Neri Merhav

Statistical Physics for Electrical Engineering

Springer

Statistical Physics for Electrical Engineering

Neri Merhav

Statistical Physics for Electrical Engineering

Springer

Neri Merhav
The Andrew and Erna Viterbi Faculty
of Electrical Engineering
Technion—Israel Institute of Technology
Technion City, Haifa
Israel

ISBN 978-3-319-62062-6 ISBN 978-3-319-62063-3 (eBook)
DOI 10.1007/978-3-319-62063-3

Library of Congress Control Number: 2017945248

© The Author(s) 2018
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

https://engineersreferencebookspdf.com

Preface

This short book is based on lecture notes of a course on statistical physics and thermodynamics, which is oriented, to a certain extent, toward electrical engineering students. The course has been taught in the Electrical Engineering department of the Technion (Haifa, Israel) ever since year 2013. The main body of the book is devoted to statistical physics, whereas much less emphasis is given to the thermodynamics part. In particular, the idea is to let the important results of thermodynamics (most notably, the laws of thermodynamics) to be obtained as conclusions from the derivations in statistical physics.

Beyond the variety of central topics in statistical physics that are important to the general scientific education of the electrical engineering student, special emphasis is devoted to subjects that are vital to the engineering education concretely. These include, first of all, quantum statistics, like the Fermi–Dirac distribution, as well as diffusion processes, which are both fundamental for deep understanding of semiconductor devices. Another important issue for the electrical engineering student is to understand mechanisms of noise generation and stochastic dynamics in physical systems, most notably, in electric circuitry. Accordingly, the fluctuation–dissipation theorem of statistical mechanics, which is the theoretical basis for understanding thermal noise processes in systems, is presented from a signals-and-systems point of view, in a way that would hopefully be understandable and useful for an engineering student, and well connected to some other important courses learned by students of electrical engineering, like courses on random processes. The quantum regime, in this context, is important too and hence provided as well. Finally, we touch very briefly upon some relationships between statistical mechanics and information theory, which is the theoretical basis for communications engineering, and demonstrate how statistical–mechanical approach can be useful for the study of information–theoretic problems. These relationships are further explored in [1], and in a much deeper manner.

In the table of contents below, chapters and sections, marked by asterisks, can be skipped without loss of continuity.

Technion City, Haifa, Israel   Neri Merhav

Reference

1. N. Merhav, Statistical physics and information theory. Foundat. Trends Commun. Inf. Theor. **6** (1–2), pp. 1–212, 2009.

Acknowledgements

I would first like to express my deep gratitude to several colleagues in the Technion Physics Department, including Profs. Dov Levine, Shmuel Fishman, and Yariv Kafri, for many fruitful discussions, and for relevant courses that they have delivered and that I have listened to. I have certainly learned a lot from them. I would like to thank Profs. Nir Tessler and Baruch Fischer, of my department, for their encouragement to develop a statistical physics course for our students. I am also grateful to Prof. Yuval Yaish of my department, who has been teaching the course too (in alternate years), for sharing with me his thoughtful ideas about the course. The lecture notes of the course have served as the basis for this book. Finally, I thank my dear wife, Ilana, and a student of mine, Mr. Aviv Lewis, for their useful comments on the English grammar, typographical errors, and style.

Contents

Introduction

Statistical physics is a branch in physics which deals with systems with a huge number of particles (or any other elementary units). For example, *Avogadro's number*, which is about 6×10^{23}, is the number of molecules in 22.4 liters of ideal gas at standard temperature and pressure. Evidently, when it comes to systems with such an enormous number of particles, there is no hope to keep track of the physical state (e.g., position and momentum) of each and every individual particle by means of the classical methods in physics, that is, by solving a gigantic system of differential equations pertaining to Newton's laws for all particles. Moreover, even if those differential equations could have been solved somehow (at least approximately), the information that they would have given us would be virtually useless. What we normally really want to know about our physical system boils down to a fairly short list of *macroscopic* quantities, such as energy, heat, pressure, temperature, volume, magnetization, and the like. In other words, while we continue to use the well-known laws of physics, even the classical ones, we no longer use them in the ordinary manner that we have known from elementary physics courses. Instead, we think of the state of the system, at any given moment, as a realization of a certain *probabilistic ensemble*. This is to say that we approach the problem from a probabilistic (or a statistical) point of view. The beauty of statistical physics is that it derives the *macroscopic* theory of thermodynamics (i.e., the relationships between thermodynamic potentials, temperature, pressure, etc.) as *ensemble averages* that stem from this probabilistic *microscopic* theory, in the limit of an infinite number of particles, that is, the *thermodynamic limit*.

The purpose of this book is to teach statistical mechanics and thermodynamics, with some degree of orientation toward students in electrical engineering. The main body of the lectures is devoted to statistical mechanics, whereas much less emphasis is given to the thermodynamics part. In particular, the idea is to let the laws of thermodynamics to be obtained as conclusions from the derivations in statistical mechanics.

Beyond the variety of central topics in statistical physics that are important to the general scientific education of the electrical engineering student, special emphasis is devoted to subjects that are vital to the engineering education concretely. These include, first of all, quantum statistics, like the Fermi–Dirac distribution, as well as diffusion processes, which are both fundamental for understanding semiconductor devices. Another important issue for the electrical engineering student is to understand mechanisms of noise generation and stochastic dynamics in physical systems, most notably, in electric circuitry. Accordingly, the fluctuation–dissipation theorem of statistical mechanics, which is the theoretical basis for understanding thermal noise processes and physical systems, is presented from the standpoint of a system with an input and output, in a way that would be understandable and useful for an engineer, and well related to other courses in the undergraduate curriculum of electrical engineering, like courses on random processes. This engineering perspective is not available in standard physics textbooks. The quantum regime, in this context, is important and hence provided as well. Finally, we touch upon some relationships between statistical mechanics and information theory, and demonstrate how the statistical–mechanical approach can be useful for the study of information theoretic problems. These relationships are further explored, and in a much deeper manner, in [1].

Most of the topics in this book are covered on the basis of several other well-known books on statistical mechanics. However, several perspectives and mathematical derivations are original and new (to the best of the author's knowledge). The book includes fairly many examples, exercises, and figures, which will hopefully help the student to grasp the material better.

It is assumed that the reader has prior background in the following subjects: (i) elementary calculus and linear algebra, (ii) basics of quantum mechanics, and (iii) fundamentals of probability theory. Chapter 7 assumes also basic background in signals-and-systems theory, as well as the theory of random processes, including the response of linear systems to random input signals.

Reference

1. N. Merhav, Statistical physics and information theory. Foundat. Trends Commun. Inf. Theor. **6** (1–2), pp. 1–212, 2009.

Chapter 1
Kinetic Theory and the Maxwell Distribution

The concept that a gas consists of many small mobile mass particles is very old–it dates back to the Greek philosophers. It has been periodically rejected and revived throughout many generations of the history of science. Around the middle of the 19th century, against the general trend of rejecting the atomistic approach, Clausius,[1] Maxwell[2] and Boltzmann[3] succeeded in developing a kinetic theory for the motion of gas molecules, which was mathematically solid, on the one hand, and agreed satisfactorily with the experimental evidence (at least in simple cases), on the other hand.

In this chapter, we present some elements of Maxwell's formalism and derivation that builds the kinetic theory of the ideal gas. It derives some rather useful results from first principles. While the main results that we shall see in this section can be viewed as a special case of the more general concepts and principles that will be provided later on, the purpose here is to give a quick snapshot on the taste of this matter and to demonstrate how the statistical approach to physics, which is based on very few reasonable assumptions, gives rise to rather far–reaching results and conclusions.

The choice of the ideal gas, as a system of many mobile particles, is a good test-bed to begin with, as on the one hand, it is simple, and on the other hand, it is not irrelevant to electrical engineering and electronics in particular. For example, the free electrons in a metal can often be considered a "gas" (albeit not an ideal gas), as we shall see later on.

---

[1] Rudolf Julius Emanuel Clausius (1822–1888) was a German physicist and mathematician who is considered to be one of the central pioneers of thermodynamics.

[2] James Clerk Maxwell (1831–1879) was a Scottish physicist and mathematician, whose other prominent achievement was formulating classical electromagnetic theory.

[3] Ludwig Eduard Boltzmann (1844–1906) was an Austrian physicist, who has founded contributions in statistical mechanics and thermodynamics. He was one of the advocators of the atomic theory when it was still very controversial.

N. Merhav, *Statistical Physics for Electrical Engineering*,
DOI 10.1007/978-3-319-62063-3_1

1.1 The Statistical Nature of the Ideal Gas

From the statistical–mechanical perspective, an ideal gas is a system of mobile particles, which interact with one another only via elastic collisions, whose duration is extremely short compared to the time elapsed between two consecutive collisions in which a given particle is involved. This basic assumption is valid as long as the gas is not too dense and the pressure that it exerts is not too high. As explained in the Introduction, the underlying idea of statistical mechanics in general, is that instead of hopelessly trying to keep track of the motion of each individual molecule, using differential equations that are based on Newton's laws, one treats the population of molecules as a statistical ensemble using tools from probability theory, hence the name statistical mechanics (or statistical physics).

What is the probability distribution of the state of the molecules of an ideal gas in equilibrium? Here, by "state" we refer to the positions and the velocities (or momenta) of all molecules at any given time. As for the positions, if gravity is neglected, and assuming that the gas is contained in a given box (container) of volume V, there is no apparent reason to believe that one region is preferable over others, so the distribution of the locations is assumed uniform across the container, and independently of one another. Thus, if there are N molecules, the joint probability density of their positions is $1/V^N$ everywhere within the container and zero outside. It is therefore natural to define the density of particles per unit volume as $\rho = N/V$.

What about the distribution of velocities? This is slightly more involved, but as we shall see, still rather simple, and the interesting point is that once we derive this distribution, we will be able to derive some interesting relationships between macroscopic quantities pertaining to the equilibrium state of the system (pressure, density, energy, temperature, etc.). As for the velocity of each particle, we will make two assumptions:

1. All possible directions of motion in space are equally likely. In other words, there are no preferred directions (as gravity is neglected). Thus, the probability density function (pdf) of the velocity vector $\vec{v} = v_x\hat{x} + v_y\hat{y} + v_z\hat{z}$ depends on $\vec{v}$ only via its magnitude, i.e., the speed $s = \|\vec{v}\| = \sqrt{v_x^2 + v_y^2 + v_z^2}$, or in mathematical terms:
$$f(v_x, v_y, v_z) = g(v_x^2 + v_y^2 + v_z^2) \tag{1.1.1}$$
for some function g.
2. The various components v_x, v_y and v_z are identically distributed and independent, i.e.,
$$f(v_x, v_y, v_z) = f(v_x)f(v_y)f(v_z). \tag{1.1.2}$$
The rationale behind identical distributions is, like in item 1 above, namely, the isotropic nature of the pdf. The rationale behind the independence assumption is that in each collision between two particles, the total momentum is conserved

> and in each component (x, y, and z) separately, so there are actually no interactions among the component momenta. Each three–dimensional particle actually behaves like three independent one–dimensional particles, as far as the momentum is concerned.

We now argue that there is only one kind of (differentiable) joint pdf $f(v_x, v_y, v_z)$ that complies with both assumptions at the same time, and this is the Gaussian density where all three components of $\vec{v}$ are independent, zero–mean and with the same variance.

To see why this is true, consider the equation

$$f(v_x)f(v_y)f(v_z) = g(v_x^2 + v_y^2 + v_z^2) \tag{1.1.3}$$

which combines both requirements. Let us assume that both f and g are differentiable. Taking now partial derivatives w.r.t. v_x, v_y and v_z, we obtain

$$f'(v_x)f(v_y)f(v_z) = 2v_x g'(v_x^2 + v_y^2 + v_z^2) \tag{1.1.4}$$

$$f(v_x)f'(v_y)f(v_z) = 2v_y g'(v_x^2 + v_y^2 + v_z^2) \tag{1.1.5}$$

$$f(v_x)f(v_y)f'(v_z) = 2v_z g'(v_x^2 + v_y^2 + v_z^2) \tag{1.1.6}$$

implying that

$$\frac{f'(v_x)f(v_y)f(v_z)}{v_x} = \frac{f(v_x)f'(v_y)f(v_z)}{v_y} = \frac{f(v_x)f(v_y)f'(v_z)}{v_z} = 2g'(v_x^2 + v_y^2 + v_z^2), \tag{1.1.7}$$

and in particular,

$$\frac{f'(v_x)}{v_x f(v_x)} = \frac{f'(v_y)}{v_y f(v_y)}. \tag{1.1.8}$$

Since the l.h.s. depends only on v_x and the r.h.s. depends only on v_y, the last identity can hold only if $f'(v_x)/[v_x f(v_x)] = \text{const}$. Let us denote this constant by -2α. Then, we have a simple differential equation,

$$\frac{f'(v_x)}{f(v_x)} = -2\alpha v_x, \tag{1.1.9}$$

whose solution is easily found to be

$$f(v_x) = Be^{-\alpha v_x^2}, \tag{1.1.10}$$

and similar relations hold also for v_y and v_z. For f to be a valid pdf, α must be positive and B must be the appropriate constant of normalization, which gives

$$f(v_x) = \sqrt{\frac{\alpha}{\pi}} e^{-\alpha v_x^2} \tag{1.1.11}$$

and the same applies to v_y and v_z. Thus, we finally obtain

$$f(v_x, v_y, v_z) = \left(\frac{\alpha}{\pi}\right)^{3/2} e^{-\alpha(v_x^2+v_y^2+v_z^2)} \tag{1.1.12}$$

namely, a Gaussian pdf of three zero–mean independent variables with the same variance, and it only remains to determine one constant, α, which is related to this variance. We would like now to express α in terms of some physical parameters of the system.

To this end, we adopt the following consideration. Assume, without essential loss of generality, that the container is a box of sizes $L_x \times L_y \times L_z$, whose walls are parallel to the axes of the coordinate system. Consider a molecule with velocity $\vec{v} = v_x\hat{x} + v_y\hat{y} + v_z\hat{z}$ hitting a wall parallel to the YZ plane from the inner side (left side) of the box. The molecule bounces elastically with a new velocity vector $\vec{v'} = -v_x\hat{x} + v_y\hat{y} + v_z\hat{z}$, and so, the change in momentum, which is also the impulse that the molecule exerts on the wall, is $\Delta p = 2mv_x$, where m is the mass of the molecule. For a molecule of velocity v_x in the x–direction to hit the wall within time duration τ, its initial distance from the wall must not exceed $v_x\tau$ in the x–direction. Thus, the total average impulse contributed by a molecule with an x–component velocity ranging between v_x and $v_x + \mathrm{d}v_x$, is given by

$$2mv_x \cdot \frac{v_x\tau}{L_x} \cdot \sqrt{\frac{\alpha}{\pi}} e^{-\alpha v_x^2} \mathrm{d}v_x.$$

Consequently, the average impulse of a single molecule, exerted within time τ, is the integral, given by

$$\frac{2m\tau}{L_x} \cdot \sqrt{\frac{\alpha}{\pi}} \int_0^\infty v_x^2 e^{-\alpha v_x^2} \mathrm{d}v_x = \frac{m\tau}{2\alpha L_x}.$$

It follows that the average[4] force exerted on the YZ wall is obtained by dividing the last expression by τ, namely, it is $m/(2\alpha L_x)$, and then the average pressure contributed by a single molecule is $m/(2\alpha L_x L_y L_z)$. Therefore, the total pressure contributed by all N molecules is

$$P = \frac{mN}{2\alpha L_x L_y L_z} = \frac{mN}{2\alpha V} = \frac{\rho m}{2\alpha} \tag{1.1.13}$$

---

[4] Average – over time as well.

and so, we can determine α in terms of the physical quantities P and ρ and m:

$$\alpha = \frac{\rho m}{2P}. \tag{1.1.14}$$

From the equation of state of the ideal gas[5]

$$P = \rho kT \tag{1.1.15}$$

where k is Boltzmann's constant ($= 1.381 \times 10^{-23}$ Joules/degree) and T is absolute temperature. Thus, an alternative expression for α is:

$$\alpha = \frac{m}{2kT}. \tag{1.1.16}$$

On substituting this into the general Gaussian form of the pdf, we finally obtain

$$f(\vec{v}) = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left[-\frac{m}{2kT}(v_x^2 + v_y^2 + v_z^2)\right] = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left[-\frac{\epsilon}{kT}\right], \tag{1.1.17}$$

where ϵ is the (kinetic) energy of the molecule. This form of a pdf, that is proportional to $e^{-\epsilon/(kT)}$, where ϵ is the energy, is not a coincidence. We shall see it again and again later on, and in much greater generality, as a fact that stems from much deeper and more fundamental principles. It is called the *Boltzmann–Gibbs distribution*.[6]

Having derived the pdf of $\vec{v}$, we can now calculate a few moments. Throughout this book, we will denote the expectation operator by $\langle \cdot \rangle$, which is the customary notation used by physicists. Since

$$\langle v_x^2 \rangle = \langle v_y^2 \rangle = \langle v_z^2 \rangle = \frac{kT}{m} \tag{1.1.18}$$

we readily have

$$\langle \|\vec{v}\|^2 \rangle = \langle v_x^2 + v_y^2 + v_z^2 \rangle = \frac{3kT}{m}, \tag{1.1.19}$$

---

[5] Consider this as an experimental fact.

[6] Another example of the Boltzmann form $e^{-\epsilon/(kT)}$ is the barometric formula: considering gravity, the pressure increment $\mathrm{d}P$ between height h and height $h + \mathrm{d}h$ in an ideal–gas atmosphere, must be equal to $-\mu g \mathrm{d}h$, which is the pressure contributed by a layer of thickness $\mathrm{d}h$, where μ is the mass density. Thus, $\mathrm{d}P/\mathrm{d}h = -\mu g$. But by the equation of state, $\mu = Nm/V = mP/kT$, which gives the simple differential equation $\mathrm{d}P/\mathrm{d}h = -Pmg/kT$, whose solution is $P = P_0 e^{-mgh/kT}$, and so, $\mu = -P'/g = \rho_0 e^{-mgh/kT}$, which is proportional to the probability density. Then, here we have $\epsilon = mgh$, the gravitational potential energy of one particle.

and so the root mean square (RMS) speed is given by

$$s_{\text{RMS}} \stackrel{\Delta}{=} \sqrt{\langle \|\vec{v}\|^2 \rangle} = \sqrt{\frac{3kT}{m}}. \tag{1.1.20}$$

Other related statistical quantities, that can be derived from $f(\vec{v})$, are the average speed $\langle s \rangle$ and the most likely speed. Like s_{RMS}, they are also proportional to $\sqrt{kT/m}$ but with different constants of proportionality (see Exercise 1.1 below). The average kinetic energy per molecule is

$$\langle \epsilon \rangle = \left\langle \frac{1}{2} m \|\vec{v}\|^2 \right\rangle = \frac{3kT}{2}, \tag{1.1.21}$$

independent of m. This relation gives the basic significance to the notion of temperature: at least in the case of the ideal gas, temperature is simply a quantity that is directly proportional to the average kinetic energy of each particle. In other words, temperature and kinetic energy are almost synonyms in this case. In the sequel, we will see a more general definition of temperature. The factor of 3 at the numerator is due to the fact that space has three dimensions, and so, each molecule has 3 degrees of freedom. Every degree of freedom contributes an amount of energy given by $kT/2$. This will turn out later to be a special case of a more general principle called the *equipartition of energy*.

The pdf of the speed, $s = \|\vec{v}\|$, can be derived from the pdf of the velocity $\vec{v}$ using the obvious consideration that all vectors $\vec{v}$ of the same norm correspond to the same speed. Thus, the pdf of s is simply the pdf of $\vec{v}$ (which depends solely on $\|\vec{v}\| = s$) multiplied by the surface area of a three–dimensional sphere of radius s, which is $4\pi s^2$, i.e.,

$$f(s) = 4\pi s^2 \left(\frac{m}{2\pi kT}\right)^{3/2} e^{-ms^2/(2kT)} = \sqrt{\frac{2}{\pi}\left(\frac{m}{kT}\right)^3} \cdot s^2 e^{-ms^2/(2kT)} \tag{1.1.22}$$

This is called the *Maxwell distribution* and it is depicted in Fig. 1.1 for various values of the parameter kT/m. To obtain the pdf of the energy ϵ, we should change variables according to $s = \sqrt{2\epsilon/m}$ and $\mathrm{d}s = \mathrm{d}\epsilon/\sqrt{2m\epsilon}$. The result is

$$f(\epsilon) = \frac{2\sqrt{\epsilon}}{\sqrt{\pi}(kT)^{3/2}} \cdot e^{-\epsilon/(kT)}. \tag{1.1.23}$$

Exercise 1.1 Use the above to calculate: (i) the average speed $\langle s \rangle$, (ii) the most likely speed, $\text{argmax}_s f(s)$, and (iii) the most likely energy $\text{argmax}_\epsilon f(\epsilon)$.

An interesting relation, that will be referred to later on, links between the average energy per particle $\bar{\epsilon} = \langle \epsilon \rangle$, the density ρ, and the pressure P, or equivalently, the total energy $E = N\bar{\epsilon}$, the volume V and P:

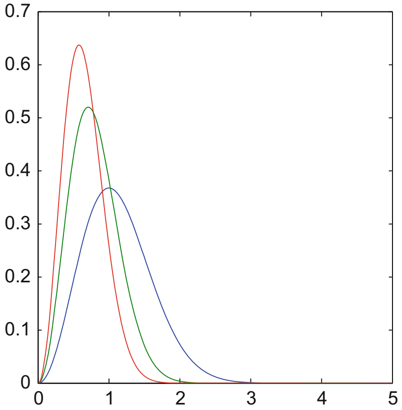


Fig. 1.1 Demonstration of the Maxwell distribution for various values of the parameter kT/m. The *red curve* (tall and narrow) corresponds to the smallest value and the *blue curve* (short and wide) – to the highest value

$$P = \rho kT = \frac{2\rho}{3} \cdot \frac{3kT}{2} = \frac{2\rho}{3} \cdot \bar{\epsilon}, \tag{1.1.24}$$

which after multiplying by V becomes

$$PV = \frac{2E}{3}. \tag{1.1.25}$$

It is interesting to note that this relation can be obtained directly from the analysis of the impulse exerted by the particles on the walls, similarly as in the earlier derivation of the parameter α, and without recourse to the equation of state (see, for example, [1, Sect. 20–4, pp. 353–355]). This is because the parameter α of the Gaussian pdf of each component of $\vec{v}$ has the obvious meaning of $1/(2\sigma_v^2)$, where σ_v^2 is the common variance of each component of $\vec{v}$. Thus, $\sigma_v^2 = 1/(2\alpha)$ and so, $\langle \|\vec{v}\|^2 \rangle = 3\sigma_v^2 = 3/(2\alpha)$, which in turn implies that

$$\bar{\epsilon} = \left\langle \frac{m}{2} \|\vec{v}\|^2 \right\rangle = \frac{3m}{4\alpha} = \frac{3m}{4\rho m/(2P)} = \frac{3P}{2\rho}, \tag{1.1.26}$$

in equivalence to the above.

