simeq 1.6 v^{1/3}$, which is of the order of the interatomic distance $v^{1/3}$.

The vibrational state $|\{n_{\vec{k}\alpha}\}\rangle$ of the lattice is defined univocally by the occupation numbers $n_{\vec{k}\alpha}$ of all phonons modes. The total energy is given by

$$E(\{n_{\vec{k}\alpha}\}) = \sum_{\vec{k},\alpha} (n_{\vec{k}\alpha} + \frac{1}{2}) \hbar\omega_{\alpha}(\vec{k}).$$

In the limit of $V \rightarrow \infty$ we can write

$$\begin{aligned} E &= \int_0^{\omega_D} \rho(\omega) \left(\frac{1}{e^{\beta\hbar\omega} - 1} + \frac{1}{2} \right) \hbar\omega d\omega \\ &= E_0 + \frac{3V}{2\pi^2 c^3} \int_0^{\omega_D} \omega^2 \frac{\hbar\omega}{e^{\beta\hbar\omega} - 1} d\omega, \end{aligned} \quad (13.6)$$

where $E_0 = (3V\hbar/4\pi^2 c^3) \int_0^{\omega_D} \omega^3 d\omega = (9/8)N\hbar\omega_D$ stands for the ground-state zero-point-motion energy, which is of the order of $\hbar\omega_D$ per atom. From Eq. (13.6) one obtains the heat capacity at constant volume $C_V = (\partial E/\partial T)|_V$. It has the form

$$C_V = \begin{cases} 3Nk_B \left[1 - \frac{1}{20} \left(\frac{T_D}{T} \right)^2 + \dots \right] & T \gg T_D \\ \frac{12\pi^4}{5} Nk_B \left(\frac{T}{T_D} \right)^3 + O(e^{-T_D/T}) & T \ll T_D \end{cases}$$

where $T_D = \hbar\omega_D/k_B$ is the *Debye temperature*. Typical values of T_D for pure elements are in the range of 200–600 K although some elements can have values as low as 40 K (Cs) or more than 2000 K (C). Notice that C_V vanishes as T^3 for $T \rightarrow 0$, as in the case of photons. In metals, and at sufficiently low temperatures, the free-electron-like contribution to C_V , which is linear in T , always dominates over the phonon contribution. At temperatures that are large in comparison with T_D we approach the classical limit, in which the heat capacity takes a constant value equal to k_B per degree of freedom. This is known as the Dulong-Petit law. The high-temperature value of C_V follows from the equipartition theorem which states that the classical averages of the kinetic energy and of the interaction energy in an harmonic potential are both equal to $k_B T/2$ for each degree of freedom (see Sec. 10.3). Finally, notice the profound contrast to photons, whose heat capacity diverges as T^3 since their energies are unbounded.

14 Bose-Einstein condensation

As a final example of a non-interacting Bose gas we consider non-relativistic particles having a finite mass m whose number N is conserved. We are interested in understanding the properties when the total average number of particles $N = \langle \hat{N} \rangle$ is fixed, rather than in the case where the chemical potential is held constant. The thermodynamic potentials, however, can be calculated in the grand canonical ensemble, which is by far simpler than in the canonical ensemble where no particle exchange with the environment is allowed. Moreover, we are interested in the macroscopic limit. Therefore, consider the limit of V and $N \rightarrow \infty$ keeping constant the particle density $\nu = N/V$ or, equivalently, the volume per particle $v = V/N$.

As already discussed, the grand canonical potential, equation of state and particle numbers are given by

$$\begin{aligned}\Phi &= k_B T \sum_{\vec{p}} \ln \left[1 - e^{-\beta(\varepsilon_{\vec{p}} - \mu)} \right], \\ \Phi &= k_B T \sum_{\vec{p}} \ln \left[1 - z e^{-\beta \varepsilon_{\vec{p}}} \right],\end{aligned}\tag{14.1}$$

$$\frac{pV}{k_B T} = \ln Z_{gc} = - \sum_{\vec{p}} \ln \left[1 - z e^{-\beta \varepsilon_{\vec{p}}} \right],\tag{14.2}$$

$$N = - \frac{\partial \Phi}{\partial \mu} = \sum_{\vec{p}} \frac{1}{z^{-1} e^{\beta \varepsilon_{\vec{p}}} - 1},\tag{14.3}$$

and the average occupation of the single-particle state \vec{p} is

$$\langle n_{\vec{p}} \rangle = \frac{1}{z^{-1} e^{\beta \varepsilon_p} - 1}.\tag{14.4}$$