1.2 Collisions

We now take a closer look into the issue of collisions. We first define the concept of *collision cross–section*, which we denote by σ. Referring to Fig. 1.2, consider a situation, where two hard spheres, labeled A and B, with diameters $2a$ and $2b$, respectively, are approaching each other, and let c be the projection of the distance between their centers in the direction perpendicular to the direction of their relative

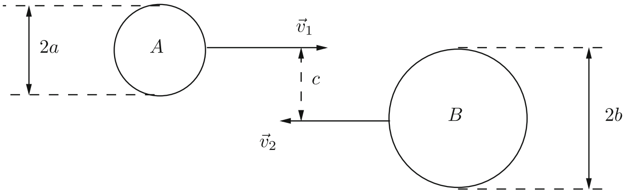


Fig. 1.2 Hard sphere collision

motion, $\vec{v}_1 - \vec{v}_2$. Clearly, collision will occur if and only if $c < a + b$. In other words, the two spheres would collide only if the center of B lies inside a volume whose cross sectional area is $\sigma = \pi(a+b)^2$, or for identical spheres, $\sigma = 4\pi a^2$. Let the colliding particles have relative velocity $\Delta\vec{v} = \vec{v}_1 - \vec{v}_2$. Passing to the coordinate system of the center of mass of the two particles, this is equivalent to the motion of one particle with the reduced mass $\mu = m_1 m_2/(m_1 + m_2)$, and so, in the case of identical particles, $\mu = m/2$. The average relative speed is easily calculated from the Maxwell distribution, but with m being replaced by $\mu = m/2$, i.e.,

$$\langle\|\Delta\vec{v}\|\rangle = 4\pi\left(\frac{m}{4\pi kT}\right)^{3/2}\int_0^\infty (\Delta v)^3 e^{-m(\Delta v)^2/(4kT)}\mathrm{d}(\Delta v) = 4\cdot\sqrt{\frac{kT}{\pi m}} = \sqrt{2}\langle s\rangle. \tag{1.2.1}$$

The total number of particles per unit volume that collide with a particular particle within time τ is

$$N_{\text{col}}(\tau) = \rho\sigma\langle\|\Delta\vec{v}\|\rangle\tau = 4\rho\sigma\tau\sqrt{\frac{kT}{\pi m}} \tag{1.2.2}$$

and so, the collision rate of each particle is

$$\nu = 4\rho\sigma\sqrt{\frac{kT}{\pi m}}. \tag{1.2.3}$$

The mean distance between collisions (a.k.a. the mean free path) is therefore

$$\lambda = \frac{\langle\|\vec{v}\|\rangle}{\nu} = \frac{1}{\sqrt{2}\rho\sigma} = \frac{kT}{\sqrt{2}P\sigma}. \tag{1.2.4}$$

What is the probability distribution of the random distance L between two consecutive collisions of a given particle? In particular, what is $p(l) \stackrel{\Delta}{=} \Pr\{L \geq l\}$? Let us assume that the collision process is memoryless in the sense that the event of not colliding before distance $l_1 + l_2$ is the intersection of two *independent* events, the first one being the event of not colliding before distance l_1, and the second one being the event of not colliding before the additional distance l_2. That is

$$p(l_1 + l_2) = p(l_1)p(l_2). \tag{1.2.5}$$

We argue that under this assumption, $p(l)$ must be exponential in l. This follows from the following consideration.[7] Taking partial derivatives of both sides w.r.t. both l_1 and l_2, we get

$$p'(l_1 + l_2) = p'(l_1)p(l_2) = p(l_1)p'(l_2). \tag{1.2.6}$$

Thus,

$$\frac{p'(l_1)}{p(l_1)} = \frac{p'(l_2)}{p(l_2)} \tag{1.2.7}$$

for all non-negative l_1 and l_2. Thus, $p'(l)/p(l)$ must be a constant, which we shall denote by $-a$. This trivial differential equation has only one solution which obeys the obvious initial condition $p(0) = 1$:

$$p(l) = e^{-al} \quad l \geq 0 \tag{1.2.8}$$

so it only remains to determine the parameter a, which must be positive since the function $p(l)$ must be monotonically non–increasing by definition. This can easily be found by using the fact that $\langle L \rangle = 1/a = \lambda$, and so,

$$p(l) = e^{-l/\lambda} = \exp\left(-\frac{\sqrt{2}P\sigma l}{kT}\right). \tag{1.2.9}$$

1.3 Dynamical Aspects

The discussion thus far focused on the static (equilibrium) behavior of the ideal gas. In this subsection, we will briefly touch upon dynamical issues pertaining to non–equilibrium situations. These issues will be further developed in Chap. 7, and with much greater generality.

Consider two adjacent containers separated by a wall. Both have the same volume V of the same ideal gas at the same temperature T, but with different densities ρ_1 and ρ_2, and hence different pressures P_1 and P_2. Let us assume that $P_1 > P_2$. At time $t = 0$, a small hole is created in the separating wall. The area of this hole is A (see Fig. 1.3).

If the mean free distances λ_1 and λ_2 are relatively large compared to the dimensions of the hole, it is safe to assume that every molecule that reaches the hole, passes

---

[7] Similar idea to the one of the earlier derivation of the Gaussian pdf of the ideal gas.

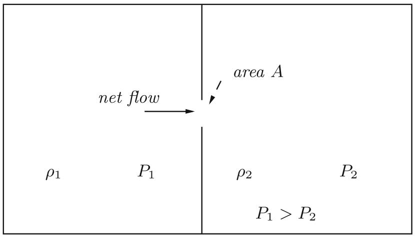


Fig. 1.3 Gas leakage through a small hole

through it. The mean number of molecules that pass from left to right within time τ is given by[8]

$$N_{\rightarrow} = \rho_1 V \cdot \int_0^\infty \mathrm{d}v_x \sqrt{\frac{\alpha}{\pi}} e^{-\alpha v_x^2} \cdot \frac{v_x \tau A}{V} = \frac{\rho_1 \tau A}{2\sqrt{\pi\alpha}} \tag{1.3.1}$$

and so the number of particles per second, flowing from left to right is

$$\frac{\mathrm{d}N_{\rightarrow}}{\mathrm{d}t} = \frac{\rho_1 A}{2\sqrt{\pi\alpha}}. \tag{1.3.2}$$

Similarly, in the opposite direction, we have

$$\frac{\mathrm{d}N_{\leftarrow}}{\mathrm{d}t} = \frac{\rho_2 A}{2\sqrt{\pi\alpha}}, \tag{1.3.3}$$

and so, the net left–to–right current is

$$I \stackrel{\Delta}{=} \frac{\mathrm{d}N}{\mathrm{d}t} = \frac{(\rho_1 - \rho_2)A}{2\sqrt{\pi\alpha}} = (\rho_1 - \rho_2)A\sqrt{\frac{kT}{2\pi m}}. \tag{1.3.4}$$

An important point here is that the current is proportional to the difference between densities $(\rho_1 - \rho_2)$, and considering the equation of state of the ideal gas, it is therefore also proportional to the pressure difference, $(P_1 - P_2)$. This rings the bell of the well known analogous fact that the electric current is proportional to the voltage, which in turn is the difference between the electric potentials at two points. Considering the fact that $\rho \stackrel{\Delta}{=} (\rho_1 + \rho_2)/2$ is constant, we obtain a simple differential equation

$$\frac{\mathrm{d}\rho_1}{\mathrm{d}t} = (\rho_2 - \rho_1)\frac{A}{V}\sqrt{\frac{kT}{2\pi m}} \stackrel{\Delta}{=} C(\rho_2 - \rho_1) = 2C(\rho - \rho_1) \tag{1.3.5}$$

---

[8] Note that for $v_y = v_z = 0$, the factor $v_x \tau A/V$, in the forthcoming equation, is clearly the relative volume (and hence the probability) of being in the 'box' in which a particle must be found in order to pass the hole within τ seconds. When v_y and v_z are non–zero, instead of a rectangular box, this region becomes a parallelepiped, but the relative volume remains $v_x \tau A/V$ independently of v_y and v_z.

whose solution is

$$\rho_1(t) = \rho + [\rho_1(0) - \rho]e^{-2Ct} \tag{1.3.6}$$

which means that equilibrium is approached exponentially fast with time constant

$$\tau = \frac{1}{2C} = \frac{V}{2A}\sqrt{\frac{2\pi m}{kT}}. \tag{1.3.7}$$

Imagine now a situation, where there is a long pipe aligned along the x–direction. The pipe is divided into a chain of cells in a linear fashion, and in the wall between each two consecutive cells there is a hole of area A. The length of each cell (i.e., the distance between consecutive walls) is the mean free distance λ, so that collisions within each cell can be neglected. Assume further that λ is so small that the density of each cell at time t can be approximated using a continuous function $\rho(x,t)$. Let x_0 be the location of one of the walls. Then, according to the above derivation, the current at $x = x_0$ is

$$\begin{aligned} I(x_0) &= \left[\rho\left(x_0 - \frac{\lambda}{2}, t\right) - \rho\left(x_0 + \frac{\lambda}{2}, t\right)\right] A\sqrt{\frac{kT}{2\pi m}} \\ &\approx -A\lambda\sqrt{\frac{kT}{2\pi m}} \cdot \left.\frac{\partial\rho(x,t)}{\partial x}\right|_{x=x_0}. \end{aligned} \tag{1.3.8}$$

Thus, the current is proportional to the negative gradient of the density. This is quite a fundamental result which holds with much greater generality. In the more general context, it is known as *Fick's law*.

Consider next two close points x_0 and $x_0 + \mathrm{d}x$, with possibly different current densities (i.e., currents per unit area) $J(x_0)$ and $J(x_0 + \Delta x)$. The difference $J(x_0) - J(x_0 + \Delta x)$ is the rate at which matter accumulates along the interval $[x_0, x_0 + \Delta x]$ per unit area in the perpendicular plane. Within Δt seconds, the number of particles per unit area within this interval has grown by $[J(x_0) - J(x_0 + \Delta x)]\Delta t$. But this amount is also $[\rho(x_0, t + \Delta t) - \rho(x_0, t)]\Delta x$, Taking the appropriate limits, we get

$$\frac{\partial J(x)}{\partial x} = -\frac{\partial\rho(x,t)}{\partial t}, \tag{1.3.9}$$

which is a one–dimensional version of the so called *equation of continuity*. Differentiating now Eq. (1.3.8) w.r.t. x and comparing with (1.3.9), we obtain the *diffusion equation* (in one dimension):

$$\frac{\partial\rho(x,t)}{\partial t} = D\frac{\partial^2\rho(x,t)}{\partial x^2} \tag{1.3.10}$$

where the constant D, in this case,

$$D = \frac{A\lambda}{S} \cdot \sqrt{\frac{kT}{2\pi m}}, \tag{1.3.11}$$

which is called the *diffusion coefficient*. Here S is the cross–section area.

This is, of course, merely a toy model – it is a caricature of a real diffusion process, but it captures its essence. Diffusion processes are central in irreversible statistical mechanics, since the solution to the diffusion equation is sensitive to the sign of time. This is different from the Newtonian equations of frictionless motion, which have a time reversal symmetry and hence are reversible. We will touch upon these issues in Chap. 7.

The equation of continuity, Fick's law, the diffusion equation and its extension, the Fokker–Planck equation (which will also be discussed), are all very central in physics in general and in semiconductor physics, in particular, as they describe processes of propagation of concentrations of electrons and holes in semiconductor materials. Another branch of physics where these equations play an important role is fluid mechanics.

1.4 Suggestions for Supplementary Reading

The following books are recommended for further reading and for related material. Beck [2, Sect. 2.2] derives the pdf of the particle momenta in a manner somewhat different than here. Other parts of this section are quite similar to those of [2, Chap. 2]. A much more detailed exposition of the kinetic theory of gases appears also in many other textbooks, including: Huang [3, Chap. 4], Kardar [4, Chap. 3], Kittel [5, Part I, Chap. 13], Mandl [6, Chap. 7], Reif [7, Chap. 9], and Tolman [8, Chap. IV], to name a few.

References

1. F.W. Sears, M.W. Zemansky, H.D. Young, *University Physics* (Addison-Wesley, Reading, 1976)
2. A.H.W. Beck, *Statistical Mechanics, Fluctuations and Noise* (Edward Arnold Publishers, London, 1976)
3. K. Huang, *Statistical Mechanics*, 2nd edn. (Wiley, New York, 1987)
4. M. Kardar, *Statistical Physics of Particles* (Cambridge University Press, Cambridge, 2007)
5. C. Kittel, *Elementary Statistical Physics* (Wiley, New York, 1958)
6. F. Mandl, *Statistical Physics* (Wiley, Chichester, 1971)
7. F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1965)
8. R.C. Tolman, *The Principles of Statistical Mechanics* (Dover Publications, New York, 1979)

Chapter 2
Elementary Statistical Physics

In this chapter, we provide the formalism and the elementary background in statistical physics. We first define the basic postulates of statistical mechanics, and then define various ensembles. Finally, we shall derive some of the thermodynamic potentials and their properties, as well as the relationships among them. The important laws of thermodynamics will also be pointed out. The contents of this chapter has a considerable overlap with Chap. 2 of [1] (though there are also considerable differences). It is provided in this book too, mostly for the sake of completeness.

2.1 Basic Postulates

As explained in the Introduction, statistical physics is about a probabilistic approach to systems of many particles. While our discussion here will no longer be specific to the ideal gas as before, we will nonetheless start again with this example in mind, just for the sake of concreteness, Consider then a system with a very large number N of mobile particles, which are free to move in a given volume. The *microscopic state* (or *microstate*, for short) of the system, at each time instant t, consists, in this example, of the position vector $\vec{r}_i(t)$ and the momentum vector $\vec{p}_i(t)$ of each and every particle, $1 \leq i \leq N$. Since each one of these is a vector of three components, the microstate is then given by a $(6N)$–dimensional vector $\vec{\boldsymbol{x}}(t) = \{(\vec{r}_i(t), \vec{p}_i(t)) : i = 1, 2, \ldots, N\}$, whose trajectory along the time axis, in the *phase space* $\mathbb{R}^{6N}$, is called the *phase trajectory*.

Let us assume that the system is closed, i.e., *isolated* from its environment, in the sense that no energy flows inside or out. Imagine that the phase space $\mathbb{R}^{6N}$ is partitioned into very small hypercubes (or cells) $\Delta\vec{p} \times \Delta\vec{r}$. One of the basic postulates of statistical mechanics is the following: in the long run, the relative amount of time at which $\vec{\boldsymbol{x}}(t)$ spends within each such cell, converges to a certain number between 0 and 1, which can be given the meaning of the *probability* of this cell.

N. Merhav, *Statistical Physics for Electrical Engineering*,
DOI 10.1007/978-3-319-62063-3_2

Thus, there is an underlying assumption of equivalence between temporal averages and ensemble averages, namely, this is the postulate of *ergodicity*. Considerable efforts were dedicated to the proof of the ergodic hypothesis at least in some cases. As reasonable and natural as it may seem, the ergodic hypothesis should not be taken for granted. It does not hold for every system but only if no other conservation law holds. For example, a single molecule ($N = 1$) of a gas in a box is non–ergodic. The reason is simple: assuming elastic collisions with the walls, the kinetic energy of the molecule is conserved, and hence also the speed s, rather than sampling the Maxwell distribution over time.

What are then the probabilities of the above–mentioned phase–space cells? We would like to derive these probabilities from first principles, based on as few as possible basic postulates. Our second postulate is that for an isolated system (i.e., whose energy is fixed) all microscopic states $\{\vec{\boldsymbol{x}}(t)\}$ are equiprobable. The rationale behind this postulate is twofold:

- In the absence of additional information, there is no apparent reason that certain regions in phase space would have preference relative to any others.
- This postulate is in harmony with a basic result in kinetic theory of gases – *the Liouville theorem* (see e.g., [2]), which we will not touch upon in this book, but in a nutshell, it asserts that the phase trajectories must lie along hyper-surfaces of constant probability density.[1]

2.2 Statistical Ensembles

2.2.1 The Microcanonical Ensemble

Before we proceed, let us slightly broaden the scope of our discussion. In a more general context, associated with our N–particle physical system, is a certain instantaneous microstate, generically denoted by $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$, where each x_i, $1 \leq i \leq N$, may itself be a vector of several physical quantities associated with particle number i, e.g., its position, momentum, angular momentum, magnetic moment, spin, and so on, depending on the type and the nature of the physical system. According to the physical model of the given system, there is a certain energy function, a.k.a. *Hamiltonian*, that assigns to every $\boldsymbol{x}$ a certain energy $\mathcal{E}(\boldsymbol{x})$.[2] Now, let us denote by $A(E)$ the volume of the shell of energy about E. This means

---

[1] This is a result of the energy conservation law along with the fact that probability mass behaves like an incompressible fluid in the sense that whatever mass that flows into a certain region from some direction must be equal to the outgoing flow from some other direction. This is reflected in the equation of continuity, which was demonstrated earlier.

[2] For example, in the case of an *ideal gas*, $\mathcal{E}(\boldsymbol{x}) = \sum_{i=1}^{N} \|\vec{p}_i\|^2/(2m)$, where m is the mass of each molecule, namely, it accounts for the contribution of the kinetic energies only. In more complicated situations, there might be additional contributions of potential energy, which depend on the positions.

$$A(E) = \text{Vol}\{\boldsymbol{x}: \ E \le \mathcal{E}(\boldsymbol{x}) \le E + \Delta E\} = \int_{\{\boldsymbol{x}: \ E \le \mathcal{E}(\boldsymbol{x}) \le E+\Delta E\}} \mathrm{d}\boldsymbol{x}, \tag{2.2.1}$$

where ΔE is a very small (but fixed) energy increment, which is immaterial when N is large. Then, our above postulate concerning the ensemble of an isolated system, which is called the *microcanonical ensemble*, is that the probability density $P(\boldsymbol{x})$ is given by

$$P(\boldsymbol{x}) = \begin{cases} \frac{1}{A(E)} & E \le \mathcal{E}(\boldsymbol{x}) \le E + \Delta E \\ 0 & \text{elsewhere} \end{cases} \tag{2.2.2}$$

In the discrete case, things are simpler: here, $A(E)$ is the number of microstates with $\mathcal{E}(\boldsymbol{x}) = E$ (exactly) and $P(\boldsymbol{x})$ is the uniform probability mass function over this set of states.

Returning to the general case, we next define the notion of the *density of states*, $\omega(E)$, which is intimately related to $A(E)$. Basically, in simple cases, $\omega(E)$ is defined such that $\omega(E)\Delta E = A(E)$ where ΔE is very small, but there might be a few minor corrections, depending on the concrete system being addressed. More generally, we define the density of states such that $\omega(E)\Delta E = \Omega(E)$, where $\Omega(E)$ will be the relevant (possibly corrected) function. The first correction has to do with the fact that $A(E)$ is, in general, not dimensionless: in the above example of a gas, it has the physical units of [length × momentum]^{3N} = [J · s]^{3N}, but we must eliminate these physical units because we will have to apply on it non–linear functions like the logarithmic function. To this end, we normalize the volume $A(E)$ by an elementary reference volume. In the gas example, this reference volume is taken to be h^{3N}, where h is *Planck's constant* ($h \approx 6.62 \times 10^{-34}$ J·s). Informally, the intuition comes from the fact that h is our best available "resolution" in the plane spanned by each component of $\vec{r}_i$ and the corresponding component of $\vec{p}_i$, owing to the *uncertainty principle* in quantum mechanics, which tells that the product of the standard deviations $\Delta p_a \cdot \Delta r_a$ of each component a ($a = x, y, z$) is lower bounded by $\hbar/2$, where $\hbar = h/(2\pi)$. More formally, this reference volume is obtained in a natural manner from quantum statistical mechanics: by changing the integration variable $\vec{p}$ to $\vec{k}$ using the relation $\vec{p} = \hbar\vec{k}$, where $\vec{k}$ is the wave vector. This is a well–known relation (one of the de Broglie relations) pertaining to particle–wave duality. The second correction that is needed to pass from $A(E)$ to $\Omega(E)$ is applicable when the particles are indistinguishable[3]: In these cases, we do not consider permutations between particles in a given configuration as distinct microstates. Thus, we have to divide also by N! Taking into account both corrections, we find that in the example of the ideal gas,

---

[3] In the example of the ideal gas, since the particles are mobile and since they have no colors and no identity certificates, there is no distinction between a state where particle no. 15 has position $\vec{r}$ and momentum $\vec{p}$ while particle no. 437 has position $\vec{r}'$ and momentum $\vec{p}'$ and a state where these two particles are swapped.

$$\Omega(E) = \frac{A(E)}{N!h^{3N}}. \tag{2.2.3}$$

Once again, it should be understood that both of these corrections are optional and their applicability depends on the system in question: the first correction is applicable only if $A(E)$ has physical units and the second correction is applicable only if the particles are indistinguishable. For example, if $\boldsymbol{x}$ is discrete, in which case the integral defining $A(E)$ is replaced by a sum (that counts $\boldsymbol{x}$'s with $\mathcal{E}(\boldsymbol{x}) = E$), and the particles are distinguishable, then no corrections are needed at all, i.e.,

$$\Omega(E) = A(E). \tag{2.2.4}$$

Now, the *entropy* is defined as

$$S(E) = k \ln \Omega(E), \tag{2.2.5}$$

where k is *Boltzmann's constant*. We will see later what is the relationship between $S(E)$ and the classical thermodynamic entropy, due to Clausius (1850), as well as the information–theoretic entropy, due to Shannon (1948). As will turn out, all three are equivalent to one another. Here, a comment on the notation is in order: the entropy S may depend on additional quantities, other than the energy E, like the volume V and the number of particles N. When this dependence will be relevant and important, we will use the more complete form of notation $S(E, V, N)$. If only the dependence on E is relevant in a certain context, we use the simpler notation $S(E)$.

To get some insight into the behavior of the entropy, it should be noted that normally, $\Omega(E)$ (and hence also $\omega(E)$) behaves as an exponential function of N (at least asymptotically), and so, $S(E)$ is roughly linear in N. For example, if $\mathcal{E}(\boldsymbol{x}) = \sum_{i=1}^{N} \frac{\|\vec{p}_i\|^2}{2m}$, then $\Omega(E)$ is the volume of a thin shell about the surface of a $(3N)$–dimensional sphere with radius $\sqrt{2mE}$, divided by $N!h^{3N}$, which is proportional to $(2mE)^{3N/2}V^N/N!h^{3N}$, where V is the volume. The quantity $\omega(E)$ is then associated with the surface area of this $(3N)$–dimensional sphere. Specifically (ignoring the contribution of the factor ΔE), we get

$$\begin{aligned} S(E, V, N) &= k \ln \left[\left(\frac{4\pi m E}{3N} \right)^{3N/2} \cdot \frac{V^N}{N!h^{3N}} \right] + \frac{3}{2}Nk \\ &\approx Nk \ln \left[\left(\frac{4\pi m E}{3N} \right)^{3/2} \cdot \frac{V}{Nh^3} \right] + \frac{5}{2}Nk. \end{aligned} \tag{2.2.6}$$

Assuming that E and V are both proportional to N ($E = N\epsilon$ and $V = N/\rho$), it is readily seen that $S(E, V, N)$ is also proportional to N. A physical quantity that has a linear dependence on the size of the system N, is called an *extensive quantity*. Energy, volume and entropy are then extensive quantities. Other quantities, which are not extensive, i.e., independent of the system size, like temperature and pressure, are called *intensive*.