In principle, in order to calculate the properties of the gas, we need to solve for $z = e^{\beta \mu}$ as a function of N from Eq. (14.3) by using the dispersion relation

$$\varepsilon_{\vec{p}} = \frac{p^2}{2m},$$

and the fact that the momenta take the values

$$\vec{p} = \frac{2\pi\hbar}{L} \vec{k}$$

with $\vec{k} \in \mathbb{Z}^3$. One would then replace z as function of N , V and T in Eqs. (14.1) and (14.2) in order to obtain Φ and the equation of state. However, before doing that (or in order to succeed in doing that) we need to analyze the properties of $\langle n_{\vec{p}} \rangle$ and N as given by Eqs. (14.3) and (14.4) in some detail. Since

$$\langle n_{\vec{p}} \rangle = \frac{1}{e^{\beta(\varepsilon_{\vec{p}} - \mu)} - 1}$$

is positive for all \vec{p} , we must have $\mu \leq \varepsilon_{\vec{p}}$ for all \vec{p} , which implies that $\mu \leq 0$. Consequently, the fugacity $z = e^{\beta\mu}$ satisfies

$$0 < z \leq 1.$$

Clearly, both limits can be approached arbitrarily close: $z \rightarrow 0$ for $T \rightarrow \infty$, since $\mu \rightarrow -\infty$ when $\beta \rightarrow 0$, while $z \rightarrow 1$ for $T \rightarrow 0$, since the ground-state of the ideal Bose gas has all its particles occupying the $p = 0$ state ($\langle n_{\vec{p}} \rangle = 0, \forall p \neq 0$). In order to see this more clearly we express z in terms of occupation $\langle n_0 \rangle$ of the $p = 0$ state:

$$\langle n_0 \rangle = \frac{1}{z^{-1} - 1} = \frac{z}{1 - z} \quad \Rightarrow \quad z = \frac{\langle n_0 \rangle}{\langle n_0 \rangle + 1}.$$

For $T \rightarrow \infty$, $\langle n_0 \rangle \rightarrow 0$, since $\langle n_{\vec{p}} \rangle$ becomes independent of \vec{p} and $\sum_{\vec{p}} \langle n_{\vec{p}} \rangle = N$. Thus, $z \rightarrow 0$ for $T \rightarrow \infty$, irrespectively of the value of N . For $T \rightarrow 0$, $\langle n_{\vec{p}} \rangle \rightarrow 0$ for all $\vec{p} \neq 0$. Consequently, $\langle n_0 \rangle \rightarrow N$ and $z \rightarrow N/(N + 1)$, which tends to 1 for N or $V \rightarrow \infty$. We will actually see that in the limit of $V \rightarrow \infty$ the fugacity z is strictly equal to 1 ($\mu = 0$) below a given finite critical temperature T_c . Therefore, we must consider $z = 1$ as a possible attainable value in the macroscopic limit.

In order to be able to take the limit $V \rightarrow \infty$ in Eqs. (14.2) and (14.3) we need to verify if any element of the sum can diverge for $V \rightarrow \infty$, since these terms need to be taken off the sum, before replacing the latter by an integral. We must therefore pay attention to the terms having $\varepsilon_{\vec{p}} \rightarrow 0$ when $z \rightarrow 1$.

Let us consider the density of bosons occupying a given state $\vec{p} \neq 0$, which is given by

$$\frac{\langle n_{\vec{p}} \rangle}{V} = \frac{1}{V} \frac{1}{z^{-1} e^{\beta\varepsilon_p} - 1} \leq \frac{1}{V} \frac{1}{e^{\beta\varepsilon_p} - 1}, \quad (14.5)$$

where in the last inequality we have used that $z \leq 1$ and $\varepsilon_p \geq 0$. In the macroscopic limit the smallest ε_p tends to 0, even for $p \neq 0$. Therefore, $(e^{\beta\varepsilon_p} - 1)^{-1}$ becomes divergent for $p \rightarrow 0$. In order to determine $\langle n_{\vec{p}} \rangle/V$ for large V , we calculate the leading contributions:

$$\begin{aligned} e^{\beta\varepsilon_p} - 1 &= \beta \frac{p^2}{2m} + O(p^4) \\ &= \frac{\beta}{2m} \frac{(2\pi\hbar)^2}{V^{2/3}} k^2 + O\left(\frac{1}{V^{4/3}}\right), \end{aligned}$$

where we have used that $p^2 = \frac{(2\pi\hbar)^2}{V^{2/3}} k^2$ with $k^2 = k_x^2 + k_y^2 + k_z^2$ and $\vec{k} \in \mathbb{Z}^3$. Replacing in (14.5) we obtain

$$\frac{\langle n_{\vec{p}} \rangle}{V} \leq \frac{2m}{\beta(2\pi\hbar)^2 k^2} \frac{1}{V^{1/3}} \xrightarrow{V \rightarrow \infty} 0$$

for all $\vec{p} \neq 0$. One concludes that none of the occupation densities $\langle n_{\vec{p}} \rangle/V$ diverges for $\vec{p} \neq 0$. However, this is obviously not true for the $p = 0$ state, since

$$\langle n_0 \rangle = \frac{1}{z^{-1} - 1} \quad (14.6)$$

diverges for $z \rightarrow 1$. This term (and only this term) needs to be singled-out before replacing the sum over \vec{p} by an integral in Eq. (14.3). In the limit of large V we therefore have

$$\frac{N}{V} = \frac{4\pi}{(2\pi\hbar)^3} \int_0^\infty p^2 \frac{1}{z^{-1}e^{\beta p^2/2m} - 1} dp + \frac{1}{V} \frac{z}{1-z}.$$

Replacing with the variable $x = p/\sqrt{2mk_B T}$ we have

$$\begin{aligned} \frac{N}{V} &= 4\pi \frac{(\sqrt{2mk_B T})^3}{(2\pi\hbar)^3} \int_0^\infty \frac{x^2 dx}{z^{-1}e^{x^2} - 1} + \frac{1}{V} \frac{z}{1-z} \\ &= \frac{1}{\lambda^3} g_{3/2}(z) + \frac{1}{V} \frac{z}{1-z}, \end{aligned} \quad (14.7)$$

where

$$g_{3/2}(z) = \frac{4}{\sqrt{\pi}} \int_0^\infty \frac{x^2}{z^{-1}e^{x^2} - 1} dx \quad (14.8)$$

is a simple function of z . Introducing the density of particles per unit volume $\nu = N/V = 1/v$ we can write

$$\nu = \frac{N}{V} = \frac{\langle n^* \rangle}{V} + \frac{\langle n_0 \rangle}{V}, \quad (14.9)$$

where

$$\nu^* = \frac{\langle n^* \rangle}{V} = \frac{1}{V} \sum_{\vec{p} \neq 0} \langle n_{\vec{p}} \rangle = \frac{1}{\lambda^3} g_{3/2}(z). \quad (14.10)$$

is the density of excited particles occupying states having $\vec{p} \neq 0$, and

$$\nu_0 = \frac{\langle n_0 \rangle}{V} = \frac{1}{V} \frac{z}{1-z} \quad (14.11)$$

is the density of particles occupying the lowest single-particle state with $p = 0$. Notice that ν_0 may well be different from zero in the macroscopic limit, if $z \rightarrow 1$ for $V \rightarrow \infty$. This is certainly the case for $T = 0$, where $\nu_0 = \nu$. In general, having $\nu_0 \neq 0$ means that a finite macroscopic fraction of the bosons occupies the $p = 0$ state.

Let us now analyze the functions

$$g_{3/2}(z) = \frac{4}{\sqrt{\pi}} \int_0^\infty \frac{x^2}{z^{-1}e^{x^2} - 1} dx = \sum_{l=1}^\infty \frac{z^l}{l^{3/2}} \quad (14.12)$$

and

$$\frac{dg_{3/2}}{dz} = \frac{4}{\sqrt{\pi}} \int_0^\infty \frac{x^2 e^{x^2} (1/z)^2}{(z^{-1}e^{x^2} - 1)^2} dx > 0 \quad (14.13)$$

for $0 \leq z \leq 1$. $g_{3/2}(z)$ is always finite as can be easily seen by looking at the integral or the Taylor expansion. However, $\frac{dg_{3/2}}{dz} \rightarrow \infty$ for $z \rightarrow 1$ since the integral (14.13) diverges at its lower bound. Moreover, $g_{3/2}(z)$ is a monotonously increasing function of z , which reaches its maximal finite value $g_{3/2}(1) = 2.612$ for $z = 1$. The fact that the largest possible value of $g_{3/2}(z)$ and thus of ν^* is finite has very important consequences.

Consider a system with a given density of particles $\nu = N/V$ or volume per particle $v = V/N = 1/\nu$ at a sufficiently high temperature T , so that

$$\lambda^3 \nu = \frac{\lambda^3}{v} < g_{3/2}(1) = 2.612.$$

Recalling that

$$\lambda = \frac{2\pi\hbar}{\sqrt{2\pi m k_B T}} = \sqrt{\frac{2\pi}{m k_B T}} \hbar,$$

this corresponds to a temperature

$$T > T_c = \frac{2\pi\hbar^2}{m k_B} \frac{1}{[v g_{3/2}(1)]^{2/3}}. \quad (14.14)$$

From Eq. (14.7) we have

$$\lambda^3 \nu = \frac{\lambda^3}{v} = g_{3/2}(z) + \frac{\lambda^3}{V} \frac{z}{1-z} \quad (14.15)$$

$$= \lambda^3 \left(\frac{\langle n^* \rangle}{V} + \frac{\langle n_0 \rangle}{V} \right). \quad (14.16)$$

Since $\lambda^3/v < g_{3/2}(1)$ we must have $z < 1$ and

$$g_{3/2}(z) = \frac{\lambda^3}{v} \quad \text{for} \quad \frac{\lambda^3}{v} < g_{3/2}(1) = 2.612.$$

Therefore, in the temperature range $T > T_c$ we have

$$\frac{\langle n_0 \rangle}{V} = \frac{1}{V} \frac{z}{1-z} \rightarrow 0 \quad \text{for} \quad V \rightarrow \infty.$$