It is interesting to point out that from the function $S(E, V, N)$, one can obtain the entire information about the relevant macroscopic physical quantities of the system, e.g., temperature, pressure, and so on. Specifically, the *temperature* T of the system is defined according to:

$$\frac{1}{T} = \left[\frac{\partial S(E, V, N)}{\partial E}\right]_{V,N} \tag{2.2.7}$$

where $[\cdot]_{V,N}$ emphasizes that the derivative is taken while keeping V and N constant. One may wonder, at this point, what is the justification for *defining* temperature this way. We will get back to this point a bit later, but for now, we can easily see that this is indeed true at least for the ideal gas, as by taking the derivative of (2.2.6) w.r.t. E, we get

$$\frac{\partial S(E, V, N)}{\partial E} = \frac{3Nk}{2E} = \frac{1}{T}, \tag{2.2.8}$$

where the second equality has been shown already in Chap. 1.

Intuitively, in most situations, we expect that $S(E)$ would be an increasing function of E for fixed V and N (although this is not strictly always the case), which means $T \geq 0$. But T is also expected to increase with E (or equivalently, E is increasing with T, as otherwise, the heat capacity $\mathrm{d}E/\mathrm{d}T < 0$). Thus, $1/T$ should decrease with E, which means that the increase of S in E slows down as E grows. In other words, we expect $S(E)$ to be a concave function of E. In the above example, indeed, $S(E)$ is logarithmic and $E = 3NkT/2$, as we have seen.

How can we be convinced, in mathematical terms, that under certain regularity conditions, $S(E)$ is a concave function in E? The answer may be given by a simple superadditivity argument: As both E and S are extensive quantities, let us define $E = N\epsilon$ and for a given density ρ,

$$s(\epsilon) = \lim_{N\to\infty} \frac{S(N\epsilon)}{N}, \tag{2.2.9}$$

i.e., the per–particle entropy as a function of the per–particle energy, where we assume that the limit exists. Consider the case where the Hamiltonian is additive, i.e.,

$$\mathcal{E}(\boldsymbol{x}) = \sum_{i=1}^{N} \mathcal{E}(x_i) \tag{2.2.10}$$

just like in the above example where $\mathcal{E}(\boldsymbol{x}) = \sum_{i=1}^{N} \frac{\|\vec{p}_i\|^2}{2m}$. Then, the inequality

$$\Omega(N_1\epsilon_1 + N_2\epsilon_2) \geq \Omega(N_1\epsilon_1) \cdot \Omega(N_2\epsilon_2), \tag{2.2.11}$$

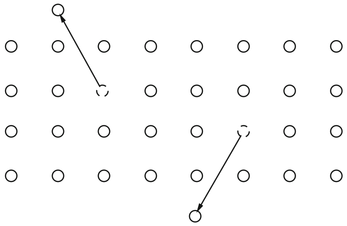


Fig. 2.1 Schottky defects in a crystal lattice

expresses the simple fact that if our system is partitioned into two parts,[4] one with N_1 particles, and the other with $N_2 = N - N_1$ particles, then every combination of individual microstates with energies $N_1\epsilon_1$ and $N_2\epsilon_2$ corresponds to a combined microstate with a total energy of $N_1\epsilon_1 + N_2\epsilon_2$ (but there are more ways to split this total energy between the two parts). Thus,

$$\begin{aligned}\frac{k\ln\Omega(N_1\epsilon_1 + N_2\epsilon_2)}{N_1 + N_2} &\geq \frac{k\ln\Omega(N_1\epsilon_1)}{N_1 + N_2} + \frac{k\ln\Omega(N_2\epsilon_2)}{N_1 + N_2}\\ &= \frac{N_1}{N_1 + N_2}\cdot\frac{k\ln\Omega(N_1\epsilon_1)}{N_1} +\\ &\quad \frac{N_2}{N_1 + N_2}\cdot\frac{k\ln\Omega(N_2\epsilon_2)}{N_2}. \end{aligned} \tag{2.2.12}$$

and so, by taking N_1 and N_2 to ∞, with $N_1/(N_1 + N_2) \to \lambda \in (0, 1)$, we get:

$$s(\lambda\epsilon_1 + (1-\lambda)\epsilon_2) \geq \lambda s(\epsilon_1) + (1-\lambda)s(\epsilon_2), \tag{2.2.13}$$

which establishes the concavity of $s(\cdot)$ at least in the case of an additive Hamiltonian, which means that the entropy of mixing two systems of particles is greater than the total entropy before the mix. A similar proof can be generalized to the case where $\mathcal{E}(\boldsymbol{x})$ includes also a limited degree of interactions (short range interactions), e.g., $\mathcal{E}(\boldsymbol{x}) = \sum_{i=1}^{N}\mathcal{E}(x_i, x_{i+1})$, but this requires somewhat more caution. In general, however, concavity may no longer hold when there are long range interactions, e.g., where some terms of $\mathcal{E}(\boldsymbol{x})$ depend on a linear subset of particles.

Example 2.1 (*Schottky defects*) In a certain crystal, the atoms are located in a lattice, and at any positive temperature there may be defects, where some of the atoms are dislocated (see Fig. 2.1). Assuming that defects are sparse enough, such that around each dislocated atom all neighbors are in place, the activation energy, ϵ_0, required for dislocation is fixed. Denoting the total number of atoms by N and the number of defected ones by n, the total energy is then $E = n\epsilon_0$, and so,

---

[4] This argument works for distinguishable particles. Later on, a more general argument will be presented, that holds for indistinguishable particles too.

$$\Omega(E) = \binom{N}{n} = \frac{N!}{n!(N-n)!}, \tag{2.2.14}$$

or, equivalently,

$$\begin{aligned} S(E) = k \ln \Omega(E) &= k \ln \left[\frac{N!}{n!(N-n)!} \right] \\ &\approx k[N \ln N - n \ln n - (N-n) \ln(N-n)] \end{aligned} \tag{2.2.15}$$

where in the last passage we have used the Stirling approximation. It is important to point out that here, unlike in the example of the ideal gas, we have not divided $A(E)$ by N! The reason is that we do distinguish between two different configurations where the same number of particles were dislocated but the sites of dislocation are different. Yet, we do not distinguish between two microstates whose only difference is two (identical) particles which are not dislocated but swapped. This is the reason for the denominator $n!(N-n)!$ in the expression of $\Omega(E)$. Now,[5]

$$\frac{1}{T} = \frac{\partial S}{\partial E} = \frac{\mathrm{d}n}{\mathrm{d}E} \cdot \frac{\mathrm{d}S}{\mathrm{d}n} = \frac{1}{\epsilon_0} \cdot k \ln \frac{N-n}{n}, \tag{2.2.16}$$

which gives the number of defects as

$$n = \frac{N}{\exp(\epsilon_0/kT) + 1}. \tag{2.2.17}$$

At $T = 0$, there are no defects, but their number increases gradually with T, approximately according to $\exp(-\epsilon_0/kT)$. Note also that

$$\begin{aligned} S(E) = k \ln \binom{N}{n} &\approx kNh_2\left(\frac{n}{N}\right) \\ &= kNh_2\left(\frac{E}{N\epsilon_0}\right) = kNh_2\left(\frac{\epsilon}{\epsilon_0}\right), \end{aligned} \tag{2.2.18}$$

where

$$h_2(x) \stackrel{\Delta}{=} -x \ln x - (1-x) \ln(1-x), \qquad 0 \le x \le 1$$

is the so called binary entropy function. Note also that $s(\epsilon) = kh_2(\epsilon/\epsilon_0)$ is indeed concave in this example. □

---

[5] Here and in the sequel, the reader might wonder about the meaning of taking derivatives of, and with respect to, integer valued variables, like the number of dislocated particles, n. To this end, imagine an approximation where n is interpolated to be a continuous valued variable.

Suppose we have two systems that are initially at certain temperatures (and with corresponding energies). At a certain time instant, the two systems are brought into thermal contact with one another, but their combination remains isolated. What happens after a long time? How does the total energy E, split and what is the final temperature T of the combined system? The number of combined microstates where subsystem no. 1 has energy E_1 and subsystem no. 2 has energy $E_2 = E - E_1$ is $\Omega_1(E_1) \cdot \Omega_2(E - E_1)$. As the combined system is isolated, the probability of such a combined macrostate is proportional to $\Omega_1(E_1) \cdot \Omega_2(E - E_1)$. Keeping in mind that, normally, Ω_1 and Ω_2 are exponential in N, then for large N, this product is dominated by the value of E_1 for which it is maximum, or equivalently, the sum of logarithms, $S_1(E_1) + S_2(E - E_1)$, is maximum, i.e., it is a **maximum entropy** situation, which is **the second law of thermodynamics**, asserting that an isolated system (in this case, combined of two subsystems) achieves its maximum possible entropy in equilibrium. This maximum is normally achieved at the value of E_1 for which the derivative vanishes, i.e.,

$$S_1'(E_1) - S_2'(E - E_1) = 0 \tag{2.2.19}$$

or

$$S_1'(E_1) - S_2'(E_2) = 0 \tag{2.2.20}$$

which means

$$\frac{1}{T_1} \equiv S_1'(E_1) = S_2'(E_2) \equiv \frac{1}{T_2}. \tag{2.2.21}$$

Thus, in equilibrium, which is the maximum entropy situation, the energy splits in a way that temperatures are the same. Now, we can understand the concavity of entropy more generally: $\lambda s(\epsilon_1) + (1-\lambda)s(\epsilon_2)$ was the total entropy per particle when two subsystems (with the same entropy function) were isolated from one another, whereas $s(\lambda\epsilon_1 + (1-\lambda)\epsilon_2)$ is the equilibrium entropy per particle after we let them interact thermally.

At this point, we are ready to justify why $S'(E)$ is equal to $1/T$ in general, as was promised earlier. Although it is natural to expect that equality between $S_1'(E_1)$ and $S_2'(E_2)$, in thermal equilibrium, is related to equality between T_1 and T_2, this does not automatically mean that the derivative of each entropy is given by one over its temperature. On the face of it, for the purpose of this implication, this derivative could have been equal any one–to–one function of temperature $f(T)$. To see why $f(T) = 1/T$ indeed, imagine that we have a system with an entropy function $S_0(E)$ and that we let it interact thermally with an ideal gas whose entropy function, which we shall denote now by $S_g(E)$, is given as in Eq. (2.2.6). Now, at equilibrium $S_0'(E_0) = S_g'(E_g)$, but as we have seen already, $S_g'(E_g) = 1/T_g$, where T_g is the temperature of the ideal gas. But in thermal equilibrium the temperatures equalize, i.e., $T_g = T_0$, where T_0 is the temperature of the system of interest. It then follows eventually that $S_0'(E_0) = 1/T_0$, which now means that in equilibrium, the derivative of entropy of the system of interest is equal to the reciprocal of its temperature *in general*, and not only for the ideal gas! At this point, the fact that our system has interacted and equilibrated with an ideal gas is not important anymore and it does not limit the generality of this statement. In simple words, our system does not 'care' what kind of system it has interacted with, whether ideal gas or any other. This follows from a fundamental principle in thermodynamics, called **the zero–th law of thermodynamics**, which states that thermal equilibrium has a transitive property: If system A is in equilibrium with system B and system B is in equilibrium with system C, then A is in equilibrium with C.

So we have seen that $\partial S/\partial E = 1/T$, or equivalently, $\delta S = \delta E/T$. But in the absence of any mechanical work (V is fixed) applied to the system and any chemical energy injected into the system (N is fixed), any change in energy must be in the form of *heat*,[6] thus we denote $\delta E = \delta Q$, where Q is the heat intake. Consequently,

$$\delta S = \frac{\delta Q}{T}, \tag{2.2.22}$$

This is exactly the definition of the classical thermodynamic entropy due to Clausius. Thus, at least for the case where no mechanical work is involved, we have demonstrated the equivalence of the two notions of entropy, the statistical notion due to Boltzmann $S = k \ln \Omega$, and the thermodynamic entropy due to Clausius, $S = \int \mathrm{d}Q/T$, where the integration should be understood to be taken along a slow (quasi–static) process, where after each small increase in the heat intake, the system is allowed to equilibrate, which means that T is given enough time to adjust before more heat is further added. For a given V and N, the difference ΔS between the entropies S_A and S_B associated with two temperatures T_A and T_B (pertaining to internal energies E_A and E_B, respectively) is given by $\Delta S = \int_A^B \mathrm{d}Q/T$ along such a quasi–static process. This is a rule that defines entropy differences, but not absolute levels. A reference value is determined by the **third law of thermodynamics**, which asserts that as T tends to zero, the entropy tends to zero as well.[7]

We have seen what is the meaning of the partial derivative of $S(E, V, N)$ w.r.t. E. Is there also a simple meaning to the partial derivative w.r.t. V? Again, let us begin by examining the ideal gas. Differentiating the expression of $S(E, V, N)$ of the ideal gas w.r.t. V, we obtain

$$\frac{\partial S(E, V, N)}{\partial V} = \frac{Nk}{V} = \frac{P}{T}, \tag{2.2.23}$$

---

[6] Heat is a form of energy that is transferred neither by mechanical work nor by matter. It is the type of energy that flows spontaneously from a system/body at a higher temperature to one with a lower temperature (and this transfer is accompanied by an increase in the total entropy).

[7] In this context, it should be understood that the results we derived for the ideal gas hold only for high enough temperatures: since S was found proportional to $\ln E$ and E is proportional to T, then S is proportional to $\ln T$, but this cannot be true for small T as it contradicts (among other things) the third law.

where the second equality follows again from the equation of state. So at least for the ideal gas, this partial derivative is related to the pressure P. For similar considerations as before, the relation

$$\frac{\partial S(E, V, N)}{\partial V} = \frac{P}{T} \tag{2.2.24}$$

is true not only for the ideal gas, but in general. Consider again an isolated system that consists of two subsystems, separated by a wall (or a piston). Initially, this wall is fixed and the volumes are V_1 and V_2. At a certain moment, this wall is released and allowed to be pushed in either direction. How would the total volume $V = V_1 + V_2$ divide between the two subsystems in equilibrium? Again, the total entropy $S_1(E_1, V_1) + S_2(E - E_1, V - V_1)$ would tend to its maximum for the same reasoning as before. The maximum will be reached when the partial derivatives of this sum w.r.t. both E_1 and V_1 would vanish. The partial derivative w.r.t. E_1 has already been addressed. The partial derivative w.r.t. V_1 gives

$$\frac{P_1}{T_1} = \frac{\partial S_1(E_1, V_1)}{\partial V_1} = \frac{\partial S_2(E_2, V_2)}{\partial V_2} = \frac{P_2}{T_2} \tag{2.2.25}$$

Since $T_1 = T_2$ by the thermal equilibrium pertaining to derivatives w.r.t. energies, it follows that $P_1 = P_2$, which means mechanical equilibrium: the wall will be pushed to the point where the pressures from both sides are equal. We now have the following differential relationship:

$$\begin{aligned}\delta S &= \frac{\partial S}{\partial E}\delta E + \frac{\partial S}{\partial V}\delta V \\ &= \frac{\delta E}{T} + \frac{P\delta V}{T}\end{aligned} \tag{2.2.26}$$

or

$$\delta E = T\delta S - P\delta V = \delta Q - \delta W, \tag{2.2.27}$$

which is the **the first law of thermodynamics**, asserting that the change in the energy δE of a system with a fixed number of particles is equal to the difference between the incremental heat intake δQ and the incremental mechanical work δW carried out by the system. This is nothing but a restatement of the law of energy conservation.

Example 2.2 (*Compression of ideal gas*) Consider again an ideal gas of N particles at constant temperature T. The energy is $E = 3NkT/2$ regardless of the volume. This means that if we compress (slowly) the gas from volume V_1 to volume V_2 ($V_2 < V_1$), the energy remains the same, in spite of the fact that we injected energy by applying mechanical work

$$W = -\int_{V_1}^{V_2} P\mathrm{d}V = -NkT\int_{V_1}^{V_2}\frac{\mathrm{d}V}{V} = NkT\ln\frac{V_1}{V_2}. \tag{2.2.28}$$

What happened to that energy? The answer is that it was transformed into heat as the entropy of the system (which is proportional to $\ln V$) has changed by the amount $\Delta S = -Nk\ln(V_1/V_2)$, and so, the heat intake $\Delta Q = T\Delta S = -NkT\ln(V_1/V_2)$ exactly balances the work. □

Finally, we should consider the partial derivative of S w.r.t. N. This is given by

$$\frac{\partial S(E, V, N)}{\partial N} = -\frac{\mu}{T}, \tag{2.2.29}$$

where μ is called the *chemical potential*. If we now consider again the isolated system, which consists of two subsystems that are allowed to exchange, not only heat and volume, but also particles (of the same kind), whose total number is $N = N_1 + N_2$, then again, maximum entropy considerations would yield an additional equality between the chemical potentials, $\mu_1 = \mu_2$ (chemical equilibrium).[8] The chemical potential should be understood as a kind of a force that controls the ability to inject particles into the system. For example, if the particles are electrically charged, then the chemical potential has a simple analogy to the electrical potential. The first law is now extended to have an additional term, pertaining to an increment of chemical energy, and it now reads:

$$\delta E = T\delta S - P\delta V + \mu\delta N. \tag{2.2.30}$$

Equation (2.2.30) can be used to derive a variety of relations. For example, $\mu = (\partial E/\partial N)_{S,V}$, $T = (\partial H/\partial S)_N$, where $H = E + PV$ is called the enthalpy, $P = \mu(\partial N/\partial V)_{E,S}$, and so on.

2.2.2 The Canonical Ensemble

So far we have assumed that our system is isolated, and therefore has a strictly fixed energy E. Let us now relax this assumption and assume instead that our system is free to exchange energy with its very large environment (heat bath) and that the total energy of the system plus heat bath, E_0, is by far larger than the typical energy of the system. The combined system, composed of our original system plus the heat bath, is now an isolated system at temperature T.

---

[8] Equality of chemical potentials is, in fact, the general principle of chemical equilibrium, and not equality of concentrations or densities. In Sect. 1.3, we saw equality of densities, because in the case of the ideal gas, the chemical potential is a function of the density, so equality of chemical potentials happens to be equivalent to equality of densities in this case.

Similarly as before, since the combined system is isolated, it is governed by the microcanonical ensemble. The only difference is that now we assume that one of the systems (the heat bath) is very large compared to the other (our test system). This means that if our small system is in microstate $\boldsymbol{x}$ (for whatever definition of the microstate vector) with energy $\mathcal{E}(\boldsymbol{x})$, then the heat bath must have energy $E_0 - \mathcal{E}(\boldsymbol{x})$ to complement the total energy to E_0. The number of ways that the heat bath may have energy $E_0 - \mathcal{E}(\boldsymbol{x})$ is $\Omega_B(E_0 - \mathcal{E}(\boldsymbol{x}))$, where $\Omega_B(\cdot)$ is the state volume function pertaining to the entropy of the heat bath. In other words, the volume/number of microstates of the *combined* system for which the small subsystem is in microstate $\boldsymbol{x}$ is $\Omega_B(E_0 - \mathcal{E}(\boldsymbol{x}))$. Since the combined system is governed by the microcanonical ensemble, the probability of this is proportional to $\Omega_B(E_0 - \mathcal{E}(\boldsymbol{x}))$. More precisely:

$$P(\boldsymbol{x}) = \frac{\Omega_B(E_0 - \mathcal{E}(\boldsymbol{x}))}{\sum_{\boldsymbol{x}'} \Omega_B(E_0 - \mathcal{E}(\boldsymbol{x}'))}. \tag{2.2.31}$$

Let us focus on the numerator for now, and normalize the result at the end. Then,

$$\begin{aligned} P(\boldsymbol{x}) &\propto \Omega_B(E_0 - \mathcal{E}(\boldsymbol{x})) \\ &= \exp\{S_B(E_0 - \mathcal{E}(\boldsymbol{x}))/k\} \\ &\approx \exp\left\{\frac{S_B(E_0)}{k} - \frac{1}{k}\frac{\partial S_B(E)}{\partial E}\bigg|_{E=E_0} \cdot \mathcal{E}(\boldsymbol{x})\right\} \\ &= \exp\left\{\frac{S_B(E_0)}{k} - \frac{1}{kT} \cdot \mathcal{E}(\boldsymbol{x})\right\} \\ &\propto \exp\{-\mathcal{E}(\boldsymbol{x})/(kT)\}. \end{aligned} \tag{2.2.32}$$

It is customary to work with the so called *inverse temperature*:

$$\beta = \frac{1}{kT} \tag{2.2.33}$$

and so,

$$P(\boldsymbol{x}) \propto e^{-\beta\mathcal{E}(\boldsymbol{x})}, \tag{2.2.34}$$

as we have already seen in the example of the ideal gas (where $\mathcal{E}(\boldsymbol{x})$ was the kinetic energy), but now it is much more general. Thus, all that remains to do is to normalize, and we then obtain the *Boltzmann–Gibbs* (B–G) distribution, or the *canonical ensemble*, which describes the underlying probability law in equilibrium:

$$P(\boldsymbol{x}) = \frac{\exp\{-\beta\mathcal{E}(\boldsymbol{x})\}}{Z(\beta)} \tag{2.2.35}$$

where $Z(\beta)$ is the normalization factor:

$$Z(\beta) = \sum_{\boldsymbol{x}} \exp\{-\beta\mathcal{E}(\boldsymbol{x})\} \tag{2.2.36}$$

in the discrete case, or

$$Z(\beta) = \int \mathrm{d}\boldsymbol{x} \exp\{-\beta\mathcal{E}(\boldsymbol{x})\} \tag{2.2.37}$$

in the continuous case. This function is called the *partition function*. As with the function $\Omega(E)$, similar comments apply to the partition function: it must be dimensionless, so if the components of $\boldsymbol{x}$ do have physical units, we must normalize by a 'reference' volume, which in the case of the (ideal) gas is again h^{3N}. By the same token, for indistinguishable particles, it should be divided by $N!$ While the microcanonical ensemble was defined in terms of the extensive variables E, V and N, in the canonical ensemble, we replaced the variable E by the intensive variable that controls it, namely, β (or T). Thus, the full notation of the partition function should be $Z_N(\beta, V)$ or $Z_N(T, V)$.