In particular for $T \rightarrow T_c$ ($T > T_c$) we have $\lambda^3/v \rightarrow g_{3/2}(1)$ and $z \rightarrow 1$. Precisely at $T = T_c$, where $\lambda^3/v = g_{3/2}(1)$, we still have $\nu_0 = \langle n_0 \rangle/V = 0$, since $\nu^* = \langle n^* \rangle/V = g_{3/2}(z)/\lambda^3$ is equal to $g_{3/2}(1)/\lambda^3 = \nu$. In other words, at $T \geq T_c$ the density of excited particles ν^* (i.e., occupying $\vec{p} \neq 0$ states) is equal to the total density ν . However, for $T < T_c$ the temperature is so low that $\lambda^3/v > g_{3/2}(1)$, a value that $g_{3/2}(z)$ cannot reach. The only possibility of satisfying Eq. (14.15) is to set $z = 1$, which gives the maximum possible density of excited particles

$$\frac{\langle n^* \rangle}{V} = \frac{g_{3/2}(1)}{\lambda^3}, \quad (14.17)$$

and to place all the remaining particles

$$\frac{\langle n_0 \rangle}{V} = \frac{N}{V} - \frac{\langle n^* \rangle}{V} = \frac{1}{v} - \frac{g_{3/2}(1)}{\lambda^3} \quad (14.18)$$

in the $\vec{p} = 0$ state. Thus, for $T < T_c$, a finite macroscopic fraction of the particles

$$\frac{\langle n_0 \rangle}{V} = \frac{N}{V} - \frac{g_{3/2}(1)}{\lambda^3} = \frac{1}{v} - \frac{g_{3/2}(1)}{\lambda^3} \quad (14.19)$$

condensates in the lowest-energy $\vec{p} = 0$ state. This remarkable phenomenon is known as Bose-Einstein condensation. In the so-called mixed phase, i.e., for $T < T_c$ we have a macroscopic occupation $\langle n_0 \rangle$ of the $\vec{p} = 0$ state. $\langle n_0 \rangle/V$ increases proportional to $T^{3/2}$ with decreasing temperature until at $T = 0$ we reach $\langle n_0 \rangle = N$.

Knowing that $\lambda = 2\pi\hbar/\sqrt{2\pi mk_B T}$ and using Eq. (14.14) for T_c we have

$$\frac{\lambda^3(T_c)}{v} = g_{3/2}(1), \quad \left(\frac{2\pi\hbar^2}{mk_B T_c} \right)^{3/2} = v g_{3/2}(1), \quad T_c = \frac{2\pi\hbar^2}{mk_B} \frac{1}{[v g_{3/2}(1)]^{2/3}} \quad (14.20)$$

and thus

$$\left(\frac{T}{T_c} \right)^{3/2} = \frac{v g_{3/2}(1)}{\lambda^3}, \quad (14.21)$$

which allows us to bring Eq. (14.19) in the form

$$\frac{\langle n_0 \rangle}{V} = \frac{1}{v} \left(1 - \frac{v g_{3/2}(1)}{\lambda^3} \right) = \frac{N}{V} \left[1 - \left(\frac{T}{T_c} \right)^{3/2} \right] \quad (14.22)$$

for $T \leq T_c$. For $T > T_c$, $\langle n_0 \rangle/V = 0$. In the condensed phase ($T < T_c$) the fugacity is fixed to $z = 1$, while in the normal phase ($T > T_c$) z is obtained as usual as the root of $\lambda^3/v = g_{3/2}(z)$.

[Grafik: mixed phase: normal gas + zero pressure "liquid"]

For each temperature T there is a critical density $\nu_c = 1/v_c$ above which (or a critical volume per particle v_c below which) the Bose-Einstein condensation takes place. This is given by

$$v_c(T) = \frac{\lambda^3}{g_{3/2}(1)} \quad (14.23)$$

In this context it is meaningful to define the order parameter

$$\eta = \frac{\langle n_0 \rangle}{N} = \begin{cases} 0 & \text{for } T > T_C \\ 1 - \frac{v g_{3/2}(1)}{\lambda^3} = 1 - \left(\frac{T}{T_c} \right)^{3/2} = 1 - \frac{v}{v_c} & \text{for } T < T_c, \end{cases} \quad (14.24)$$

which represents the proportion of the total number of bosons which are condensed in the $p = 0$ state. In the normal phase ($T > T_c$ or $v > v_c$) we have $\eta = 0$, whereas in the condensed phase we have $\eta > 0$. From Eq. (14.14), taking into account that $T_c \propto v^{-2/3}$, we can draw the phase boundary in a T - V diagram.

Before looking at the equation of state, it is useful to derive some relations valid in the normal phase, i.e., $\lambda^3/v \leq g_{3/2}(1)$, $T > T_c$ or $v > v_c$. For $\lambda^3/v \leq g_{3/2}(1)$, z is given by

$$g_{3/2}(z) = \frac{\lambda^3}{v}. \quad (14.25)$$

At the transition we have

$$g_{3/2}(1) = \frac{\lambda^3}{v_c} \quad (14.26)$$

for all T , and

$$\begin{aligned} g_{3/2}(1) &= \frac{\lambda(T_c)^3}{v} \\ &= \frac{1}{v} \left(\frac{2\pi\hbar^2}{mk_B T_c} \right)^{3/2} \end{aligned} \quad (14.27)$$

for all v . Combining (14.25) and (14.26) we have

$$\frac{g_{3/2}(z)}{g_{3/2}(1)} = \frac{v_c}{v}, \quad (14.28)$$

and combining (14.25) and (14.27) we have

$$\frac{g_{3/2}(z)}{g_{3/2}(1)} = \frac{\lambda^3}{\lambda(T_c)^3} = \left(\frac{T_c}{T} \right)^{3/2}. \quad (14.29)$$

Let us now verify the behavior of the different terms in the sum over \vec{p} , which gives the grand canonical potential Φ , for $V \rightarrow +\infty$. Before replacing the sum over \vec{p} by an integral we must single out any possibly diverging terms. For finite V we have

$$\Phi = k_B T \sum_{\vec{p}} \ln \left(1 - z e^{-\beta \varepsilon_{\vec{p}}} \right) < 0. \quad (14.30)$$

It is easy to see that $|\ln(1 - z e^{-\beta \varepsilon_p})|/V \rightarrow 0$ for $V \rightarrow \infty$ for all $\vec{p} \neq 0$, since $|\ln(1 - z e^{-\beta \varepsilon_p})| \leq |\ln(1 - e^{-\beta \varepsilon_p})| \simeq \ln(\beta \varepsilon_p) \propto (-2/3) \ln(V)$. In the limit of $V \rightarrow \infty$, we need to separate only the potentially diverging $\vec{p} = 0$ term. We can thus write

$$-p = \frac{\Phi}{V} = k_B T \left\{ \frac{4\pi}{(2\pi\hbar)^3} \int_0^\infty p^2 \ln \left(1 - z e^{-\beta p^2/2m} \right) dp + \frac{1}{V} \ln(1 - z) \right\} \quad (14.31)$$

Within the usual change of variable $x = \frac{p}{\sqrt{2mk_B T}}$ and knowing that $\lambda = \sqrt{2\pi\hbar/mk_B T}$, we have

$$\frac{\Phi}{V} = -k_B T \left\{ \frac{1}{\lambda^3} g_{5/2}(z) - \frac{1}{V} \ln(1-z) \right\} \quad (14.32)$$

where

$$g_{5/2}(z) = -\frac{4}{\sqrt{\pi}} \int_0^\infty x^2 \ln(1 - ze^{-x^2}) dx = \sum_{l=1}^{\infty} \frac{z^l}{l^{5/2}}. \quad (14.33)$$

Knowing that $pV = -\Phi$ we can write more simply

$$\frac{p}{k_B T} = \frac{1}{\lambda^3} g_{5/2}(z) - \frac{1}{V} \ln(1-z). \quad (14.34)$$

Since in the normal phase [i.e., for $\lambda^3/v < g_{3/2}(1)$] we have $z < 1$, it is clear that the $p = 0$ term vanishes. But also in the condensed (mixed) phase this term does not contribute. In fact,

$$\langle n_0 \rangle = \frac{1}{z^{-1} - 1} \Rightarrow 1 - z = \frac{1}{\langle n_0 \rangle + 1} \propto \frac{1}{V} \quad (14.35)$$

in the condensed phase. Consequently,

$$\frac{1}{V} \ln(1-z) \approx -\frac{1}{V} \ln V \xrightarrow{V \rightarrow \infty} 0. \quad (14.36)$$

The equation of state reads therefore

$$\frac{p}{k_B T} = \frac{1}{\lambda^3} g_{5/2}(z) = \begin{cases} \frac{1}{\lambda^3} g_{5/2}(z) & \text{for } v > v_c \\ \frac{1}{\lambda^3} g_{5/2}(1) & \text{for } v < v_c. \end{cases} \quad (14.37)$$

We should therefore analyze $g_{5/2}(z)$ for $0 \leq z \leq 1$. For $z \ll 1$ we have $g_{5/2}(z) \approx z + O(z^2)$. Clearly,

$$g_{5/2}(z) = \sum_{l \geq 1} \frac{z^l}{l^{5/2}} < \sum_{l \geq 1} \frac{z^l}{l^{3/2}} = g_{3/2}(z) \quad \forall z. \quad (14.38)$$

Consequently, $g_{5/2}(z)$ is finite $\forall z \in [0, 1]$, and also $\frac{dg_{5/2}}{dz}$ is finite, since $g_{3/2}(z) = z \frac{dg_{5/2}}{dz}$ is finite. Thus, $g_{5/2}$ increases monotonously. Its largest value in $[0, 1]$ is $g_{5/2}(1) = 1,3415$.

[Grafik]

Remarkably, the pressure is independent of V or $v = V/N$ in the mixed phase. Indeed, for $v < v_c$ we can write

$$\frac{p}{k_B T} = \frac{1}{\lambda^3} g_{5/2}(1) = \frac{1}{v_c} \frac{g_{5/2}(1)}{g_{3/2}(1)}.$$

The reason is that only the excited particles ($p \neq 0$) contribute to the pressure, since for $v < v_c$ we can write

$$\frac{pv}{k_B T} = \frac{v}{v_c} \frac{g_{5/2}(1)}{g_{3/2}(1)} = \frac{\langle n^* \rangle}{N} \frac{g_{5/2}(1)}{g_{3/2}(1)}$$

Remember that

$$\frac{\langle n_0 \rangle}{V} = 1 - \frac{v}{v_c} \quad \Rightarrow \quad \frac{\langle n^* \rangle}{N} = 1 - \frac{n_0}{V} = \frac{v}{v_c}.$$

Maybe more clearly

$$pV = \langle n^* \rangle k_B T \frac{g_{5/2}(1)}{g_{3/2}(1)}$$

for $v < v_c$ or $T < T_c$, where $g_{5/2}(1)/g_{3/2}(1) \simeq 0.5135 < 1$. Each excited particle contributes to the pressure by about half the contribution of the classical Boltzmann gas. This reflects again the effective attraction between Bose particles, which we already observed in the high-temperature limit. The condensed part of the mixed phase has zero pressure. One therefore speaks about the condensate as a zero pressure liquid. This is of course an idealization in the limit $v \rightarrow 0$, due to the neglected interactions among the particles.