Exercise 2.1 Show that for the ideal gas

$$Z_N(T, V) = \frac{1}{N!h^{3N}} V^N (2\pi mkT)^{3N/2} = \frac{1}{N!}\left(\frac{V}{\lambda^3}\right)^N \tag{2.2.38}$$

where

$$\lambda \triangleq \frac{h}{\sqrt{2\pi mkT}}. \tag{2.2.39}$$

λ is called the *thermal de Broglie wavelength*.[9]

The formula of the B–G distribution is one of the most fundamental results in statistical mechanics, obtained solely from the energy conservation law and the postulate of the uniform distribution in an isolated system. As we shall see, the meaning of the partition function is by far deeper than just being a normalization constant. Interestingly, a great deal of the macroscopic physical quantities, like the internal energy, the free energy, the entropy, the heat capacity, the pressure, etc., can be obtained from the partition function. This is in analogy to the fact that in the microcanonical ensemble, $S(E)$ (or, more generally, $S(E, V, N)$) was pivotal to the derivation of all macroscopic physical quantities of interest.

---

[9] The origin of this name comes from the wave–particle de Broglie relation $\lambda = h/p$ together with the fact that the denominator, $\sqrt{2\pi mkT}$, can be viewed as a notion of thermal momentum of the ideal gas, given the fact that the average molecular speed is proportional to $\sqrt{kT/m}$ (see Sect. 1.1).

Several comments are in order:

- The B–G distribution tells us that the system "prefers" to visit its low energy states more than the high energy states, and what counts is only energy differences, not absolute energies: If we add to all states a fixed amount of energy E_0, this will result in an extra factor of $e^{-\beta E_0}$ both in the numerator and in the denominator of the B–G distribution, which of course will cancel out.
- In many physical systems, the Hamiltonian is a quadratic (or "harmonic") function, e.g., $\frac{1}{2}mv^2$, $\frac{1}{2}kx^2$, $\frac{1}{2}CV^2$, $\frac{1}{2}LI^2$, $\frac{1}{2}I\omega^2$, etc., in which case the resulting B–G distribution turns out to be Gaussian. This is at least part of the explanation why the Gaussian distribution is so frequently encountered in Nature.
- When the Hamiltonian is additive, that is, $\mathcal{E}(\boldsymbol{x}) = \sum_{i=1}^{N} \mathcal{E}(x_i)$, the various particles are statistically independent: Additive Hamiltonians correspond to non–interacting particles. In other words, the $\{x_i\}$'s behave as if they were drawn from an i.i.d. probability distribution. By the law of large numbers $\frac{1}{N}\sum_{i=1}^{N}\mathcal{E}(x_i)$ will tend (almost surely) to $\epsilon = \langle \mathcal{E}(X_i)\rangle$. Thus, the average energy of the system is about $N \cdot \epsilon$, not only on the average, but moreover, with an overwhelmingly *high probability* for large N. Nonetheless, this is different from the microcanonical ensemble where $\frac{1}{N}\sum_{i=1}^{N}\mathcal{E}(x_i)$ was held *strictly* at the value of ϵ.

One of the important principles of statistical mechanics is that the microcanonical ensemble and the canonical ensemble (with the corresponding temperature) are asymptotically equivalent (in the thermodynamic limit) as far as macroscopic quantities go. They continue to be such, even in cases of interactions, as long as these are short range[10] and the same is true with the other ensembles that we will encounter later in this chapter. This is an important and useful fact, because more often than not, it is more convenient to analyze things in one ensemble rather than in others, and so it is appropriate to pass to another ensemble for the purpose of the analysis, even though the "real system" is in the other ensemble. We will use this *ensemble equivalence principle* many times later on. The important thing, however, is to be consistent and not to mix up two ensembles or more. Having moved to the other ensemble, it is recommended to keep all further analysis in that ensemble.

Exercise 2.2 Consider the ideal gas with gravitation, where the Hamiltonian includes, in addition to the kinetic energy term for each molecule, also an additive term of potential energy mgz_i for the i–th molecule (z_i being its height). Suppose that an ideal gas of N molecules of mass m is confined to a room whose floor and ceiling areas are both A and whose height is h: (i) Write an expression for the joint pdf of the location $\vec{r}$ and the momentum $\vec{p}$ of each molecule. (ii) Use this expression to show that the gas pressures at the floor and the ceiling are given by

$$P_{\text{floor}} = \frac{mgN}{A(1-e^{-mgh/kT})}; \quad P_{\text{ceiling}} = \frac{mgN}{A(e^{mgh/kT}-1)}. \tag{2.2.40}$$

---

[10] This is related to the concavity of $s(\epsilon)$ [3, 4].

Properties of the Partition Function and the Free Energy

Let us now examine more closely the partition function and make a few observations about its basic properties. For simplicity, we shall assume that $\boldsymbol{x}$ is discrete. First, let's look at the limits: obviously, $Z(0)$ is equal to the size of the entire set of microstates, which is also $\sum_E \Omega(E)$. This is the high temperature limit, where all microstates are equiprobable. At the other extreme, we have:

$$\lim_{\beta\to\infty}\frac{\ln Z(\beta)}{\beta} = -\min_{\boldsymbol{x}} \mathcal{E}(\boldsymbol{x}) \stackrel{\Delta}{=} -E_{\text{GS}} \tag{2.2.41}$$

which describes the situation where the system is frozen to the absolute zero. Only states with minimum energy – the *ground–state energy*, prevail.

Another important property of $Z(\beta)$, or more precisely, of $\ln Z(\beta)$, is that it is a cumulant generating function: by taking derivatives of $\ln Z(\beta)$, we can obtain cumulants of $\mathcal{E}(\boldsymbol{x})$. For the first cumulant, we have

$$\langle \mathcal{E}(\boldsymbol{X})\rangle = \frac{\sum_{\boldsymbol{x}} \mathcal{E}(\boldsymbol{x})e^{-\beta\mathcal{E}(\boldsymbol{x})}}{\sum_{\boldsymbol{x}} e^{-\beta\mathcal{E}(\boldsymbol{x})}} = -\frac{\mathrm{d}\ln Z(\beta)}{\mathrm{d}\beta}. \tag{2.2.42}$$

For example, referring to Exercise 2.1, for the ideal gas,

$$Z_N(\beta, V) = \frac{1}{N!}\left(\frac{V}{\lambda^3}\right)^N = \frac{1}{N!}\frac{V^N}{h^{3N}}\cdot\left(\frac{2\pi m}{\beta}\right)^{3N/2}, \tag{2.2.43}$$

thus, $\langle \mathcal{E}(\boldsymbol{X})\rangle = -\mathrm{d}\ln Z_N(\beta, V)/\mathrm{d}\beta = 3N/(2\beta) = 3NkT/2$ in agreement with the result we obtained both in Chap. 1 and in the microcanonical ensemble, thus demonstrating the ensemble equivalence principle. Similarly, it is easy to show that

$$\mathrm{Var}\{\mathcal{E}(\boldsymbol{X})\} = \langle \mathcal{E}^2(\boldsymbol{X})\rangle - \langle \mathcal{E}(\boldsymbol{X})\rangle^2 = \frac{\mathrm{d}^2\ln Z(\beta)}{\mathrm{d}\beta^2}. \tag{2.2.44}$$

This in turn implies that

$$\frac{\mathrm{d}^2\ln Z(\beta)}{\mathrm{d}\beta^2} \geq 0, \tag{2.2.45}$$

which means that $\ln Z(\beta)$ must always be a convex function. Note also that

$$\begin{aligned}\frac{\mathrm{d}^2\ln Z(\beta)}{\mathrm{d}\beta^2} &= -\frac{\mathrm{d}\langle \mathcal{E}(\boldsymbol{x})\rangle}{\mathrm{d}\beta}\\ &= -\frac{\mathrm{d}\langle \mathcal{E}(\boldsymbol{x})\rangle}{\mathrm{d}T}\cdot\frac{\mathrm{d}T}{\mathrm{d}\beta}\\ &= kT^2C(T)\end{aligned} \tag{2.2.46}$$

where $C(T) = \mathrm{d}\langle \mathcal{E}(\boldsymbol{x})\rangle/\mathrm{d}T$ is the heat capacity (at constant volume). Thus, the convexity of $\ln Z(\beta)$ is intimately related to the physical fact that the heat capacity of the system is positive.

Next, we look at the function $Z(\beta)$ slightly differently. Instead of summing the terms $\{e^{-\beta\mathcal{E}(\boldsymbol{x})}\}$ over all states individually, we sum them by energy levels, in a collective manner. This amounts to:

$$\begin{aligned} Z(\beta) &= \sum_{\boldsymbol{x}} e^{-\beta\mathcal{E}(\boldsymbol{x})} \\ &= \sum_{E} \Omega(E) e^{-\beta E} \\ &\approx \sum_{\epsilon} e^{Ns(\epsilon)/k} \cdot e^{-\beta N\epsilon} \\ &= \sum_{\epsilon} \exp\{-N\beta[\epsilon - Ts(\epsilon)]\} \\ &\doteq \max_{\epsilon} \exp\{-N\beta[\epsilon - Ts(\epsilon)]\} \\ &= \exp\{-N\beta \min_{\epsilon}[\epsilon - Ts(\epsilon)]\} \\ &\stackrel{\Delta}{=} \exp\{-N\beta[\epsilon^* - Ts(\epsilon^*)]\} \\ &\stackrel{\Delta}{=} e^{-\beta F}, \end{aligned} \tag{2.2.47}$$

where here and throughout the sequel, the notation $\doteq$ means asymptotic equivalence in the exponential scale. More precisely, $a_N \doteq b_N$ for two positive sequences $\{a_N\}$ and $\{b_N\}$, means that $\lim_{N\to\infty} \frac{1}{N} \ln \frac{a_N}{b_N} = 0$.

The quantity $f \stackrel{\Delta}{=} \epsilon^* - Ts(\epsilon^*)$ is the (per–particle) *free energy*. Similarly, the entire free energy, F, is defined as

$$F \stackrel{\Delta}{=} E - TS = -\frac{\ln Z(\beta)}{\beta} = -kT \ln Z(\beta). \tag{2.2.48}$$

Once again, due to the exponentiality of (2.2.47) in N, with *very high probability* the system would be found in a microstate $\boldsymbol{x}$ whose normalized energy $\epsilon(\boldsymbol{x}) = \mathcal{E}(\boldsymbol{x})/N$ is very close to ϵ^*, the normalized energy that minimizes $\epsilon - Ts(\epsilon)$ and hence achieves f. Note that the minimizing ϵ^* (obtained by equating the derivative of $\epsilon - Ts(\epsilon)$ to zero), is the solution to the equation $s'(\epsilon^*) = 1/T$, which conforms with the definition of temperature. We see then that equilibrium in the canonical ensemble amounts to **minimum free energy.** This extends the second law of thermodynamics from isolated systems to non–isolated ones. While in an isolated system, the second law asserts the principle of maximum entropy, when it comes to a non–isolated system, this rule is replaced by the principle of minimum free energy.

Exercise 2.3 Show that the canonical average pressure is given by

$$P = -\frac{\partial F}{\partial V} = kT \cdot \frac{\partial \ln Z_N(\beta, V)}{\partial V}.$$

Examine this formula for the canonical ensemble of the ideal gas. Compare to the equation of state.

The physical meaning of the free energy, or more precisely, the difference between two free energies F_1 and F_2, is the minimum amount of work that it takes to transfer the system from equilibrium state 1 to another equilibrium state 2 in an isothermal (fixed temperature) process. This minimum is achieved when the process is *quasi–static*, i.e., so slow that the system is always almost in equilibrium. Equivalently, $-\Delta F$ is the maximum amount of energy in the system, that is *free* and useful for performing work (i.e., not dissipated as heat) in fixed temperature.

To demonstrate this point, let us consider the case where $\mathcal{E}(\boldsymbol{x})$ includes a term of a potential energy that is given by the (scalar) product of a certain external force and the conjugate physical variable at which this force is exerted (e.g., pressure times volume, gravitational force times height, moment times angle, magnetic field times magnetic moment, voltage times electric charge, etc.), i.e.,

$$\mathcal{E}(\boldsymbol{x}) = \mathcal{E}_0(\boldsymbol{x}) - \lambda \cdot L(\boldsymbol{x}) \tag{2.2.49}$$

where λ is the force and $L(\boldsymbol{x})$ is the conjugate physical variable, which depends on (some coordinates of) the microstate. The partition function then depends on both β and λ and hence will be denoted[11] $Z(\beta, \lambda)$. It is easy to see (similarly as before) that $\ln Z(\beta, \lambda)$ is convex in λ for fixed β. Also,

$$\langle L(\boldsymbol{x})\rangle = kT \cdot \frac{\partial \ln Z(\beta, \lambda)}{\partial \lambda}. \tag{2.2.50}$$

The free energy is given by[12]

---

[11] Since the term $\lambda \cdot L(\boldsymbol{x})$ is not considered part of the internal energy (but rather an external energy resource), formally, this ensemble is no longer the canonical ensemble, but a somewhat different ensemble, called the Gibbs ensemble, which will be discussed later on.

[12] At this point, there is a distinction between the *Helmholtz free energy* and the *Gibbs free energy*. The former is defined as $F = E - TS$ in general, as mentioned earlier. The latter is defined as $G = E - TS - \lambda L = -kT \ln Z$, where L is shorthand notation for $\langle L(\boldsymbol{x})\rangle$ (the quantity $H = E - \lambda L$ is called the enthalpy). The physical significance of the Gibbs free energy is similar to that of the Helmholtz free energy, except that it refers to the total work of all other external forces in the system (if there are any), except the work contributed by the force λ (**Exercise 2.4** show this!). The passage to the Gibbs ensemble, which replaces a fixed value of $L(\boldsymbol{x})$ (say, constant volume of a gas) by the control of the conjugate external force λ, (say, pressure in the example of a gas) can be carried out by another Legendre–Fenchel transform (see, e.g., [5, Sect. 1.14]) as well as Sect. 2.2.3 in the sequel.

$$\begin{aligned} F &= E - TS \\ &= -kT \ln Z + \lambda \langle L(\boldsymbol{X}) \rangle \\ &= kT \left(\lambda \cdot \frac{\partial \ln Z}{\partial \lambda} - \ln Z \right). \end{aligned} \tag{2.2.51}$$

Now, let F_1 and F_2 be the equilibrium free energies pertaining to two values of λ, denoted λ_1 and λ_2. Then,

$$\begin{aligned} F_2 - F_1 &= \int_{\lambda_1}^{\lambda_2} \mathrm{d}\lambda \cdot \frac{\partial F}{\partial \lambda} \\ &= kT \cdot \int_{\lambda_1}^{\lambda_2} \mathrm{d}\lambda \cdot \lambda \cdot \frac{\partial^2 \ln Z}{\partial \lambda^2} \\ &= \int_{\lambda_1}^{\lambda_2} \mathrm{d}\lambda \cdot \lambda \cdot \frac{\partial \langle L(\boldsymbol{X}) \rangle}{\partial \lambda} \\ &= \int_{\langle L(X) \rangle_{\lambda_1}}^{\langle L(X) \rangle_{\lambda_2}} \lambda \cdot \mathrm{d}\langle L(\boldsymbol{X}) \rangle \end{aligned} \tag{2.2.52}$$

The product $\lambda \cdot \mathrm{d}\langle L(\boldsymbol{X}) \rangle$ designates an infinitesimal amount of (average) work performed by the force λ on a small change in the average of the conjugate variable $\langle L(\boldsymbol{X}) \rangle$, where the expectation is taken w.r.t. the actual value of λ. Thus, the last integral expresses the total work along a slow process of changing the force λ in small steps and letting the system adapt and equilibrate after this small change every time. On the other hand, it is easy to show (using the convexity of $\ln Z$ in λ), that if λ varies in large steps, the resulting amount of work will always be larger.

Let us define

$$\phi(\beta) = \lim_{N \to \infty} \frac{\ln Z(\beta)}{N} \tag{2.2.53}$$

and, in order to avoid dragging the constant k, let us define

$$\Sigma(\epsilon) = \lim_{N \to \infty} \frac{\ln \Omega(N\epsilon)}{N} = \frac{s(\epsilon)}{k}. \tag{2.2.54}$$

Then, the chain of equalities (2.2.47), written slightly differently, gives

$$\begin{aligned} \phi(\beta) &= \lim_{N \to \infty} \frac{\ln Z(\beta)}{N} \\ &= \lim_{N \to \infty} \frac{1}{N} \ln \left\{ \sum_{\epsilon} e^{N[\Sigma(\epsilon) - \beta\epsilon]} \right\} \\ &= \max_{\epsilon} [\Sigma(\epsilon) - \beta\epsilon]. \end{aligned}$$

Thus, $\phi(\beta)$ is (a certain variant of) the *Legendre–Fenchel transform*[13] of $\Sigma(\epsilon)$. As $\Sigma(\epsilon)$ is (normally) a concave, monotonically increasing function, then it can readily be shown that the inverse transform is:

$$\Sigma(\epsilon) = \min_{\beta}[\beta\epsilon + \phi(\beta)]. \tag{2.2.55}$$

The achiever, $\epsilon^*(\beta)$, of $\phi(\beta)$ in the forward transform is obtained by equating the derivative to zero, i.e., it is the solution to the equation

$$\beta = \Sigma'(\epsilon), \tag{2.2.56}$$

where $\Sigma'(\epsilon)$ is the derivative of $\Sigma(\epsilon)$. In other words, $\epsilon^*(\beta)$ the inverse function of $\Sigma'(\cdot)$. By the same token, the achiever, $\beta^*(\epsilon)$, of $\Sigma(\epsilon)$ in the backward transform is obtained by equating the other derivative to zero, i.e., it is the solution to the equation

$$\epsilon = -\phi'(\beta) \tag{2.2.57}$$

or in other words, the inverse function of $-\phi'(\cdot)$. This establishes a relationship between the typical per–particle energy ϵ and the inverse temperature β that gives rise to ϵ (cf. the Lagrange interpretation above, where we said that β controls the average energy).

Example 2.3 (*Two level system*) Similarly to the earlier example of Schottky defects, which was previously given in the context of the microcanonical ensemble, consider now a system of N independent particles, each having two possible states: state 0 of zero energy and state 1, whose energy is ϵ_0, i.e., $\mathcal{E}(x) = \epsilon_0 x$, $x \in \{0, 1\}$. The x_i's are independent, each having a marginal[14]:

$$P(x) = \frac{e^{-\beta\epsilon_0 x}}{1 + e^{-\beta\epsilon_0}} \quad x \in \{0, 1\}. \tag{2.2.58}$$

In this case,

$$\phi(\beta) = \ln(1 + e^{-\beta\epsilon_0}) \tag{2.2.59}$$

---

[13] More precisely, the one–dimensional Legendre–Fenchel transform of a real function $f(x)$ is defined as $g(y) = \sup_x[xy - f(x)]$. If f is convex, it can readily be shown that: (i) The inverse transform has the very same form, i.e., $f(x) = \sup_y[xy - g(y)]$, and (ii) The derivatives $f'(x)$ and $g'(y)$ are inverses of each other.

[14] Note that the expected number of 'activated' particles $\langle n\rangle = NP(1) = Ne^{-\beta\epsilon_0}/(1 + e^{-\beta\epsilon_0}) = N/(e^{\beta\epsilon_0} + 1)$, in agreement with the result of Example 2.1 (Eq. (2.2.17)). This demonstrates the ensemble equivalence principle.

and

$$\Sigma(\epsilon) = \min_{\beta \geq 0}[\beta\epsilon + \ln(1 + e^{-\beta\epsilon_0})]. \tag{2.2.60}$$

To find $\beta^*(\epsilon)$, we take the derivative and equate to zero:

$$\epsilon - \frac{\epsilon_0 e^{-\beta\epsilon_0}}{1 + e^{-\beta\epsilon_0}} = 0 \tag{2.2.61}$$

which gives

$$\beta^*(\epsilon) = \frac{\ln(\epsilon_0/\epsilon - 1)}{\epsilon_0}. \tag{2.2.62}$$

On substituting this back into the above expression of $\Sigma(\epsilon)$, we get:

$$\Sigma(\epsilon) = \frac{\epsilon}{\epsilon_0} \ln\left(\frac{\epsilon}{\epsilon_0} - 1\right) + \ln\left[1 + \exp\left\{-\ln\left(\frac{\epsilon}{\epsilon_0} - 1\right)\right\}\right], \tag{2.2.63}$$

which after a short algebraic manipulation, becomes

$$\Sigma(\epsilon) = h_2\left(\frac{\epsilon}{\epsilon_0}\right), \tag{2.2.64}$$

just like in the Schottky example. In the other direction:

$$\phi(\beta) = \max_{\epsilon}\left[h_2\left(\frac{\epsilon}{\epsilon_0}\right) - \beta\epsilon\right], \tag{2.2.65}$$

whose achiever $\epsilon^*(\beta)$ solves the zero–derivative equation:

$$\frac{1}{\epsilon_0} \ln\left[\frac{1 - \epsilon/\epsilon_0}{\epsilon/\epsilon_0}\right] = \beta \tag{2.2.66}$$

or equivalently,

$$\epsilon^*(\beta) = \frac{\epsilon_0}{1 + e^{-\beta\epsilon_0}}, \tag{2.2.67}$$

which is exactly the inverse function of $\beta^*(\epsilon)$ above, and which when substituted back into the expression of $\phi(\beta)$, indeed gives

$$\phi(\beta) = \ln(1 + e^{-\beta\epsilon_0}). \tag{2.2.68}$$

This concludes Example 2.3.