Along the transition line we have

$$\frac{\lambda^3}{v_c} = g_{3/2}(1) \Rightarrow \lambda^2 = \frac{2\pi\hbar^2}{mk_B T} = [g_{3/2}(1)v_c]^{2/3} \quad \Rightarrow \quad k_B T = \frac{2\pi\hbar^2}{m} \frac{1}{[g_{3/2}(1)v_c]^{3/2}}.$$

The pressure along the transition line is

$$p(v_c) = p_c = \frac{k_B T}{v_c} \frac{g_{5/2}(1)}{g_{3/2}(1)} = \frac{2\pi\hbar^2}{m} \frac{g_{5/2}(1)}{g_{3/2}(1)^{5/3}} \frac{1}{v_c^{5/3}}$$

In order to obtain the pressure as a function of temperature we use, for $T \leq T_c$, the relation

$$\begin{aligned} \frac{p}{k_B T} &= \frac{1}{\lambda^3} g_{5/2}(1) \\ p &= \left(\frac{m}{2\pi\hbar^2} \right)^{3/2} (k_B T)^{5/2} g_{5/2}(1), \end{aligned}$$

and for $T = T_c$ we have

$$p(T_c) = p_c = \frac{N}{V} k_B T_c \frac{g_{5/2}(1)}{g_{3/2}(1)} = 0.5135 \frac{N}{V} k_B T_c.$$

For $T > T_c$, p approaches the classical limit from below.

One may easily express all the other thermodynamic properties in terms of $g_{5/2}$ and $g_{3/2}$ by taking the appropriate derivatives of $\Phi = -pV$. For example, the entropy is given by

$$\frac{S}{Nk_B} = \begin{cases} \frac{5}{2} \frac{v}{\lambda^3} g_{5/2}(z) - \ln z & T \geq T_c \\ \frac{5}{2} \frac{v}{\lambda^3} g_{5/2}(1) & T \leq T_c \end{cases} \quad (14.39)$$

Since S is continuous at T_c , there is no latent heat as in a first order solid-liquid transition. In the mixed phase, $T \leq T_c$, we have

$$\begin{aligned} S &= \frac{5}{2} k_B N \frac{v}{v_c} \frac{g_{5/2}(1)}{g_{3/2}(1)} = \frac{5}{2} k_B \langle n^* \rangle \frac{g_{5/2}(1)}{g_{3/2}(1)} \\ &= \frac{5}{2} k_B N \left(\frac{T}{T_c} \right)^{3/2} \frac{g_{5/2}(1)}{g_{3/2}(1)}. \end{aligned}$$

Only the excited particles ($p \neq 0$) contribute to the entropy, with an entropy per excited particle

$$\frac{S}{\langle n^* \rangle} = k_B \frac{5}{2} \frac{g_{5/2}(1)}{g_{3/2}(1)}$$

that is independent of $\langle n^* \rangle$ and T . The particles in the “liquid” (i.e., having $p = 0$) do not contribute at all to S . They have zero entropy, since they all form one and the same many-body state. Every time a particle condensates (i.e., $\langle n^* \rangle V \rightarrow \langle n^* \rangle V - 1$) there is a finite change $\Delta S = k_B \frac{5}{2} \frac{g_{5/2}(1)}{g_{3/2}(1)}$ in the total entropy of the system (not in the entropy per unit volume for $V \rightarrow \infty$). This supports the interpretation or analogy with a first-order phase transition. The Bose-Einstein condensation has also features that are characteristic of a second-order phase transition. One of the them is the continuous increase of the order parameter $\eta = \langle n_0 \rangle / N$ with decreasing T , starting at the transition temperature T_c . Another one is the divergence of the compressibility κ_T at v_c for all T .