Comment A very similar model (and hence with similar results) pertains to non–interacting spins (magnetic moments), where the only difference is that $x \in \{-1, +1\}$ rather than $x \in \{0, 1\}$. Here, the meaning of the parameter ϵ_0 becomes that of a magnetic field, which is more customarily denoted by B (or H), and which is either parallel or anti-parallel to that of the spin, and so the potential energy (in the appropriate physical units), $\vec{B} \cdot \vec{x}$, is either Bx or $-Bx$. Thus,

$$P(x) = \frac{e^{\beta Bx}}{2\cosh(\beta B)}; \quad Z(\beta) = 2\cosh(\beta B). \tag{2.2.69}$$

The net *magnetization* per–spin is defined as

$$m \stackrel{\Delta}{=} \left\langle \frac{1}{N} \sum_{i=1}^{N} X_i \right\rangle = \langle X_1 \rangle = \frac{\partial \phi}{\partial(\beta B)} = \tanh(\beta B). \tag{2.2.70}$$

This is the paramagnetic characteristic of the magnetization as a function of the magnetic field: as $B \to \pm\infty$, the magnetization $m \to \pm 1$ accordingly. When the magnetic field is removed ($B = 0$), the magnetization vanishes too. We will get back to this model and its extensions in Chap. 5. □

Now, observe that whenever β and ϵ are related as explained above, we have:

$$\Sigma(\epsilon) = \beta\epsilon + \phi(\beta) = \phi(\beta) - \beta \cdot \phi'(\beta). \tag{2.2.71}$$

The Gibbs entropy per particle is defined in its normalized form as

$$\bar{H} = -\lim_{N\to\infty} \frac{1}{N} \sum_{\boldsymbol{x}} P(\boldsymbol{x}) \ln P(\boldsymbol{x}) = -\lim_{N\to\infty} \frac{1}{N} \langle \ln P(\boldsymbol{x}) \rangle, \tag{2.2.72}$$

which in the case of the B–G distribution amounts to

$$\begin{aligned}
\bar{H} &= \lim_{N\to\infty} \frac{1}{N} \left\langle \ln \frac{Z(\beta)}{e^{-\beta \mathcal{E}(\boldsymbol{X})}} \right\rangle \\
&= \lim_{N\to\infty} \left[\frac{\ln Z(\beta)}{N} + \frac{\beta \langle \mathcal{E}(\boldsymbol{X}) \rangle}{N} \right] \\
&= \phi(\beta) - \beta \cdot \phi'(\beta),
\end{aligned}$$

but this is exactly the same expression as in (2.2.71), and so, $\Sigma(\epsilon)$ and $\bar{H}$ are identical whenever β and ϵ are related accordingly. The former, as we recall, we defined as the normalized logarithm of the number of microstates with per–particle energy ϵ. Thus, we have learned that the number of such microstates is of the exponential order of $e^{N\bar{H}}$. Another look at this relation is the following:

$$\begin{aligned}1 \geq \sum_{\boldsymbol{x}:\ \mathcal{E}(\boldsymbol{x})\approx N\epsilon} P(\boldsymbol{x}) &= \sum_{\boldsymbol{x}:\ \mathcal{E}(\boldsymbol{x})\approx N\epsilon} \frac{\exp\{-\beta\sum_i \mathcal{E}(x_i)\}}{Z^N(\beta)} \\ &\doteq \sum_{\boldsymbol{x}:\ \mathcal{E}(\boldsymbol{x})\approx N\epsilon} \exp\{-\beta N\epsilon - N\phi(\beta)\} \\ &= \Omega(N\epsilon)\cdot\exp\{-N[\beta\epsilon+\phi(\beta)]\} \end{aligned} \tag{2.2.73}$$

which means that

$$\Omega(N\epsilon) \leq \exp\{N[\beta\epsilon+\phi(\beta)]\} \tag{2.2.74}$$

for all β, and so,

$$\Omega(N\epsilon) \leq \exp\{N\min_\beta[\beta\epsilon+\phi(\beta)]\} = e^{N\Sigma(\epsilon)} = e^{N\bar{H}}. \tag{2.2.75}$$

A compatible lower bound is obtained by observing that the minimizing β gives rise to $\langle\mathcal{E}(X_1)\rangle = \epsilon$, which makes the event $\{\boldsymbol{x}:\ \mathcal{E}(\boldsymbol{x}) \approx N\epsilon\}$ a high–probability event, by the weak law of large numbers. A good reference for further study, and from a more general perspective, is the article by Hall [6]. See also [7].

Now, that we identified the Gibbs entropy with the Boltzmann entropy, it is instructive to point out that the B–G distribution could have been obtained also in a different manner, owing to the maximum–entropy principle that stems from the second law, or the minimum free–energy principle. Specifically, let us denote the Gibbs entropy as

$$H(P) = -\sum_{\boldsymbol{x}} P(\boldsymbol{x})\ln P(\boldsymbol{x}) \tag{2.2.76}$$

and consider the following optimization problem:

$$\begin{aligned} &\max\ H(P) \\ &\text{s.t. } \langle\mathcal{E}(\boldsymbol{X})\rangle = E \end{aligned} \tag{2.2.77}$$

By formalizing the equivalent Lagrange problem, where β now plays the role of a Lagrange multiplier:

$$\max\left\{H(P) + \beta\left[E - \sum_{\boldsymbol{x}} P(\boldsymbol{x})\mathcal{E}(\boldsymbol{x})\right]\right\}, \tag{2.2.78}$$

or equivalently,

$$\min\left\{\sum_{\boldsymbol{x}} P(\boldsymbol{x})\mathcal{E}(\boldsymbol{x}) - \frac{H(P)}{\beta}\right\} \tag{2.2.79}$$

one readily verifies that the solution to this problem is the B–G distribution where the choice of the (Lagrange multiplier) β controls the average energy E. If β is identified with the inverse temperature, the above is nothing but the minimization of the free energy.

Note also that Eq. (2.2.71), which we will rewrite, with a slight abuse of notation as

$$\phi(\beta) - \beta\phi'(\beta) = \Sigma(\beta) \tag{2.2.80}$$

can be viewed in two ways. The first suggests to take derivatives of both sides w.r.t. β and then obtain $\Sigma'(\beta) = -\beta\phi''(\beta)$ and so,

$$\begin{aligned} s(\beta) &= k\Sigma(\beta) \\ &= k\int_{\beta}^{\infty} \tilde{\beta}\phi''(\tilde{\beta})\mathrm{d}\tilde{\beta} \qquad \text{3rd law} \\ &= k\int_0^T \frac{1}{k\tilde{T}} \cdot k\tilde{T}^2 c(\tilde{T}) \cdot \frac{\mathrm{d}\tilde{T}}{k\tilde{T}^2} \qquad c(\tilde{T}) \stackrel{\Delta}{=} \text{heat capacity per particle} \\ &= \int_0^T \frac{c(\tilde{T})\mathrm{d}\tilde{T}}{\tilde{T}} \end{aligned} \tag{2.2.81}$$

recovering the Clausius entropy as $c(\tilde{T})\mathrm{d}\tilde{T}$ is the increment of heat intake per particle $\mathrm{d}q$. The second way to look at Eq. (2.2.80) is as a first order differential equation in $\phi(\beta)$, whose solution is easily found to be

$$\phi(\beta) = -\beta\epsilon_{\text{GS}} + \beta \cdot \int_{\beta}^{\infty} \frac{\mathrm{d}\hat{\beta}\Sigma(\hat{\beta})}{\hat{\beta}^2}, \tag{2.2.82}$$

where $\epsilon_{\text{GS}} = \lim_{N\to\infty} E_{\text{GS}}/N$. Equivalently,

$$Z(\beta) \doteq \exp\left\{-\beta E_{\text{GS}} + N\beta \cdot \int_{\beta}^{\infty} \frac{\mathrm{d}\hat{\beta}\Sigma(\hat{\beta})}{\hat{\beta}^2}\right\}, \tag{2.2.83}$$

namely, the partition function at a certain temperature can be expressed as a functional of the entropy pertaining to all temperatures lower than that temperature. Changing the integration variable from β to T, this readily gives the relation

$$F = E_{\text{GS}} - \int_0^T S(T')\mathrm{d}T'. \tag{2.2.84}$$

Since $F = E - ST$, we have

$$E = E_{\text{GS}} + ST - \int_0^T S(T')\mathrm{d}T' = E_{\text{GS}} + \int_0^S T(S')\mathrm{d}S', \tag{2.2.85}$$

where the second term amounts to the heat Q that accumulates in the system, as the temperature is raised from 0 to T. This is a special case of the first law of thermodynamics. The more general form, as said, takes into account also possible work performed on (or by) the system.

Let us now summarize the main properties of the partition function that we have seen thus far:

1. $Z(\beta)$ is a continuous function. $Z(0) = |\mathcal{X}^n|$ and $\lim_{\beta\to\infty} \frac{\ln Z(\beta)}{\beta} = -E_{\mathrm{GS}}$.
2. Generating cumulants: $\langle \mathcal{E}(\boldsymbol{X}) \rangle = -\mathrm{d}\ln Z/\mathrm{d}\beta$, $\mathrm{Var}\{\mathcal{E}(\boldsymbol{X})\} = \mathrm{d}^2 \ln Z/\mathrm{d}\beta^2$, which implies convexity of $\ln Z$, and hence also of $\phi(\beta)$.
3. ϕ and Σ are a Legendre–Fenchel transform pair. Σ is concave.

We have also seen that Boltzmann's entropy is not only equivalent to the Clausius entropy, but also to the Gibbs/Shannon entropy. Thus, there are actually three different forms of the expression of entropy.

Comment Consider $Z(\beta)$ for an *imaginary temperature* $\beta = i\omega$, where $i = \sqrt{-1}$, and define $z(E)$ as the inverse Fourier transform of $Z(i\omega)$. It can readily be seen that $z(E) = \omega(E)$, i.e., for $E_1 < E_2$, the number of states with energy between E_1 and E_2 is given by $\int_{E_1}^{E_2} z(E)\mathrm{d}E$. Thus, $Z(\cdot)$ can be related to energy enumeration in two different ways: one is by the Legendre–Fenchel transform of $\ln Z(\beta)$ for real β, and the other is by the inverse Fourier transform of $Z(\beta)$ for imaginary β. It should be kept in mind, however, that while the latter relation holds for every system size N, the former is true only in the thermodynamic limit, as mentioned.

The Energy Equipartition Theorem

It turns out that in the case of a quadratic Hamiltonian, $\mathcal{E}(x) = \frac{1}{2}\alpha x^2$, which means that x is Gaussian, the average per–particle energy, is always given by $1/(2\beta) = kT/2$, independently of α. If we have N such quadratic terms, then of course, we end up with $NkT/2$. In the case of the ideal gas, we have three such terms (one for each dimension) per particle, thus a total of $3N$ terms, and so, $E = 3NkT/2$, which is exactly the expression we obtained also from the microcanonical ensemble as well as in the previous chapter. In fact, we observe that in the canonical ensemble, whenever we have an Hamiltonian of the form $\frac{\alpha}{2}x_i^2$ plus some arbitrary terms that do not depend on x_i, then x_i is Gaussian (with variance kT/α) and independent of the other variables, i.e., $p(x_i) \propto e^{-\alpha x_i^2/(2kT)}$. Hence it contributes an amount of

$$\left\langle \frac{1}{2}\alpha X_i^2 \right\rangle = \frac{1}{2}\alpha \cdot \frac{kT}{\alpha} = \frac{kT}{2} \tag{2.2.86}$$

to the total average energy, independently of α. It is more precise to refer to this x_i as a *degree of freedom* rather than a particle. This is because in the three–dimensional world, the kinetic energy, for example, is given by $p_x^2/(2m) + p_y^2/(2m) + p_z^2/(2m)$, that is, each particle contributes *three* additive quadratic terms rather than one (just like three independent one–dimensional particles) and so, it contributes $3kT/2$. This principle is called the *energy equipartition theorem*.

Below is a direct derivation of the equipartition theorem:

$$\begin{aligned}\left\langle\frac{1}{2}\alpha X^2\right\rangle &= \frac{\int_{-\infty}^{\infty}\mathrm{d}x(\alpha x^2/2)e^{-\beta\alpha x^2/2}}{\int_{-\infty}^{\infty}\mathrm{d}xe^{-\beta\alpha x^2/2)}}\\ &= -\frac{\partial}{\partial\beta}\ln\left[\int_{-\infty}^{\infty}\mathrm{d}xe^{-\beta\alpha x^2/2}\right]\\ &= -\frac{\partial}{\partial\beta}\ln\left[\frac{1}{\sqrt{\beta}}\int_{-\infty}^{\infty}\mathrm{d}(\sqrt{\beta}x)e^{-\alpha(\sqrt{\beta}x)^2/2}\right]\\ &= -\frac{\partial}{\partial\beta}\ln\left[\frac{1}{\sqrt{\beta}}\int_{-\infty}^{\infty}\mathrm{d}ue^{-\alpha u^2/2}\right]\\ &= \frac{1}{2}\frac{\mathrm{d}\ln\beta}{\mathrm{d}\beta} = \frac{1}{2\beta} = \frac{kT}{2}.\end{aligned}$$

Note that although we could have used closed–form expressions for both the numerator and the denominator of the first line, we have deliberately taken a somewhat different route in the second line, where we have presented it as the derivative of the denominator of the first line. Also, rather than calculating the Gaussian integral explicitly, we only figured out how it scales with β, because this is the only thing that matters after taking the derivative relative to β. The reason for using this trick of bypassing the need to calculate integrals, is that it can easily be extended in two directions at least:

1. Let $\boldsymbol{x} \in \mathbb{R}^N$ and let $\mathcal{E}(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x}$, where A is a $N \times N$ positive definite matrix. This corresponds to a physical system with a quadratic Hamiltonian, which includes also interactions between pairs (e.g., harmonic oscillators or springs, which are coupled because they are tied to one another). It turns out that here, regardless of A, we get:

$$\langle\mathcal{E}(\boldsymbol{X})\rangle = \left\langle\frac{1}{2}\boldsymbol{X}^T A\boldsymbol{X}\right\rangle = N\cdot\frac{kT}{2}. \tag{2.2.87}$$

2. Back to the case of a scalar x, but suppose now a more general power–law Hamiltonian, $\mathcal{E}(x) = \alpha|x|^\theta$. In this case, we get

$$\langle\mathcal{E}(X)\rangle = \langle\alpha|X|^\theta\rangle = \frac{kT}{\theta}. \tag{2.2.88}$$

Moreover, if $\lim_{x\to\pm\infty} xe^{-\beta\mathcal{E}(x)} = 0$ for all $\beta > 0$, and we denote $\mathcal{E}'(x) \stackrel{\Delta}{=} \mathrm{d}\mathcal{E}(x)/\mathrm{d}x$, then

$$\langle X\cdot\mathcal{E}'(X)\rangle = kT. \tag{2.2.89}$$

It is easy to see that the earlier power–law result is obtained as a special case of this, as $\mathcal{E}'(x) = \alpha\theta|x|^{\theta-1}\text{sgn}(x)$ in this case.

Example 2.4 (*Ideal gas with gravitation*) Let

$$\mathcal{E}(x) = \frac{p_x^2 + p_y^2 + p_z^2}{2m} + mgz. \tag{2.2.90}$$

The average kinetic energy of each particle is $3kT/2$, as said before. The contribution of the average potential energy is kT (one degree of freedom with $\theta = 1$). Thus, the total is $5kT/2$, where 60% come from kinetic energy and 40% come from potential energy, universally, that is, independent of T, m, and g. □

2.2.3 The Grand–Canonical Ensemble and the Gibbs Ensemble

A brief summary of what we have done thus far, is the following: we started with the microcanonical ensemble, which was very restrictive in the sense that the energy was held strictly fixed to the value of E, the number of particles was held strictly fixed to the value of N, and at least in the example of a gas, the volume was also held strictly fixed to a certain value V. In the passage from the microcanonical ensemble to the canonical one, we slightly relaxed the first of these parameters, E: rather than insisting on a fixed value of E, we allowed energy to be exchanged back and forth with the environment, and thereby to slightly fluctuate (for large N) around a certain average value, which was controlled by temperature, or equivalently, by the choice of β. This was done while keeping in mind that the total energy of both system and heat bath must be kept fixed, by the law of energy conservation, which allowed us to look at the combined system as an isolated one, thus obeying the microcanonical ensemble. We then had a one–to–one correspondence between the extensive quantity E and the intensive variable β, that adjusted its average value. But the other extensive variables, like N and V were still kept strictly fixed.

It turns out, that we can continue in this spirit, and 'relax' also either one of the other variables N or V (but not both at the same time), allowing it to fluctuate around a typical average value, and controlling it by a corresponding intensive variable. Like E, both N and V are also subjected to conservation laws when the combined system is considered. Each one of these relaxations leads to a new ensemble, in addition to the microcanonical and the canonical ensembles that we have already seen. In the case where it is the variable V that is allowed to be flexible, this ensemble is called the *Gibbs ensemble*. In the case where it is the variable N, this ensemble is called the *grand–canonical ensemble*. There are, of course, additional ensembles based on this principle, depending on the kind of the physical system.

The Grand–Canonical Ensemble

The fundamental idea is essentially the very same as the one we used to derive the canonical ensemble: let us get back to our (relatively small) subsystem, which is in contact with a heat bath, and this time, let us allow this subsystem to exchange with the heat bath, not only energy, but also matter, i.e., particles. The heat bath consists of a huge reservoir of energy and particles. The total energy is E_0 and the total number of particles is N_0. Suppose that we can calculate the number/volume of states of the heat bath as a function of both its energy E' and amount of particles N', and denote this function by $\Omega_B(E', N')$. A microstate is now a combination $(\boldsymbol{x}, N)$, where N is the (variable) number of particles in our subsystem and $\boldsymbol{x}$ is as before for a given N. From the same considerations as before, whenever our subsystem is in state $(\boldsymbol{x}, N)$, the heat bath can be in any one of $\Omega_B(E_0 - \mathcal{E}(\boldsymbol{x}), N_0 - N)$ microstates of its own. Thus, owing to the microcanonical ensemble,

$$\begin{aligned}P(\boldsymbol{x}, N) &\propto \Omega_B(E_0 - \mathcal{E}(\boldsymbol{x}), N_0 - N) \\ &= \exp\{S_B(E_0 - \mathcal{E}(\boldsymbol{x}), N_0 - N)/k\} \\ &\approx \exp\left\{\frac{S_B(E_0, N_0)}{k} - \frac{1}{k}\frac{\partial S_B}{\partial E}\cdot\mathcal{E}(\boldsymbol{x}) - \frac{1}{k}\frac{\partial S_B}{\partial N}\cdot N\right\} \\ &\propto \exp\left\{-\frac{\mathcal{E}(\boldsymbol{x})}{kT} + \frac{\mu N}{kT}\right\}\end{aligned} \tag{2.2.91}$$

where μ is the chemical potential of the heat bath. Thus, we now have the grand–canonical distribution:

$$P(\boldsymbol{x}, N) = \frac{e^{\beta[\mu N - \mathcal{E}(\boldsymbol{x})]}}{\Xi(\beta, \mu)}, \tag{2.2.92}$$

where the denominator is called the *grand partition function*:

$$\Xi(\beta, \mu) \triangleq \sum_{N=0}^{\infty} e^{\beta\mu N} \sum_{\boldsymbol{x}} e^{-\beta\mathcal{E}(\boldsymbol{x})} \triangleq \sum_{N=0}^{\infty} e^{\beta\mu N} Z_N(\beta). \tag{2.2.93}$$

Example 2.5 (*Grand partition function of the ideal gas*) Using the result of Exercise 2.1, we have for the ideal gas:

$$\begin{aligned}\Xi(\beta, \mu) &= \sum_{N=0}^{\infty} e^{\beta\mu N}\cdot\frac{1}{N!}\left(\frac{V}{\lambda^3}\right)^N \\ &= \sum_{N=0}^{\infty} \frac{1}{N!}\left(e^{\beta\mu}\cdot\frac{V}{\lambda^3}\right)^N\end{aligned}$$

$$= \exp\left(e^{\beta\mu} \cdot \frac{V}{\lambda^3}\right). \tag{2.2.94}$$

It is convenient to change variables and to define $z = e^{\beta\mu}$ (which is called the *fugacity*) and then, define

$$\tilde{\Xi}(\beta, z) = \sum_{N=0}^{\infty} z^N Z_N(\beta). \tag{2.2.95}$$

This notation emphasizes the fact that for a given β, $\tilde{\Xi}(z)$ is actually the z–transform of the sequence $\{Z_N(\beta)\}_{N\geq 0}$. A natural way to think about $P(\boldsymbol{x}, N)$ is as $P(N) \cdot P(\boldsymbol{x}|N)$, where $P(N)$ is proportional to $z^N Z_N(\beta)$ and $P(\boldsymbol{x}|N)$ corresponds to the canonical ensemble as before.

Using the grand partition function, it is now easy to obtain moments of the random variable N. For example, the first moment is:

$$\langle N\rangle = \frac{\sum_N N z^N Z_N(\beta)}{\sum_N z^N Z_N(\beta)} = z \cdot \frac{\partial \ln \tilde{\Xi}(\beta, z)}{\partial z}. \tag{2.2.96}$$

Thus, we have replaced the fixed number of particles, N, by a random number of particles, which concentrates around an average controlled by μ, or equivalently, by z. The dominant[15] value of N is the one that maximizes the product $z^N Z_N(\beta)$, or equivalently, $\beta\mu N + \ln Z_N(\beta) = \beta(\mu N - F_N)$. Thus, $\ln \tilde{\Xi}$ is related to $\ln Z_N$ by another kind of a Legendre–Fenchel transform: $\ln \tilde{\Xi}(\beta, z, V) \approx \max_N[\beta\mu N + \ln Z_N(\beta, V)]$ or equivalently

$$kT \ln \tilde{\Xi}(\beta, z, V) \approx \max_N[\mu N + kT \ln Z_N(\beta, V)].$$

Note that by passing to the grand–canonical ensemble, we have replaced two extensive quantities, E and N, by their respective conjugate intensive variables, T and μ. This means that the grand partition function depends on one remaining extensive variable only, which is V, and so, under ordinary conditions, $\ln \Xi(\beta, z)$, or in its more complete notation, $\ln \Xi(\beta, z, V)$, depends linearly on V at least in the thermodynamic limit, namely, $\lim_{V\to\infty}[\ln \Xi(\beta, z, V)]/V$ tends to a constant that depends only on β and z. What is this constant? Let us examine again the first law in its more general form, as it appears in Eq. (2.2.30). For fixed T and μ, we have the following:

---

[15] The best way to understand this is in analogy to the derivation of ϵ^* as the minimizer of the free energy in the canonical ensemble, except that now the 'big' extensive variable is V rather than N, so that $z^N Z_N(\beta, V)$ is roughly exponential in V for a given fixed $\rho = N/V$. The exponential coefficient depends on ρ, and the 'dominant' ρ^* maximizes this coefficient. Finally, the 'dominant' N is $N^* = \rho^* V$.

$$\begin{aligned} P\delta V &= \mu\delta N + T\delta S - \delta E \\ &= \delta(\mu N + TS - E) \\ &= \delta(\mu N - F) \\ &\approx kT \cdot \delta[\ln \Xi(\beta, z, V)] \quad V \text{ large} \end{aligned} \tag{2.2.97}$$

Thus, the constant of proportionality must be P. In other words, the grand–canonical formula of the pressure is:

$$P = kT \cdot \lim_{V\to\infty} \frac{\ln \Xi(\beta, z, V)}{V}. \tag{2.2.98}$$

This is different from the canonical–ensemble formula (Exercise 2.3): $P = kT \cdot \partial \ln Z_N(\beta, V)/\partial V$, and to the microcanonical–ensemble formula, $P = T \cdot \partial S(E, V, N)/\partial V$.

Example 2.6 (*more on the ideal gas*) Applying formula (2.2.96) to Eq. (2.2.94), we readily obtain

$$\langle N\rangle = \frac{zV}{\lambda^3} = \frac{e^{\mu/kT}V}{\lambda^3}. \tag{2.2.99}$$

We see that the grand–canonical factor $e^{\mu/kT}$ has the physical meaning of the average number of ideal gas atoms in a cube of size $\lambda \times \lambda \times \lambda$, where λ is the thermal de Broglie wavelength. Now, applying Eqs. (2.2.98) to (2.2.94), we get

$$P = \frac{kT \cdot e^{\mu/kT}}{\lambda^3} = \frac{\langle N\rangle \cdot kT}{V}, \tag{2.2.100}$$

recovering again the equation of state of the ideal gas. This also demonstrates the principle of ensemble equivalence.

Once again, it should be pointed out that beyond the obvious physical significance of the grand–canonical ensemble, sometimes it proves useful to work with from reasons of pure mathematical convenience, using the principle of ensemble equivalence. We will see this very clearly in the next chapters on quantum statistics.

The Gibbs Ensemble

Consider now the case where T and N are fixed, but V is allowed to fluctuate around an average volume controlled by the pressure P. Again, we can analyze our relatively small test system surrounded by a heat bath. The total energy is E_0 and the total volume of the system and the heat bath is V_0. Suppose that we can calculate the count/volume of states of the heat bath as function of both its energy E' and the volume V', call it $\Omega_B(E', V')$. A microstate is now a combination $(\boldsymbol{x}, V)$, where V is the (variable) volume of our subsystem. Once again, the same line of thought is used: whenever our subsystem is at state $(\boldsymbol{x}, V)$, the heat bath may be in any one of $\Omega_B(E_0 - \mathcal{E}(\boldsymbol{x}), V_0 - V)$ microstates of its own. Thus,

$$\begin{aligned}
P(\boldsymbol{x}, V) &\propto \Omega_B(E_0 - \mathcal{E}(\boldsymbol{x}), V_0 - V) \\
&= \exp\{S_B(E_0 - \mathcal{E}(\boldsymbol{x}), V_0 - V)/k\} \\
&\approx \exp\left\{\frac{S_B(E_0, V_0)}{k} - \frac{1}{k}\frac{\partial S_B}{\partial E} \cdot \mathcal{E}(\boldsymbol{x}) - \frac{1}{k}\frac{\partial S_B}{\partial V} \cdot V\right\} \\
&\propto \exp\left\{-\frac{\mathcal{E}(\boldsymbol{x})}{kT} - \frac{PV}{kT}\right\} \\
&= \exp\{-\beta[\mathcal{E}(\boldsymbol{x}) + PV]\}.
\end{aligned} \tag{2.2.101}$$

The corresponding partition function that normalizes this probability function is given by

$$Y_N(\beta, P) = \int_0^\infty e^{-\beta PV} Z_N(\beta, V)\mathrm{d}V = \int_0^\infty e^{-\beta PV}\mathrm{d}V \sum_{\boldsymbol{x}} e^{-\beta\mathcal{E}(\boldsymbol{x})}. \tag{2.2.102}$$

For a given N and β, the function $Y_N(\beta, P)$ can be thought of as the *Laplace transform* of $Z_N(\beta, V)$ as a function of V. In the thermodynamic limit, $\lim_{N\to\infty} \frac{1}{N} \ln Y_N(\beta, P)$ is the Legendre–Fenchel transform of $\lim_{N\to\infty} \frac{1}{N} \ln Z_N(\beta, V)$ for fixed β, similarly to the Legendre–Fenchel relationship between the entropy and the canonical log–partition function.[16] Note that analogously to Eq. (2.2.96), here the Gibbs partition function serves as a cumulant generating function for the random variable V, thus, for example,

$$\langle V \rangle = -kT \cdot \frac{\partial \ln Y_N(\beta, P)}{\partial P}. \tag{2.2.103}$$

As mentioned in footnote no. 20,

$$G = -kT \ln Y_N(\beta, P) = E - TS + PV = F + PV \tag{2.2.104}$$

is the Gibbs free energy of the system, and for the case considered here, the force is the pressure and the conjugate variable it controls is the volume. In analogy to the grand–canonical ensemble, here too, there is only one extensive variable, this time, it is N. Thus, G should be asymptotically proportional to N with a constant of proportionality that depends on the fixed values of T and P.

---

[16] **Exercise 2.5** Write explicitly the Legendre–Fenchel relation (and its inverse) between the Gibbs partition function and the canonical partition function.

Exercise 2.6 Show that this constant is the chemical potential μ.

All this is, of course, relevant when the physical system is a gas in a container. In general, the Gibbs ensemble is obtained by a similar Legendre–Fenchel transform replacing an extensive physical quantity of the canonical ensemble by its conjugate force. For example, magnetic field is conjugate to magnetization, electric field is conjugate to electric charge, mechanical force is conjugate to displacement, moment is conjugate to angular shift, and so on. By the same token, the chemical potential is a 'force' that is conjugate to the number of particles in grand–canonical ensemble, and (inverse) temperature is a 'force' that controls the heat energy.

Figure 2.2 summarizes the thermodynamic potentials associated with the various statistical ensembles. The arrow between each two connected blocks in the diagram designates a passage from one ensemble to another by a Legendre–Fenchel transform operator $\mathcal{L}$ that is defined generically at the bottom of the figure. In each passage, it is also indicated which extensive variable is replaced by its conjugate intensive variable.

It should be noted, that at least mathematically, one could have defined three more ensembles that would complete the picture of Fig. 2.2 in a symmetric manner. Two of the additional ensembles can be obtained by applying Legendre–Fenchel transforms on $S(E, V, N)$, other than the transform that takes us to the canonical ensemble. The first Legendre–Fenchel transform is w.r.t. the variable V, replacing it by P, and the second additional ensemble is w.r.t. the variable N, replacing it by μ. Let us denote the new resulting 'potentials' (minus kT times log–partition functions) by $A(E, P, N)$ and $B(E, V, \mu)$, respectively. The third ensemble, with potential $C(E, P, \mu)$, whose only extensive variable is E, could be obtained by

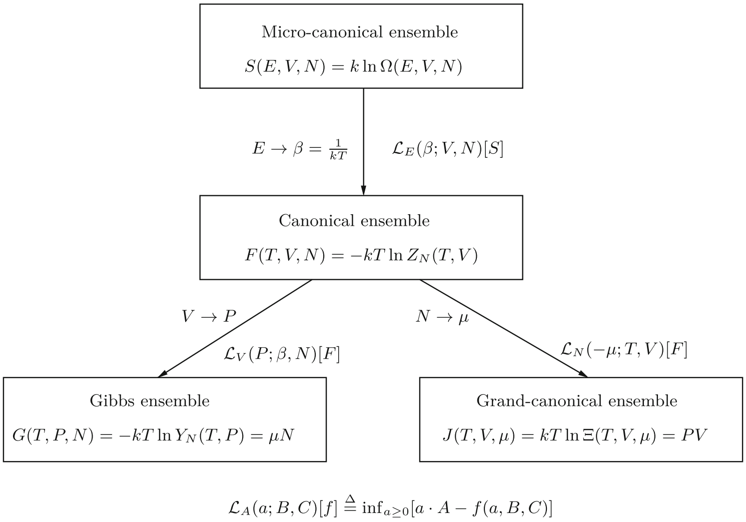


Fig. 2.2 Diagram of Legendre–Fenchel relations between the various ensembles

yet another Legendre–Fenchel transform, either on $A(E, P, N)$ or $B(E, V, \mu)$ w.r.t. the appropriate extensive variable. Of course, $A(E, P, N)$ and $B(E, V, \mu)$ are also connected directly to the Gibbs ensemble and to the grand–canonical ensemble, respectively, both by Legendre–Fenchel–transforming w.r.t. E. While these three ensembles are not really used in physics, they might prove useful to work with for calculating certain physical quantities, by taking advantage of the principle of ensemble equivalence.

Exercise 2.7 Complete the diagram of Fig. 2.2 by the three additional ensembles just defined. Can you give physical meanings to A, B and C? Also, as said, $C(E, P, \mu)$ has only E as an extensive variable. Thus, $\lim_{E\to\infty} C(E, P, \mu)/E$ should be a constant. What is this constant?

Even more generally, we could start from a system model, whose micro–canonical ensemble consists of many extensive variables $L_1, \ldots, L_n$, in addition to the internal energy E (not just V and N). The entropy function is then $S(E, L_1, \ldots, L_n, N)$. Here, L_i can be, for example, volume, mass, electric charge, electric polarization in each one of the three axes, magnetization in each one of three axes, and so on. The first Legendre–Fenchel transform takes us from the micro–canonical ensemble to the canonical one upon replacing E by β. Then we can think of various Gibbs ensembles obtained by replacing any subset of extensive variables L_i by their respective conjugate forces $\lambda_i = T\partial S/\partial L_i$, $i = 1, \ldots, n$ (in the above examples: pressure, gravitational force (weight), voltage (or electric potential), electric fields, and magnetic fields in the corresponding axes, respectively). In the extreme case, all L_i are replaced by λ_i upon applying successive Legendre–Fenchel transforms, or equivalently, a multi–dimensional Legendre–Fenchel transform:

$$G(T, \lambda_1, \ldots, \lambda_n, N) = -\sup_{L_1,\ldots,L_n} [kT \ln Z_N(\beta, L_1, \ldots, L_n) - \lambda_1 L_1 - \ldots - \lambda_n L_n]. \tag{2.2.105}$$

Once again, there must be one extensive variable at least.

2.3 Suggestions for Supplementary Reading

Part of the presentation in this chapter is similar to a corresponding chapter in [1]. The derivations associated with the various ensembles of statistical mechanics, as well as their many properties, can also be found in any textbook on elementary statistical mechanics, including: Beck [8, Chap. 3], Huang [9, Chaps. 6, 7], Honerkamp [10, Chap. 3], Landau and Lifshitz [2], Pathria [11, Chaps. 2–4], and Reif [12, Chap. 6], among many others.

References

1. N. Merhav, Statistical physics and information theory. Found. Trends Commun. Inf. Theory **6**(1–2), 1–212 (2009)
2. L.D. Landau, E.M. Lifshitz, *Course of Theoretical Physics – Volume 5: Statistical Physics, Part 1*, 3rd edn. (Elsevier: Butterworth–Heinemann, New York, 1980)
3. J. Barré, D. Mukamel, S. Ruffo, Inequivalence of ensembles in a system with long-range interactions. Phys. Rev. Lett. **87**(3), 030601 (2001)
4. D. Mukamel, Statistical mechanics of systems with long range interactions, in *AIP Conference Proceedings*, vol. 970, no. 1, ed. by A. Campa, A. Giansanti, G. Morigi, F. Sylos Labini (AIP, 2008), pp. 22–38
5. R. Kubo, *Statistical Mechanics* (North-Holland, New York, 1965)
6. M.J.W. Hall, Universal geometric approach to uncertainty, entropy, and information. Phys. Rev. A **59**(4), 2602–2615 (1999)
7. R.S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics* (Springer, New York, 1985)
8. A.H.W. Beck, *Statistical Mechanics, Fluctuations and Noise* (Edward Arnold Publishers, London, 1976)
9. K. Huang, *Statistical Mechanics*, 2nd edn. (Wiley, New York, 1987)
10. J. Honerkamp, *Statistical Physics - An Advanced Approach with Applications*, 2nd edn. (Springer, Berlin, 2002)
11. R.K. Pathria, *Statistical Mechanics*, 2nd edn. (Elsevier: Butterworth-Heinemann, Oxford, 1996)
12. F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1965)

Chapter 3
Quantum Statistics – The Fermi–Dirac Distribution

In our discussion thus far, we have largely taken for granted the assumption that our system can be analyzed in the classical regime, where quantum effects are negligible. This is, of course, not always the case, especially at very low temperatures. Also, if radiation plays a role in the physical system, then at very high frequency ν, the classical approximation also breaks down. Roughly speaking, kT should be much larger than $h\nu$ for the classical regime to be well justified.[1] It is therefore necessary to address quantum effects in statistical physics issues, most notably, the fact that certain quantities, like energy and angular momentum (or spin), no longer take on values in the continuum, but only in a discrete set, which depends on the system in question.

Consider a gas of identical particles with discrete single–particle quantum states, $1, 2, \ldots, r, \ldots$, corresponding to energies

$$\epsilon_1 \le \epsilon_2 \le \cdots \le \epsilon_r \le \cdots .$$

Since the particles are assumed indistinguishable, then for a gas of N particles, a micro–state is defined by the combination of occupation numbers

$$N_1, N_2, \ldots, N_r, \ldots,$$

where N_r is the number of particles at a single state r.

---

[1] One well–known example is black–body radiation. According to the classical theory, the radiation density per unit frequency grows proportionally to $kT\nu^2$, a function whose integral over ν, from zero to infinity, diverges ("the ultraviolet catastrophe"). This absurd is resolved by quantum mechanical considerations, according to which the factor kT should be replaced by $h\nu/[e^{h\nu/(kT)} - 1]$, which is close to kT at low frequencies, but decays exponentially for $\nu > kT/h$.

N. Merhav, *Statistical Physics for Electrical Engineering*,
DOI 10.1007/978-3-319-62063-3_3

The first fundamental question is the following: what values can the occupation numbers $N_1, N_2, \ldots$ assume? According to quantum mechanics, there might be certain restrictions on these numbers. In particular, there are two kinds of situations that may arise, which divide the various particles in the world into two mutually exclusive classes.

For the first class of particles, there are no restrictions at all. The occupation numbers can assume any non–negative integer value ($N_r = 0, 1, 2, \ldots$). Particles of this class are called *Bose–Einstein (BE) particles*[2] or *bosons* for short. Another feature of bosons is that their spins are always integral multiples of $\hbar$, namely, 0, $\hbar$, $2\hbar$, etc. Examples of bosons are photons, π mesons and K mesons. We will focus on them in the next chapter.

In the second class of particles, the occupation numbers are restricted by the *Pauli exclusion principle* (discovered in 1925), according to which no more than one particle can occupy a given quantum state r (thus N_r is either 0 or 1 for all r), since the wave function of two such particles is anti–symmetric and thus vanishes if they assume the same quantum state (unlike bosons for which the wave function is symmetric). Particles of this kind are called *Fermi–Dirac (FD) particles*[3] or *fermions* for short. Another characteristic of fermions is that their spins are always odd multiples of $\hbar/2$, namely, $\hbar/2$, $3\hbar/2$, $5\hbar/2$, etc. Examples of fermions are electrons, positrons, protons, and neutrons. The statistical mechanics of fermions will be discussed in this chapter.

3.1 Combinatorial Derivation of the FD Statistics

Consider a gas of N fermions in volume V and temperature T. In the thermodynamic limit, where the dimensions of the system are large, the discrete single–particle energy levels $\{\epsilon_r\}$ are very close to each other. Therefore, instead of considering each one of them individually, we shall consider groups of neighboring states. Since the energy levels in each group are very close, we will approximate all of them by a single energy value. Let us label these groups by $s = 1, 2, \ldots$. Let group no. s contain G_s single–particle states and let the representative energy level be $\hat{\epsilon}_s$. Let us assume that $G_s \gg 1$. A (coarse–grained) microstate of the gas is now defined by the occupation numbers

$$\hat{N}_1, \hat{N}_2, \ldots, \hat{N}_s, \ldots,$$

$\hat{N}_s$ being the total number of particles in group no. s, where, of course $\sum_s \hat{N}_s = N$.

---

[2] Bosons were first introduced by Bose (1924) in order to derive Planck's radiation law, and Einstein applied this finding in the same year to a perfect gas of particles.

[3] Introduced independently by Fermi and Dirac in 1926.

To derive the equilibrium behavior of this system, we analyze the Helmholtz free energy F as a function of the occupation numbers, and use the fact that in equilibrium, it should be minimum. Since $E = \sum_s \hat{N}_s \hat{\epsilon}_s$ and $F = E - TS$, this boils down to the evaluation of the entropy $S = k \ln \Omega(\hat{N}_1, \hat{N}_2, \ldots)$. Let $\Omega_s(\hat{N}_s)$ be the number of ways of putting $\hat{N}_s$ particles into G_s states of group no. s. Now, for fermions each one of the G_s states is either empty or occupied by one particle. Thus,

$$\Omega_s(\hat{N}_s) = \frac{G_s!}{\hat{N}_s!(G_s - \hat{N}_s)!} \tag{3.1.1}$$

and

$$\Omega(\hat{N}_1, \hat{N}_2, \ldots) = \prod_s \Omega_s(\hat{N}_s). \tag{3.1.2}$$

Therefore,

$$\begin{aligned} F(\hat{N}_1, \hat{N}_2, \ldots) &= \sum_s [\hat{N}_s \hat{\epsilon}_s - kT \ln \Omega_s(\hat{N}_s)] \\ &\approx \sum_s \left[\hat{N}_s \hat{\epsilon}_s - kTG_s h_2\left(\frac{\hat{N}_s}{G_s}\right)\right]. \end{aligned} \tag{3.1.3}$$

As said, we wish to minimize $F(\hat{N}_1, \hat{N}_2, \ldots)$ s.t. the constraint $\sum_s \hat{N}_s = N$. Consider then the minimization of the Lagrangian[4]

$$L = \sum_s \left[\hat{N}_s \hat{\epsilon}_s - kTG_s h_2\left(\frac{\hat{N}_s}{G_s}\right)\right] - \lambda\left(\sum_s \hat{N}_s - N\right). \tag{3.1.4}$$

The solution is readily obtained to read

$$\hat{N}_s = \frac{G_s}{e^{(\hat{\epsilon}_s - \lambda)/kT} + 1} \tag{3.1.5}$$

where the Lagrange multiplier λ is determined to satisfy the constraint

$$\sum_s \frac{G_s}{e^{(\hat{\epsilon}_s - \lambda)/kT} + 1} = N. \tag{3.1.6}$$

---

[4] For readers that are not familiar with Lagrangians, the minimization of F s.t. $\sum_s \hat{N}_s = N$, is equivalent to the unconstrained minimization of $F - \lambda(\sum_s \hat{N}_s - N)$ for the value of λ at which the constraint is met with equality by the minimizer $\{\hat{N}_s^*\}$. This is because $F(\hat{N}_1^*, \hat{N}_2^*, \ldots) - \lambda(\sum_s \hat{N}_s^* - N) \le F(\hat{N}_1, \hat{N}_2, \ldots) - \lambda(\sum_s \hat{N}_s - N)$, together with $\sum_s \hat{N}_s = \sum_s \hat{N}_s^* = N$, imply $F(\hat{N}_1^*, \hat{N}_2^*, \ldots) \le F(\hat{N}_1, \hat{N}_2, \ldots)$ for every $\{\hat{N}_s\}$ with $\sum_s \hat{N}_s = N$.

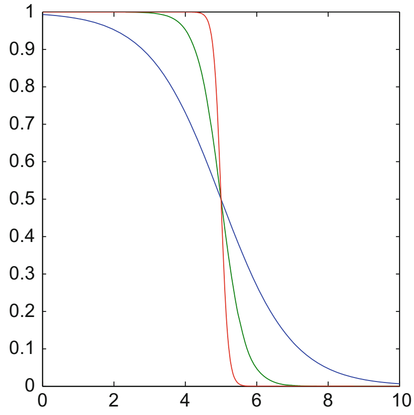


Fig. 3.1 Illustration of the FD distribution. As T decreases, the curve becomes closer to $\bar{N}_r = u(\mu - \epsilon_r)$, where $u(\cdot)$ is the unit step function

Exercise 3.1 After showing the general relation $\mu = (\partial F/\partial N)_{T,V}$, show that $\lambda = \mu$, namely, the Lagrange multiplier λ has the physical meaning of the chemical potential. From now on, then we replace the notation λ by μ.

Note that $\hat{N}_s/G_s$ is the mean occupation number $\bar{N}_r$ of a single state r within group no. s. I.e.,

$$\bar{N}_r = \frac{1}{e^{(\epsilon_r - \mu)/kT} + 1} \tag{3.1.7}$$

with the constraint

$$\sum_r \frac{1}{e^{(\epsilon_r - \mu)/kT} + 1} = N. \tag{3.1.8}$$

It is pleasing that this result no longer depends on the partition into groups. Equation (3.1.7) is the FD distribution, and it is depicted in Fig. 3.1.

3.2 FD Statistics from the Grand–Canonical Ensemble

Thanks to the principle of ensemble equivalence, an alternative, simpler derivation of the FD distribution results from the use of the grand–canonical ensemble. Beginning from the canonical partition function

$$Z_N(\beta) = \sum_{N_1=0}^{1} \sum_{N_2=0}^{1} \dots \delta\left(\sum_r N_r = N\right) e^{-\beta \sum_r N_r \epsilon_r}, \tag{3.2.1}$$

we pass to the grand–canonical ensemble in the following manner:

$$\begin{aligned}
\Xi(\beta,\mu) &= \sum_{N=0}^{\infty} e^{\beta\mu N} \sum_{N_1=0}^{1}\sum_{N_2=0}^{1}\ldots\delta\left(\sum_r N_r = N\right) e^{-\beta\sum_r N_r\epsilon_r} \\
&= \sum_{N_1=0}^{1}\sum_{N_2=0}^{1}\cdots\left[\sum_{N=0}^{\infty}\delta\left(\sum_r N_r = N\right)\right] e^{\beta\sum_r N_r(\mu-\epsilon_r)} \\
&= \sum_{N_1=0}^{1}\sum_{N_2=0}^{1}\ldots e^{\beta\sum_r N_r(\mu-\epsilon_r)} \\
&= \sum_{N_1=0}^{1}\sum_{N_2=0}^{1}\cdots\prod_r e^{\beta N_r(\mu-\epsilon_r)} \\
&= \prod_r\left[\sum_{N_r=0}^{1} e^{\beta N_r(\mu-\epsilon_r)}\right] \\
&= \prod_r\left[1+e^{\beta(\mu-\epsilon_r)}\right].
\end{aligned} \tag{3.2.2}$$

Note that this product form of the grand partition function means that under the grand–canonical ensemble the binary random variables $\{N_r\}$ are statistically independent, i.e.,

$$P(N_1, N_2, \ldots) = \prod_r P_r(N_r) \tag{3.2.3}$$

where

$$P_r(N_r) = \frac{e^{\beta N_r(\mu-\epsilon_r)}}{1+e^{\beta(\mu-\epsilon_r)}}, \qquad N_r = 0, 1, \qquad r = 1, 2, \ldots. \tag{3.2.4}$$

Thus,

$$\bar{N}_r = \Pr\{N_r = 1\} = \frac{e^{(\mu-\epsilon_r)/kT}}{1+e^{(\mu-\epsilon_r)/kT}} = \frac{1}{e^{(\epsilon_r-\mu)/kT}+1}. \tag{3.2.5}$$

Equivalently, defining $\alpha_r = \beta(\mu-\epsilon_r)$, we have $\Xi = \prod_r \sum_{N_r=0}^{1} e^{\alpha_r N_r}$, giving rise to $\bar{N}_r = \partial \ln \Xi/\partial\alpha_r = e^{\alpha_r}/(1+e^{\alpha_r})$, which is the same result.