The divergence of the compressibility

$$\kappa_T = - \frac{1}{V} \frac{\partial V}{\partial p} \Big|_T = - \frac{1}{v} \frac{\partial v}{\partial p} \Big|_T \xrightarrow{v \rightarrow v_c} +\infty$$

can be easily demonstrated by showing that the isotherms $p(v)$ approach the critical volume v_c with zero slope. First we note that

$$\frac{\partial p}{\partial v} \Big|_T = \frac{k_B T}{\lambda^3} \frac{dg_{5/2}}{dz} \frac{\partial z}{\partial v} \Big|_T = \frac{k_B T}{\lambda^3} \frac{1}{z} g_{3/2}(z) \frac{\partial z}{\partial v} \Big|_T.$$

We also know that

$$\frac{\lambda^3}{v} = g_{3/2}(z) \quad \Rightarrow \quad - \frac{\lambda^3}{v^2} = \frac{dg_{3/2}(z)}{dz} \frac{\partial z}{\partial v} \Big|_T.$$

Consequently,

$$\begin{aligned} - \frac{\partial p}{\partial v} \Big|_T &= \frac{k_B T}{\lambda^3} g_{3/2}(z) \frac{1}{z} \frac{\lambda^3}{v^2} \frac{1}{\frac{dg_{3/2}(z)}{dz}} \\ -v \frac{\partial p}{\partial v} \Big|_T &= \kappa_T^{-1} = \frac{k_B T}{v} \frac{g_{3/2}(z)}{g_{1/2}(z)}. \end{aligned}$$

Since

$$\begin{aligned} g_{3/2}(z) &\rightarrow 2.61 \\ g_{1/2}(z) &\rightarrow +\infty \end{aligned}$$

for $z \rightarrow 1$, we have

$$\kappa_T = \frac{v}{k_B T} \frac{g_{1/2}(z)}{g_{3/2}(z)} \rightarrow +\infty$$

for $v \rightarrow v_c$ (i.e., for $z \rightarrow 1$).

Finally, let us have a look at the specific heat, which can be obtained straightforwardly from S :

$$\frac{C_V}{Nk_B} = \begin{cases} \frac{15}{4} \frac{v}{\lambda^3} g_{5/2}(z) - \frac{9}{4} \frac{g_{3/2}(z)}{g_{1/2}(z)} & T > T_C \\ \frac{15}{4} \frac{v}{\lambda^3} g_{5/2}(z). & T < T_C \end{cases}$$

At low temperatures, $C_V \propto T^{3/2}$ and vanishes for $T \rightarrow 0$, in agreement with Nernst theorem. This can be easily understood by noting that at low T we have $\langle n^* \rangle / N = v/v_c = (T/T_c)^{3/2}$ excited particles, whose energy is of the order of $k_B T$. Thus, we have $E \propto T^{5/2}$ and $C_V \propto T^{3/2}$. Notice that C_V is continuous at T_c , since $g_{1/2}(z) \rightarrow \infty$ for $z \rightarrow 1$. The value at T_c is larger than the classical limit $\frac{C_V}{Nk_B} = \frac{3}{2}$, which actually corresponds to $C_V(T \rightarrow +\infty) = \frac{3}{2} Nk_B$. In fact,

$$C_V = \frac{15}{4} \frac{\lambda(T_c)^3}{\lambda} \frac{g_{5/2}(1)}{g_{3/2}(1)} \xrightarrow{T \rightarrow T_c} \frac{15}{4} \frac{g_{5/2}(1)}{g_{3/2}(1)},$$

which yields

$$C_V(T_c) \simeq 1.926 > \frac{3}{2}.$$

The fact that $C_V > 3/2$ for large T is remarkable. It can be interpreted to be the consequence of an effective attraction between bosons, which is due to the symmetry correlations (full permutation symmetry of the many-body wave functions). Let us recall that the heat capacity C_V of an interacting classical gas is enhanced by $\langle \Delta W^2 \rangle / k_B T^2$, where $\langle \Delta W^2 \rangle$ refers to the mean square deviation of the interparticle interaction energy W , irrespectively of whether the latter is attractive or repulsive (see Sec. 10.2).

[Grafik]

Exercise 14.24: Bosons in a two-level system

Consider an open system of non-interacting Bosons in which the average number of particles N is held constant for all temperatures T and volumes V . The single-particle spectrum consists of two levels: a non-degenerate ground state $|0\rangle$ with energy $E_0 = 0$ and a g -fold degenerate excited state having an energy is $E_1 = \Delta > 0$ and a degeneracy $g = V/v_0$, proportional to the volume, where v_0 is a constant representing the volume per degree of degeneracy.

- i) Express the average ground-state occupation number n_0 and the average occupation n^* of the excited states as a function of T , Δ and chemical potential μ . How are n_0 , n^* and N related? Explain why the fugacity $z = e^{\beta\mu}$ must always lie in the range $0 < z \leq 1$ and why $z = 1$ (or equivalently $\mu = 0$) is only possible in the thermodynamic limit where $N \rightarrow \infty$.
 - ii) Find the densities of bosons $\nu_0 = n_0/V$ in the single-particle ground state and $\nu^* = n^*/V$ in the excited states as a function of z , T and Δ . Express z as a function of the total boson density $\nu = N/V$. Find the density of bosons ν_0 occupying the ground state in the limit of $V \rightarrow \infty$ for $z < 1$. Show that $\lim_{V \rightarrow \infty} \nu_0 = \bar{\nu}_0 > 0$ is only possible when at the same time $z \rightarrow 1$.
Hint: For the sake of clarity one may introduce the variable $y = e^{\beta\Delta} > 1$.
 - iii) Plot or sketch ν^* as a function of z for $\Delta > 0$ at different fixed temperatures T .
 - iv) Find the critical Bose-Einstein-condensation density ν_c above which ν_0 is finite in the thermodynamic limit ($N \rightarrow \infty$).
 - v) Find the critical temperature T_c as a function of ν , v_0 and Δ below which the system is in the condensed phase.
-

Exercise 14.25: Two-level condensation: Work out Einstein's argument for the condensation of indistinguishable particles.