3.3 The Fermi Energy

Let us now examine what happens if the system is cooled to the absolute zero ($T \to 0$). It should be kept in mind that the chemical potential μ depends on T, so let μ_0 be the chemical potential at $T = 0$. It is readily seen that $\bar{N}_r$ approaches a unit step function (see Fig. 3.1), namely, all energy levels $\{\epsilon_r\}$ below μ_0 are occupied ($\bar{N}_r \approx 1$) by a fermion, whereas all those that are above μ_0 are empty ($\bar{N}_r \approx 0$). The explanation is simple: Pauli's exclusion principle does not allow all particles to reside at the ground state at $T = 0$ since then many of them would occupy the same quantum state. The minimum energy of the system that can possibly be achieved is when all energy levels are filled up, one by one, starting from the ground state up to some maximum level, which is exactly μ_0. This explains why even at the absolute zero, fermions have energy.[5] The maximum occupied energy level in a gas of non–interacting fermions at the absolute zero is called the *Fermi energy*, which we shall denote by ϵ_F. Thus, $\mu_0 = \epsilon_F$, and then the FD distribution at very low temperatures is approximately

$$\bar{N}_r = \frac{1}{e^{(\epsilon_r - \epsilon_F)/kT} + 1}. \tag{3.3.1}$$

We next take a closer look on the FD distribution, taking into account the density of states. Consider a metal box of dimensions $L_x \times L_y \times L_z$ and hence volume $V = L_x L_y L_z$. The energy level associated with quantum number (l_x, l_y, l_z) is given by

$$\epsilon_{l_x,l_y,l_z} = \frac{\pi^2\hbar^2}{2m}\left(\frac{l_x^2}{L_x^2} + \frac{l_y^2}{L_y^2} + \frac{l_z^2}{L_z^2}\right) = \frac{\hbar^2}{2m}(k_x^2 + k_y^2 + k_z^2), \tag{3.3.2}$$

where k_x, k_y and k_z are the wave numbers pertaining to the various solutions of the Schrödinger equation. First, we would like to count how many quantum states $\{(l_x, l_y, l_z)\}$ give rise to energy between ϵ and $\epsilon + \mathrm{d}\epsilon$. We denote this number by $g(\epsilon)\mathrm{d}\epsilon$, where $g(\epsilon)$ is the density of states.

$$\begin{aligned} g(\epsilon)\mathrm{d}\epsilon &= \sum_{l_x,l_y,l_z} 1\left\{\frac{2m\epsilon}{\hbar^2} \le \frac{\pi^2 l_x^2}{L_x^2} + \frac{\pi^2 l_y^2}{L_y^2} + \frac{\pi^2 l_z^2}{L_z^2} \le \frac{2m(\epsilon + \mathrm{d}\epsilon)}{\hbar^2}\right\} \\ &\approx \frac{L_x L_y L_z}{\pi^3} \cdot \mathrm{Vol}\left\{\vec{\boldsymbol{k}} : \frac{2m\epsilon}{\hbar^2} \le \|\vec{\boldsymbol{k}}\|^2 \le \frac{2m(\epsilon + \mathrm{d}\epsilon)}{\hbar^2}\right\} \\ &= \frac{V}{\pi^3} \cdot \mathrm{Vol}\left\{\vec{\boldsymbol{k}} : \frac{2m\epsilon}{\hbar^2} \le \|\vec{\boldsymbol{k}}\|^2 \le \frac{2m(\epsilon + \mathrm{d}\epsilon)}{\hbar^2}\right\}. \end{aligned} \tag{3.3.3}$$

---

[5] Indeed, free electrons in a metal continue to be mobile and free even at $T = 0$.

A volume element pertaining to a given value of $K = \|\vec{\boldsymbol{k}}\|$ ($\vec{\boldsymbol{k}}$ being $k_x\hat{x} + k_y\hat{y} + k_z\hat{z}$) is given by $\mathrm{d}K$ times the surface area of sphere of radius K, namely, $4\pi K^2\mathrm{d}K$, but it has to be divided by 8, to account for the fact that the components of $\vec{\boldsymbol{k}}$ are positive. I.e., it is $\frac{\pi}{2}K^2\mathrm{d}K$. From Eq. (3.3.2), we have $\epsilon = \hbar^2K^2/2m$, and so

$$K^2\mathrm{d}K = \frac{2m\epsilon}{\hbar^2}\cdot\frac{1}{2\hbar}\sqrt{\frac{2m}{\epsilon}}\mathrm{d}\epsilon = \frac{\sqrt{2\epsilon m^3}\mathrm{d}\epsilon}{\hbar^3} \tag{3.3.4}$$

Therefore, combining the above, we get

$$g(\epsilon) = \frac{\sqrt{2\epsilon m^3}V}{2\pi^2\hbar^3}. \tag{3.3.5}$$

For electrons, spin values of ±1/2 are allowed, so this density should be doubled, and so

$$g_e(\epsilon) = \frac{\sqrt{2\epsilon m^3}V}{\pi^2\hbar^3}. \tag{3.3.6}$$

Approximating the equation of the constraint on the total number of electrons, we get

$$\begin{aligned} N_e &= \sum_r \frac{1}{e^{(\epsilon_r-\epsilon_\mathrm{F})/kT}+1} \\ &\approx \int_0^\infty \frac{g_e(\epsilon)\mathrm{d}\epsilon}{e^{(\epsilon-\epsilon_\mathrm{F})/kT}+1} \\ &= \frac{\sqrt{2m^3}V}{\pi^2\hbar^3}\cdot\int_0^\infty \frac{\sqrt{\epsilon}\mathrm{d}\epsilon}{e^{(\epsilon-\epsilon_\mathrm{F})/kT}+1} \\ &\approx \frac{\sqrt{2m^3}V}{\pi^2\hbar^3}\cdot\int_0^{\epsilon_\mathrm{F}} \sqrt{\epsilon}\mathrm{d}\epsilon \qquad T\approx 0 \\ &= \frac{\sqrt{2m^3}V}{\pi^2\hbar^3}\cdot\frac{2\epsilon_\mathrm{F}^{3/2}}{3} \end{aligned} \tag{3.3.7}$$

which easily leads to the following simple formula for the Fermi energy:

$$\epsilon_\mathrm{F} = \frac{\hbar^2}{2m}\left(\frac{3\pi^2N_e}{V}\right)^{2/3} = \frac{\hbar^2}{2m}(3\pi^2\rho_e)^{2/3}, \tag{3.3.8}$$

where ρ_e is the electron density. In most metals ϵ_F is about the order of 5–10 electron–volts (eV's), whose equivalent temperature $T_\mathrm{F} = \epsilon_\mathrm{F}/k$ (the Fermi temperature) is of the order of magnitude of 100,000 °K. Hence, the Fermi energy is much larger than kT in laboratory conditions. In other words, electrons in a metal behave like a gas at an extremely high temperature. This means that the internal pressure in metals (the Fermi pressure) is extremely large and this a reason why metals are almost incompressible. This kind of pressure also stabilizes a neutron star (a Fermi gas of neutrons) or a white dwarf star (a Fermi gas of electrons) against the inward pull of gravity, which would ostensibly collapse the star into a Black Hole. Only when a star is sufficiently massive to overcome the degeneracy pressure can it collapse into a singularity.

Exercise 3.2 Derive an expression for $\langle \epsilon^n \rangle$ of an electron near $T = 0$, in terms of ϵ_F.

3.4 Useful Approximations of Fermi Integrals

Before considering applications, it will be instructive to develop some useful approximations for integrals associated with the Fermi function

$$f(\epsilon) \triangleq \frac{1}{e^{(\epsilon-\mu)/kT}+1}. \tag{3.4.1}$$

For example, if we wish to calculate the average energy, we have to deal with an integral like

$$\int_0^\infty \epsilon^{3/2} f(\epsilon)\mathrm{d}\epsilon.$$

Consider then, more generally, an integral of the form

$$I_n = \int_0^\infty \epsilon^n f(\epsilon)\mathrm{d}\epsilon.$$

Upon integrating by parts, we readily have

$$\begin{aligned} I_n &= f(\epsilon)\cdot\frac{\epsilon^{n+1}}{n+1}\bigg|_0^\infty - \frac{1}{n+1}\int_0^\infty \epsilon^{n+1} f'(\epsilon)\mathrm{d}\epsilon \\ &= -\frac{1}{n+1}\int_0^\infty \epsilon^{n+1} f'(\epsilon)\mathrm{d}\epsilon \end{aligned} \tag{3.4.2}$$

Changing variables to $x = (\epsilon - \mu)/kT$,

$$\begin{aligned} I_n &= -\frac{1}{n+1}\int_{-\mu/kT}^\infty (\mu + kTx)^{n+1}\phi'(x)\mathrm{d}x \\ &\approx -\frac{\mu^{n+1}}{n+1}\int_{-\infty}^\infty \left(1+\frac{kTx}{\mu}\right)^{n+1}\phi'(x)\mathrm{d}x, \end{aligned} \tag{3.4.3}$$

where we have introduced the scaled version of f, which is $\phi(x) = f(\mu + kTx) = 1/(e^x + 1)$ and ϕ' is its derivative, and where in the second line we are assuming $\mu \gg kT$. Applying the Taylor series expansion to the binomial term (recall that n is not necessarily an integer), and using the symmetry of ϕ' around the origin, we have

$$\begin{aligned} I_n &= -\frac{\mu^{n+1}}{n+1}\int_{-\infty}^{\infty}\left[1 + (n+1)\frac{kTx}{\mu} + \frac{n(n+1)}{2}\left(\frac{kTx}{\mu}\right)^2 + \cdots\right]\phi'(x)\mathrm{d}x \\ &= -\frac{\mu^{n+1}}{n+1}\left[\int_{-\infty}^{\infty}\phi'(x)\mathrm{d}x + \frac{n(n+1)}{2}\left(\frac{kT}{\mu}\right)^2\int_{-\infty}^{\infty}x^2\phi'(x)\mathrm{d}x + \cdots\right] \\ &\approx \frac{\mu^{n+1}}{n+1}\left[1 + \frac{n(n+1)\pi^2}{6}\left(\frac{kT}{\mu}\right)^2\right] \end{aligned} \tag{3.4.4}$$

where the last line was obtained by calculating the integral of $x^2\phi'(x)$ using a power series expansion. Note that this series contains only even powers of kT/μ, thus the convergence is rather fast. Let us now repeat the calculation of Eq. (3.3.7), this time at $T > 0$.

$$\begin{aligned} \rho_e &\approx \frac{\sqrt{2m^3}}{\pi^2\hbar^3}\cdot\int_0^{\infty}\frac{\sqrt{\epsilon}\mathrm{d}\epsilon}{e^{(\epsilon-\mu)/kT} + 1} \\ &= \frac{\sqrt{2m^3}}{\pi^2\hbar^3}\cdot I_{1/2} \\ &= \frac{\sqrt{2m^3}}{\pi^2\hbar^3}\cdot\frac{2}{3}\mu^{3/2}\left[1 + \frac{\pi^2}{8}\left(\frac{kT}{\mu}\right)^2\right] \end{aligned} \tag{3.4.5}$$

which gives

$$\begin{aligned} \epsilon_{\mathrm{F}} &= \mu\left[1 + \frac{\pi^2}{8}\left(\frac{kT}{\mu}\right)^2\right]^{2/3} \\ &\approx \mu\left[1 + \frac{\pi^2}{12}\left(\frac{kT}{\mu}\right)^2\right] \\ &= \mu + \frac{(\pi kT)^2}{12\mu}. \end{aligned} \tag{3.4.6}$$

This relation between μ and ϵ_{F} can be easily inverted by solving a simple quadratic equation, which yields

$$\begin{aligned}\mu &\approx \frac{\epsilon_F + \epsilon_F\sqrt{1-(\pi kT/\epsilon_F)^2/3}}{2}\\ &\approx \epsilon_F\left[1-\frac{\pi^2}{12}\cdot\left(\frac{kT}{\epsilon_F}\right)^2\right]\\ &= \epsilon_F\left[1-\frac{\pi^2}{12}\cdot\left(\frac{T}{T_F}\right)^2\right]. \end{aligned} \tag{3.4.7}$$

Since $T/T_F \ll 1$ for all T in the interesting range, we observe that the chemical potential depends extremely weakly on T. In other words, we can safely approximate $\mu \approx \epsilon_F$ for all relevant temperatures of interest. The assumption that $kT \ll \mu$ was found self–consistent with the result $\mu \approx \epsilon_F$.

Having established the approximation $\mu \approx \epsilon_F$, we can now calculate the average energy of the electron at an arbitrary temperature T:

$$\begin{aligned}\langle\epsilon\rangle &= \frac{\sqrt{2m^3}}{\pi^2\hbar^3\rho_e}\int_0^\infty \frac{\epsilon^{3/2}\mathrm{d}\epsilon}{e^{(\epsilon-\epsilon_F)/kT}+1}\\ &= \frac{\sqrt{2m^3}}{\pi^2\hbar^3\rho_e}\cdot I_{3/2}\\ &\approx \frac{\sqrt{2m^3}}{\pi^2\hbar^3\rho_e}\cdot\frac{2\epsilon_F^{5/2}}{5}\left[1+\frac{5\pi^2}{8}\left(\frac{T}{T_F}\right)^2\right]\\ &= \frac{3\hbar^2}{10m}\cdot(3\pi^2\rho_e)^{2/3}\cdot\left[1+\frac{5\pi^2}{8}\left(\frac{T}{T_F}\right)^2\right]\\ &= \frac{3\epsilon_F}{5}\cdot\left[1+\frac{5\pi^2}{8}\left(\frac{T}{T_F}\right)^2\right] \end{aligned} \tag{3.4.8}$$

Note that the dependence of the average per–particle energy on the temperature is drastically different from that of the ideal gas. While in the idea gas it was linear ($\langle\epsilon\rangle = 3kT/2$), here it is actually almost a constant, independent of the temperature (just like the chemical potential).

The same technique can be used, of course, to calculate any moment of the electron energy.

3.5 Applications of the FD Distribution

The FD distribution is at the heart of modern solid–state physics and semiconductor physics (see also, for example, [1, Sect. 4.5]) and indeed frequently encountered in related courses on semiconductor devices. It is also useful for understanding the physics of white dwarfs. We next briefly touch upon the very basics of conductance in solids, as well as on two other applications: thermionic emission and photoelectric emission.

3.5.1 Electrons in a Solid

The structure of the electron energy levels in a solid are basically obtained using quantum–mechanical considerations. In the case of a crystal, this amounts to solving the Schrödinger equation in a periodic potential, stemming from the corresponding periodic lattice structure. Its idealized form, which ignores the size of each atom, is given by a train of equispaced Dirac delta functions. This is an extreme case of the so called *Kronig–Penney model*, where the potential function is a periodic rectangular on–off function (square wave function), and it leads to a certain band structure. In particular, bands of allowed energy levels are alternately interlaced with bands of forbidden energy levels. The Fermi energy level ϵ_{F}, which depends on the overall concentration of electrons, may either fall in an allowed band or in a forbidden band. The former case is the case of a metal, whereas the latter case is the case of an insulator or a semiconductor (the difference being only how wide is the forbidden band in which ϵ_{F} lies). While in metals it is impossible to change ϵ_{F}, it is possible by doping in semiconductors.

A semiconductor can then be thought of as a system with electron orbitals grouped into two[6] energy bands separated by an energy gap. The lower band is the *valence band* (where electrons are tied to their individual atoms) and the upper band is the *conduction band*, where they are free. In a pure semiconductor at $T = 0$, all valence orbitals are occupied with electrons and all conduction orbitals are empty. A full band cannot carry any current so a pure semiconductor at $T = 0$ is an insulator. In a pure semiconductor the Fermi energy is exactly in the middle of the gap between the valence band (where $f(\epsilon)$ is very close 1) and the conduction band (where $f(\epsilon)$ is very close to 0). Finite conductivity in a semiconductor follows either from the presence of electrons in the conduction band (conduction electrons) or from unoccupied orbitals in the valence band (holes).

Two different mechanisms give rise to conduction electrons and holes: the first is thermal excitation of electrons from the valence band to the conduction band, while the second is the presence of impurities that change the balance between the number of orbitals in the valence band and the number of electrons available to fill them.

We will not delve into this too much beyond this point, since this material is normally well–covered in other courses in the standard curriculum of electrical engineering, namely, courses on solid state physics. Here we only demonstrate the use of the FD distribution in order to calculate the density of charge carriers. The density of charge carriers n of the conduction band is found by integrating up, from the conduc-

---

[6] We treat both bands as single bands for our purposes. It does not matter that both may be themselves groups of (sub)bands with additional gaps within each group.

tion band edge ϵ_C, the product of the density of states $g_e(\epsilon)$ and the FD distribution $f(\epsilon)$, i.e.,

$$n = \int_{\epsilon_C}^{\infty} \mathrm{d}\epsilon \cdot g_e(\epsilon) f(\epsilon) = \frac{\sqrt{2m^3}}{\pi^2 \hbar^3} \int_{\epsilon_C}^{\infty} \frac{\sqrt{\epsilon - \epsilon_C}\mathrm{d}\epsilon}{e^{(\epsilon-\epsilon_F)/kT} + 1}, \tag{3.5.1}$$

where here m designates the *effective mass* of the electron[7] and where we have taken the density of states to be proportional to $\sqrt{\epsilon - \epsilon_C}$ since ϵ_C is now the reference energy and only the difference $\epsilon - \epsilon_C$ goes for kinetic energy.[8] For a semiconductor at room temperature, kT is much smaller than the gap, and so

$$f(\epsilon) \approx e^{-(\epsilon-\epsilon_F)/kT} \tag{3.5.2}$$

which yields the approximation

$$\begin{aligned} n &\approx \frac{\sqrt{2m^3}}{\pi^2 \hbar^3} \cdot e^{\epsilon_F/kT} \int_{\epsilon_C}^{\infty} \mathrm{d}\epsilon \cdot \sqrt{\epsilon - \epsilon_C} \cdot e^{-\epsilon/kT} \\ &= \frac{\sqrt{2m^3}}{\pi^2 \hbar^3} \cdot e^{-(\epsilon_C-\epsilon_F)/kT} \int_0^{\infty} \mathrm{d}\epsilon \cdot \sqrt{\epsilon} e^{-\epsilon/kT} \\ &= \frac{\sqrt{2(mkT)^3}}{\pi^2 \hbar^3} \cdot e^{-(\epsilon_C-\epsilon_F)/kT} \int_0^{\infty} \mathrm{d}x \cdot \sqrt{x} e^{-x} \\ &= \frac{\sqrt{\pi}}{2} \cdot \frac{\sqrt{2(mkT)^3}}{\pi^2 \hbar^3} \cdot e^{-(\epsilon_C-\epsilon_F)/kT} \\ &= \frac{1}{4} \cdot \left(\frac{2mkT}{\pi\hbar^2}\right)^{3/2} \cdot e^{-(\epsilon_C-\epsilon_F)/kT}. \end{aligned} \tag{3.5.3}$$

We see then that the density of conduction electrons, and hence also the conduction properties, depend critically on the gap between ϵ_C and ϵ_F. A similar calculation holds for the holes, of course.

*3.5.2 Thermionic Emission**

Thermionic emission is a current of charge carriers (most notably, electrons or ions) via a surface (which acts as a potential barrier), caused by heat energy that overcomes the electrostatic forces. If a refractory metal (e.g., tungsten) is heated up to high

---

[7] The effective mass is obtained by a second order Taylor series expansion of the energy as a function of the wave-number (used to obtain the density of states), and thinking of the coefficient of the quadratic term as $\hbar^2/2m$.

[8] Recall that earlier we calculated the density of states for a simple potential well, not for a periodic potential function. Thus, the earlier expression of $g_e(\epsilon)$ is not correct here.

enough temperature (somewhat below the melting point), an emission of electrons can be obtained towards a positive anode. Quantum–mechanically speaking, this heated metal can be viewed as a potential well with finitely high walls determined by the surface potential barrier. Thus, some of the particles will be sufficiently energetic to surmount the surface barrier (a.k.a. the surface work function) and hence will be emitted. The work function ϕ varies between 2 and 6 eV for pure metals. The electron will not be emitted unless the energy component normal to the surface would exceed $\epsilon_{\mathrm{F}} + \phi$. The excess energy beyond this threshold is in the form of translational kinetic energy which dictates the velocity away from the surface.