Consider a system of $N = n_1 + n_2$ particles, which can occupy two levels with energies ε_1 and ε_2 with $\varepsilon_1 < \varepsilon_2$. The number of particles in level ε_i is denoted by n_i . In the following the thermodynamic limit $N \gg 1$ may be assumed.

i) Distinguishable particles

(a) Find the microcanonical partition function Ω and entropy S as a function of n_i , in the case of *distinguishable particles*.

(b) Determine the equilibrium value of n_1/n_2 as a function of $\Delta\varepsilon = \varepsilon_2 - \varepsilon_1$, for example, by minimizing the free energy F . What do you expect for $\Delta\varepsilon = 0$?

ii) Repeat the calculation for bosons:

(a) Find the microcanonical partition function Ω and entropy S as a function of n_i , in the case of bosons.

(b) Determine the equilibrium value of n_1/n_2 as a function of $\Delta\varepsilon = \varepsilon_2 - \varepsilon_1$, for example, by minimizing the free energy F . Analyze the cases $\Delta\varepsilon = 0$ and $\Delta\varepsilon > 0$.

iii) Condensation: One says that the two-level system is condensed, when $n_1 = N$ and $n_2 = 0$ ($\varepsilon_1 < \varepsilon_2$). Find out the conditions for condensation in the case of bosons and distinguishable particles and bosons for $T = 0$ and $T > 0$ ($\Delta\varepsilon > 0$).

Exercise 14.26: The Bose-Einstein condensation appears very clearly, mathematically, as a consequence of the fact that $g_{3/2}(z)$ is bounded for $z \in [0, 1]$. However, this does not provide any physical understanding of the reason why the normal or excited phase cannot sustain an arbitrary density of particles in equilibrium. Why is $\langle n^* \rangle / N$ bounded? Why is the system not willing to keep $\langle n^* \rangle / N = 1$ when the temperature decreases? Notice that the contribution of $\langle n_0 \rangle$ to the entropy is zero, while the contribution of $\langle n^* \rangle$ is not. One could therefore expect that having the largest $\langle n^* \rangle$ could be more favorable.

A plausible explanation (following Einstein's argument) is that the entropy S , regarded as a function of $\langle n^* \rangle$ [$\langle n_0 \rangle = N - \langle n^* \rangle$] increases too slowly with $\langle n^* \rangle$ as compared to the internal energy E . In this case a critical temperature would be needed in order that all the particles are excited (i.e., in order to have $\langle n^* \rangle = N$). In fact, it is always possible to consider a boson gas with an arbitrary number $\langle n^* \rangle$ of particles in the states having $\vec{p} \neq 0$ —even if this does not coincide with the actual value of $\langle n^* \rangle$ at equilibrium—and to compute the energy and entropy of this gas as a function of T and V under the constraint $\sum_{\vec{p} \neq 0} \langle n_{\vec{p}} \rangle = \langle n^* \rangle$. This corresponds to Landau's microcanonical calculation of the entropy of what he calls “an ideal gases out of equilibrium”.

The exercise is then to compute $S = S(\langle n^* \rangle)$, and $E = E(\langle n^* \rangle)$, to derive the free energy $F = E - TS$, and to analyze how E , S and F depend on $\langle n^* \rangle$. One expects that above T_c , or for $v > v_c$, $\partial F / \partial \langle n^* \rangle < 0$ for all $\langle n^* \rangle$, in which case $\langle n^* \rangle$ is maximal, while for $T < T_c$ one should have $\partial F / \partial \langle n^* \rangle > 0$ for $\langle n^* \rangle = N$, implying that a fraction of the particles condensates in the ground state ($p = 0$).

Exercise 14.27: Investigate the Bose-Einstein condensation in systems with arbitrary dimensions D , where $\varepsilon_{\vec{p}} = \frac{1}{2m} \sum_{\alpha=1}^D p_{\alpha}^2$. How do T_c and v_c depend on D ?

Exercise 14.28: Investigate the effects of interactions in Bose gases. For example, one could consider a two level problem and introduce a Hubbard-like local interaction U . Or one could solve numerically a bosonic Hubbard model. It would be interesting to compare repulsive and attractive cases.

15 Bibliography

- [1] K. Huang, Statistical Mechanics, 2nd Edition, (Wiley, 1987).
- [2] L.D. Landau and E.M. Lifshitz, Statistical Physics, Part 1, Course of Theoretical Physics, Vol. 5.
- [3] R.P. Feynman, Statistical Mechanics: A Set of Lectures, Frontiers in Physics (Addison Wesley, 1981).
- [4] J.J. Sakurai, Modern quantum mechanics, (Addison Wesley, Reading, 1994).
- [5] J.D. Gibbs, Elementary principles in statistical mechanics, (Scribner's sons, New York, 1902).