The analysis of this effect is made by transforming the distribution of energy into a distribution in terms of the components of the velocity, v_x, v_y, and v_z. We begin with the expression of the energy of a single electron[9]

$$\epsilon = \frac{1}{2}m(v_x^2 + v_y^2 + v_z^2) = \frac{\pi^2\hbar^2}{2m}\left(\frac{l_x^2}{L_x^2} + \frac{l_y^2}{L_y^2} + \frac{l_z^2}{L_z^2}\right). \tag{3.5.4}$$

Thus, $\mathrm{d}v_x = h\mathrm{d}l_x/(2mL_x)$ and similar relations hold for the two other components, which together yield

$$\mathrm{d}l_x\mathrm{d}l_y\mathrm{d}l_z = \left(\frac{m}{h}\right)^3 V\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z, \tag{3.5.5}$$

where we have divided by 8 since every quantum state can be occupied by only one out of 8 combinations of the signs of the three component velocities. The fraction of electrons $\mathrm{d}N$ with quantum states within the cube $\mathrm{d}l_x\mathrm{d}l_y\mathrm{d}l_z$ is simply the expected number of occupied quantum states within that cube, which is

$$\mathrm{d}N = \frac{\mathrm{d}l_x\mathrm{d}l_y\mathrm{d}l_z}{1 + \exp\{(\epsilon_{l_xl_yl_z} - \epsilon)/kT\}}.$$

Thus, we can write the distribution function of the number of electrons in a cube $\mathrm{d}v_x \times \mathrm{d}v_y \times \mathrm{d}v_z$ as

$$\mathrm{d}N = 2V\left(\frac{m}{h}\right)^3 \frac{\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z}{1 + \exp\left\{\frac{1}{kT}\left[\frac{1}{2}m(v_x^2 + v_y^2 + v_z^2) - \epsilon_{\mathrm{F}}\right]\right\}}, \tag{3.5.6}$$

where we have doubled the expression due to the spin and we have taken the chemical potential of the electron gas to be ϵ_{F}, independently of temperature, as was justified in the previous subsection. Assuming that the surface is parallel to the YZ plane, the

---

[9] We are assuming that the potential barrier ϕ is fairly large (relative to kT), such that the relationship between energy and quantum numbers is reasonably well approximated by that of a particle in a box.

minimum escape velocity in the x–direction is $v_0 = \sqrt{\frac{2}{m}(\epsilon_F + \phi)}$ and there are no restrictions on v_y and v_z. The current along the x–direction is

$$\begin{aligned} I &= \frac{\mathrm{d}q}{\mathrm{d}t} = \frac{q_e \mathrm{d}N[\text{leaving the surface}]}{\mathrm{d}t} \\ &= \frac{q_e}{\mathrm{d}t} \int_{v_0}^{\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{v_x \mathrm{d}t}{L_x} \cdot 2V \left(\frac{m}{h}\right)^3 \frac{\mathrm{d}v_x \mathrm{d}v_y \mathrm{d}v_z}{1 + \exp\left\{\frac{1}{kT}\left[\frac{1}{2}m(v_x^2 + v_y^2 + v_z^2) - \epsilon_F\right]\right\}} \\ &= 2L_y L_z q_e \left(\frac{m}{h}\right)^3 \int_{v_0}^{\infty} v_x \mathrm{d}v_x \int_{-\infty}^{+\infty} \frac{\mathrm{d}v_y \mathrm{d}v_z}{1 + \exp\left\{\frac{1}{kT}\left[\frac{1}{2}m(v_x^2 + v_y^2 + v_z^2) - \epsilon_F\right]\right\}}, \end{aligned} \tag{3.5.7}$$

where the factor $v_x \mathrm{d}t/L_x$ in the second line is the fraction of electrons close enough to the surface so as to be emitted within time $\mathrm{d}t$. Thus, the current density (current per unity area) is

$$J = 2q_e \left(\frac{m}{h}\right)^3 \int_{v_0}^{\infty} \mathrm{d}v_x \cdot v_x \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\mathrm{d}v_y \mathrm{d}v_z}{1 + \exp\left\{\frac{1}{kT}\left[\frac{1}{2}m(v_x^2 + v_y^2 + v_z^2) - \epsilon_F\right]\right\}}. \tag{3.5.8}$$

As for the inner double integral, transform to polar coordinates to obtain

$$\begin{aligned} &\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\mathrm{d}v_y \mathrm{d}v_z}{1 + \exp\left\{\frac{1}{kT}\left[\frac{1}{2}m(v_x^2 + v_y^2 + v_z^2) - \epsilon_F\right]\right\}} \\ &= 2\pi \int_0^{\infty} \frac{v_{yz} \mathrm{d}v_{yz}}{1 + e^{mv_{yz}^2/2kT} \cdot \exp\left[\frac{1}{kT}\left(\frac{1}{2}mv_x^2 - \epsilon_F\right)\right]} \\ &= \frac{2\pi kT}{m} \int_0^{\infty} \frac{\mathrm{d}u}{1 + \exp\left[\frac{1}{kT}\left(\frac{1}{2}mv_x^2 - \epsilon_F\right)\right] \cdot e^u} \qquad u = mv_{yz}^2/2kT \\ &= \frac{2\pi kT}{m} \ln\left\{1 + \exp\left[\frac{1}{kT}\left(\epsilon_F - \frac{1}{2}mv_x^2\right)\right]\right\} \end{aligned} \tag{3.5.9}$$

which yields

$$J = \frac{4\pi m^2 q_e kT}{h^3} \int_{v_0}^{\infty} \mathrm{d}v_x \cdot v_x \ln\left\{1 + \exp\left[\frac{1}{kT}\left(\epsilon_F - \frac{1}{2}mv_x^2\right)\right]\right\}. \tag{3.5.10}$$

Now, since normally[10] $\phi \gg kT$, the exponent in the integrand is very small throughout the entire range of integration and so, it is safe to approximate it by $\ln(1+x) \approx x$, i.e.,

$$\begin{aligned} J &\approx \frac{4\pi m^2 q_e kT}{h^3} e^{\epsilon_F/kT} \int_{v_0}^{\infty} \mathrm{d}v_x \cdot v_x e^{-mv_x^2/2kT} \\ &= \frac{4\pi m q_e (kT)^2}{h^3} \exp\left\{ \frac{1}{kT} \left(\epsilon_F - \frac{1}{2} m v_0^2 \right) \right\} \\ &= \frac{4\pi m q_e (kT)^2}{h^3} e^{-\phi/kT}, \end{aligned} \tag{3.5.11}$$

and thus we have obtained a simple expression for the current density as function of temperature. This result, which is known as the *Richardson–Dushman equation*, is in very good agreement with experimental evidence. Further discussion on this result can be found in [1, 2].

3.5.3 Photoelectric Emission*

An analysis based on a similar line of thought applies also to the photoelectric emission, an effect where electrons are emitted from a metal as a result of radiation at frequency beyond a certain critical frequency ν_0 (the Einstein threshold frequency), whose corresponding photon energy $h\nu_0$ is equal to the work function ϕ. Here, the electron gains an energy amount of $h\nu$ from a photon, which helps to pass the energy barrier, and so the minimum velocity of emission, after excitation by a photon of energy $h\nu$ is given by

$$h\nu + \frac{1}{2} m v_0^2 = \epsilon_F + \phi = \epsilon_F + h\nu_0. \tag{3.5.12}$$

Let α denote the probability that a photon actually excites an electron. Then, similarly as in the previous subsection,

$$J = \alpha \cdot \frac{4\pi m^2 q_e kT}{h^3} \int_{v_0}^{\infty} \mathrm{d}v_x \cdot v_x \ln\left\{ 1 + \exp\left[\frac{1}{kT} \left(\epsilon_F - \frac{1}{2} m v_x^2 \right) \right] \right\}. \tag{3.5.13}$$

where this time

$$v_0 = \sqrt{\frac{2}{m} [\epsilon_F + h(\nu_0 - \nu)]}. \tag{3.5.14}$$

---

[10] At room temperature ($T = 300\,^\circ$K), $kT \approx 4 \times 10^{-21}$ Joules ≈ 0.024eV, whereas ϕ is between 2eV and 6eV.

Changing the integration variable to

$$x = \frac{1}{kT}\left[\frac{1}{2}mv_x^2 + h(\nu - \nu_0) - \epsilon_{\mathrm{F}}\right],$$

we can write the last integral as

$$J = \alpha \cdot \frac{4\pi m q_e (kT)^2}{h^3}\int_0^\infty \mathrm{d}x \ln\left\{1 + \exp\left[\frac{h(\nu - \nu_0)}{kT} - x\right]\right\}\mathrm{d}x. \tag{3.5.15}$$

Now, let us denote

$$\Delta = \frac{h(\nu - \nu_0)}{kT}. \tag{3.5.16}$$

Integrating by parts (twice), we have

$$\begin{aligned}\int_0^\infty \mathrm{d}x \ln(1 + e^{\Delta - x}) &= \int_0^\infty \frac{x\mathrm{d}x}{e^{x-\Delta} + 1}\\ &= \frac{1}{2}\int_0^\infty \frac{x^2 e^{x-\Delta}\mathrm{d}x}{(e^{x-\Delta} + 1)^2} \triangleq f(e^\Delta).\end{aligned} \tag{3.5.17}$$

For $h(\nu - \nu_0) \gg kT$, we have $e^\Delta \gg 1$, and then it can be shown (using the same technique as in Sect. 3.4) that $f(e^\Delta) \approx \Delta^2/2$, which gives

$$J = \alpha \cdot \frac{2\pi m q_e}{h}(\nu - \nu_0)^2 \tag{3.5.18}$$

independently of T. In other words, when the energy of light quantum is much larger than the thermal energy kT, temperature becomes irrelevant. At the other extreme of very low frequency, where $h(\nu_0 - \nu) \gg kT$, and then $e^\Delta \ll 1$, we have $f(e^\Delta) \approx e^\Delta$, and then

$$J = \alpha \cdot \frac{4\pi m q_e (kT)^2}{h^3} e^{(h\nu - \phi)/kT} \tag{3.5.19}$$

which is like the thermionic current density, enhanced by a photon factor $e^{h\nu/kT}$.

3.6 Suggestions for Supplementary Reading

The Fermi–Dirac distribution, its derivation, and its various applications can also be found in many alternative textbooks, such as: Beck [1, Chap. 4], Huang [3, Chap. 11], Kittel [4, Part I, Chap. 19], Landau and Lifshitz [5, Chap. V], Mandl [6, Sect. 11.4.1], Pathria [2], and Reif [7, Chap. 9]. The exposition in this chapter is based, to a large extent, on the books by Beck, Mandl and Pathria. Applications to semiconductor physics are based also on Omar [8, Chaps. 6, 7] and Gershenfeld [9].

References

1. A.H.W. Beck, *Statistical Mechanics, Fluctuations and Noise* (Edward Arnold Publishers, London, 1976)
2. R.K. Pathria, *Statistical Mechanics*, 2nd edn. (Elsevier: Butterworth-Heinemann, Oxford, 1996)
3. K. Huang, *Statistical Mechanics*, 2nd edn. (Wiley, New York, 1987)
4. C. Kittel, *Elementary Statistical Physics* (Wiley, New York, 1958)
5. L.D. Landau, E.M. Lifshitz, *Course of Theoretical Physics – Volume 5: Statistical Physics, Part 1*, 3rd edn. Elsevier: Butterworth–Heinemann, New York (1980)
6. F. Mandl, *Statistical Physics* (Wiley, Chichester, 1971)
7. F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1965)
8. M.A. Omar, *Elementary Solid State Physics: Principles and Applications* (Addison Wesley, Reading, 1975)
9. N. Gershenfeld, *The Physics of Information Technology* (Cambridge University Press, Cambridge, 2000)

Chapter 4
Quantum Statistics – The Bose–Einstein Distribution

The general description of bosons was provided in the introductory paragraphs of Chap. 3. As said, the crucial difference between bosons and fermions is that in the case of bosons, Pauli's exclusion principle does not apply. In this chapter, we study the statistical mechanics of bosons.

4.1 Combinatorial Derivation of the BE Statistics

Using the same notation as in Chap. 3, again, we are partitioning the energy levels $\epsilon_1, \epsilon_2, \ldots$ into groups, labeled by s, where in group no. s, which has G_s quantum states, the representative energy is $\hat{\epsilon}_s$. As before, a microstate is defined in terms of $\{\hat{N}_s\}$ and $\Omega(\hat{N}_1, \hat{N}_2, \ldots) = \prod_s \Omega_s(\hat{N}_s)$, but now we need a different estimate of each factor $\Omega_s(\hat{N}_s)$, since now there are no restrictions on the occupation numbers of the quantum states. In how many ways can one partition $\hat{N}_s$ particles among G_s states? Imagine that the $\hat{N}_s$ particles of group no. s are arranged along a line. By means of $G_s - 1$ partitions we divide the particles into G_s subsets corresponding to the various states in that group. We have a total of $(\hat{N}_s + G_s - 1)$ elements, $\hat{N}_s$ of them are particles and the remaining $(G_s - 1)$ are partitions (see Fig. 4.1). In how many distinct ways can we configure them? The answer is simple:

$$\Omega_s(\hat{N}_s) = \frac{(\hat{N}_s + G_s - 1)!}{\hat{N}_s!(G_s - 1)!}. \tag{4.1.1}$$

On the account that $G_s \gg 1$, the -1 term can be safely neglected, and we approximate

$$\Omega_s(\hat{N}_s) = \frac{(\hat{N}_s + G_s)!}{\hat{N}_s!G_s!}. \tag{4.1.2}$$

N. Merhav, *Statistical Physics for Electrical Engineering*,
DOI 10.1007/978-3-319-62063-3_4

Fig. 4.1 $\hat{N}_s$ particles and $G_s - 1$ partitions

Repeating the same derivation as in Sect. 4.1, but with the above $\Omega_s(\hat{N}_s)$, we get:

$$\ln \Omega_s(\hat{N}_s) \approx (\hat{N}_s + G_s) h_2\left(\frac{G_s}{\hat{N}_s + G_s}\right), \tag{4.1.3}$$

and so the free energy is now

$$F \approx \sum_s \left[\hat{N}_s \hat{\epsilon}_s - kT(\hat{N}_s + G_s) h_2\left(\frac{G_s}{\hat{N}_s + G_s}\right)\right], \tag{4.1.4}$$

which should be minimized s.t. $\sum_s \hat{N}_s = N$. Upon carrying out the minimization of the corresponding Lagrangian, we arrive[1] at the following result for the most probable occupation numbers:

$$\hat{N}_s = \frac{G_s}{e^{\beta(\hat{\epsilon}_s - \mu)} - 1} \tag{4.1.5}$$

or, moving back to the original occupation numbers,

$$\bar{N}_r = \frac{1}{e^{\beta(\epsilon_r - \mu)} - 1}, \tag{4.1.6}$$

where μ is again the Lagrange multiplier, which has the meaning of the chemical potential. This is Bose–Einstein (BE) distribution. As we see, the formula is very similar to that of the FD distribution, the only difference is that in the denominator, $+1$ is replaced by -1. Surprisingly enough, this is a crucial difference that makes the behavior of bosons drastically different from that of fermions. Note that for this expression to make sense, μ must be smaller than the ground energy ϵ_1, otherwise the denominator either vanishes or becomes negative. If the ground–state energy is zero, this means $\mu < 0$.

4.2 Derivation Using the Grand–Canonical Ensemble

As in Sect. 4.2, an alternative derivation can be carried out using the grand–canonical ensemble. The only difference is that now, the summations over $\{N_r\}$, are not only over $\{0, 1\}$ but over all non–negative integers. In particular,

---

[1] **Exercise 4.1** Fill in the detailed derivation.

$$\Xi(\beta,\mu)=\prod_r\left[\sum_{N_r=0}^{\infty}e^{\beta N_r(\mu-\epsilon_r)}\right]. \tag{4.2.1}$$

Of course, here too, for convergence of each geometric series, we must assume $\mu<\epsilon_1$, and then the result is

$$\Xi(\beta,\mu)=\prod_r\frac{1}{1-e^{\beta(\mu-\epsilon_r)}}. \tag{4.2.2}$$

Here, under the grand–canonical ensemble, $N_1, N_2, \ldots$ are independent geometric random variables with distributions

$$P_r(N_r)=[1-e^{\beta(\mu-\epsilon_r)}]e^{\beta N_r(\mu-\epsilon_r)} \quad N_r=0,1,2,\ldots, \quad r=1,2,\ldots \tag{4.2.3}$$

Thus, $\bar{N}_r$ is just the expectation of this geometric random variable, which is readily found[2] to be as in Eq. (4.1.6).

4.3 Bose–Einstein Condensation

In analogy to the FD case, here too, the chemical potential μ is determined from the constraint on the total number of particles. In this case, it reads

$$\sum_r\frac{1}{e^{\beta(\epsilon_r-\mu)}-1}=N. \tag{4.3.1}$$

Taking into account the density of states in a potential well of sizes $L_x\times L_y\times L_z$ (as was done in Chap. 3), in the continuous limit, this yields

$$\rho=\frac{\sqrt{2m^3}}{2\pi^2\hbar^3}\cdot\int_0^{\infty}\frac{\sqrt{\epsilon}\mathrm{d}\epsilon}{e^{(\epsilon-\mu)/kT}-1}. \tag{4.3.2}$$

At this point, an important peculiarity should be discussed. Consider Eq. (4.3.2) and suppose that we are cooling the system. As T decreases, μ must adjust in order to keep Eq. (4.3.2) holding since the number of particles must be preserved. In particular, as T decreases, μ must increase, yet it must be negative. The point is that even for $\mu=0$, which is the maximum allowed value of μ, the integral at the r.h.s. of (4.3.2) is finite[3] as the density of states is proportional to $\sqrt{\epsilon}$ and hence balances the divergence of the BE integral near $\epsilon=0$. Let us define then

---

[2] **Exercise 4.2** Show this.

[3] **Exercise 4.3** Show this.

$$\varrho(T) \stackrel{\Delta}{=} \frac{\sqrt{2m^3}}{2\pi^2\hbar^3} \cdot \int_0^\infty \frac{\sqrt{\epsilon}\mathrm{d}\epsilon}{e^{\epsilon/kT} - 1} \tag{4.3.3}$$

and let T_c be the solution to the equation $\varrho(T) = \rho$, which can be found as follows. By changing the integration variable to $z = \epsilon/kT$, we can rewrite the r.h.s. as

$$\varrho(T) = \left(\frac{mkT}{2\pi\hbar^2}\right)^{3/2} \left\{\frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\sqrt{z}\mathrm{d}z}{e^z - 1}\right\} \approx 2.612 \cdot \left(\frac{mkT}{2\pi\hbar^2}\right)^{3/2}, \tag{4.3.4}$$

where the constant 2.612 is the numerical value of the expression in the curly brackets. Thus,

$$T_c \approx 0.5274 \cdot \frac{2\pi\hbar^2}{mk} \cdot \rho^{2/3} = 3.313 \cdot \frac{\hbar^2\rho^{2/3}}{mk}. \tag{4.3.5}$$

The problem is that for $T < T_c$, Eq. (4.3.2) can no longer be solved by any non–positive value of μ. So what happens below T_c?

The root of the problem is in the passage from the discrete sum over r to the integral over ϵ. The paradox is resolved when it is understood that below T_c, the contribution of $\epsilon = 0$ should be separated from the integral. That is, the correct form is

$$N = \frac{1}{e^{-\mu/kT} - 1} + \frac{\sqrt{2m^3}V}{2\pi^2\hbar^3} \cdot \int_0^\infty \frac{\sqrt{\epsilon}\mathrm{d}\epsilon}{e^{(\epsilon-\mu)/kT} - 1}. \tag{4.3.6}$$

or, after dividing by V,

$$\rho = \rho_0 + \frac{\sqrt{2m^3}}{2\pi^2\hbar^3} \cdot \int_0^\infty \frac{\sqrt{\epsilon}\mathrm{d}\epsilon}{e^{(\epsilon-\mu)/kT} - 1}, \tag{4.3.7}$$

where ρ_0 is the density of ground–state particles, and now the integral accommodates the contribution of all particles with strictly positive energy. Now, for $T < T_c$, we simply have $\rho_0 = \rho - \varrho(T)$, which means that a *macroscopic* fraction of the particles condensate at the ground state. This phenomenon is called *Bose–Einstein condensation*. Note that for $T < T_c$,

$$\begin{aligned}
\rho_0 &= \rho - \varrho(T) \\
&= \varrho(T_c) - \varrho(T) \\
&= \varrho(T_c)\left[1 - \frac{\varrho(T)}{\varrho(T_c)}\right] \\
&= \varrho(T_c)\left[1 - \left(\frac{T}{T_c}\right)^{3/2}\right]
\end{aligned}$$

$$= \rho \left[1 - \left(\frac{T}{T_c} \right)^{3/2} \right] \tag{4.3.8}$$

which gives a precise characterization of the condensation as a function of temperature. It should be pointed out that T_c is normally extremely low.[4]

One might ask why does the point $\epsilon = 0$ require special caution when $T < T_c$, but does not require such caution for $T > T_c$? The answer is that for $T > T_c$, $\rho_0 = 1/V[e^{-\mu/kT} - 1]$ tends to zero in the thermodynamic limit ($V \to \infty$) since $\mu < 0$. However, as $T \to T_c$, $\mu \to 0$, and ρ_0 becomes singular.

It is instructive to derive the pressure exerted by the ideal Boson gas for $T < T_c$. This can be obtained from the grand partition function

$$\begin{aligned} \ln \Xi &= -\sum_r \ln(1 - e^{-\epsilon_r/kT}) \qquad\qquad (\mu = 0) \\ &\sim -\frac{\sqrt{2m^3}V}{2\pi^2\hbar^3} \int_0^\infty \mathrm{d}\epsilon \cdot \sqrt{\epsilon} \ln(1 - e^{-\epsilon/kT}) \\ &= -\frac{\sqrt{2m^3}(kT)^{3/2}V}{2\pi^2\hbar^3} \int_0^\infty \mathrm{d}x \cdot \sqrt{x} \ln(1 - e^{-x}), \end{aligned} \tag{4.3.9}$$

where integral over x (including the minus sign) is just a positive constant C that we will not calculate here. Now,

$$P = \lim_{V \to \infty} \frac{kT \ln \Xi}{V} = \frac{C\sqrt{2m^3}(kT)^{5/2}}{2\pi^2\hbar^3}. \tag{4.3.10}$$

We see that the pressure is independent of the density ρ (compare with the ideal gas where $P = \rho kT$). This is because the condensed particles do not contribute to the pressure. What matters is only the density of those with positive energy, and this density in turn depends only on T.

Exercise 4.4 Why fermions do not condensate? What changes in the last derivation?

Exercise 4.5 The last derivation was in three dimensions ($d = 3$). Modify the derivation of the BE statistics to apply to a general dimension d, taking into account the dependence of the density of states upon d. For which values of d bosons condensate?

---

[4] In 1995 the first gaseous condensate was produced by Eric Cornell and Carl Wieman at the University of Colorado, using a gas of rubidium atoms cooled to 170 nanokelvin. For their achievements Cornell, Wieman, and Wolfgang Ketterle of MIT received the 2001 Nobel Prize in Physics. In November 2010 the first photon BEC was observed.

4.4 Black–Body Radiation

A *black body* is an (idealized model of an) object that absorbs all the incident electromagnetic radiation (and reflects none), regardless of the wavelength. A black body in thermal equilibrium emits radiation that is called *black–body radiation*. It should be understood that all bodies emit electromagnetic radiation whenever at positive temperature, but normally, this radiation is not in thermal equilibrium. One of the important applications of the BE statistics is to investigate the equilibrium properties of black–body radiation.

If we consider the radiation inside an opaque object whose surfaces and walls are kept at fixed temperature T, then the radiation and the surfaces arrive at thermal equilibrium and then, the radiation has properties that are appreciably close to those of a black body. To study the behavior of such a radiation, one creates a tiny hole in the surface of the enclosure (so that a photon entering the cavity will be 'trapped' within internal reflections, but will never be reflected out) it will not disturb the equilibrium of the cavity and then the emitted radiation will have the same properties as the cavity radiation, which in turn are the same as the radiation properties of a black body. The temperature of the black body is T as well, of course. In this section, we study these radiation properties using BE statistics.

We consider a radiation cavity of volume V and temperature T. Historically, Planck (1900) viewed this system as an assembly of harmonic oscillators with quantized energies $(n + 1/2)\hbar\omega$, $n = 0, 1, 2, \ldots$, where ω is the angular frequency of the oscillator. An alternative point of view is as an ideal gas of identical and indistinguishable photons, each one with energy $\hbar\omega$. Photons have integral spin and hence are bosons, but they have zero mass and zero chemical potential when they interact with a black–body. The reason is that there is no constraint that their total number would be conserved (they are emitted and absorbed in the black–body material with which they interact). Since in equilibrium F should be minimum, then $(\partial F/\partial N)_{T,V} = 0$. But $(\partial F/\partial N)_{T,V} = \mu$, and so, $\mu = 0$. It follows then that distribution of photons across the quantum states obeys BE statistics with $\mu = 0$, that is

$$\bar{N}_\omega = \frac{1}{e^{\hbar\omega/kT} - 1}. \tag{4.4.1}$$

The calculation of the density of states here is somewhat different from the one in Sect. 4.3. Earlier, we considered a particle with positive mass m, whose kinetic energy is $\|\vec{\boldsymbol{p}}\|^2/2m = \hbar^2\|\vec{\boldsymbol{k}}\|^2/2m$, whereas now we are talking about a photon whose rest mass is zero and whose energy is $\hbar\omega = \hbar\|\vec{\boldsymbol{k}}\|c = \|\vec{\boldsymbol{p}}\|c$ (c being the speed of light), so the dependence on $\|\vec{\boldsymbol{k}}\|$ is now linear rather than quadratic. This is a relativistic effect.

Assuming that V is large enough, we can pass to the continuous approximation. As in Sect. 3.3, the number of waves (i.e., the number of quantum states) whose wave–vector magnitude lies between $\|\vec{\boldsymbol{k}}\|$ and $\|\vec{\boldsymbol{k}}\| + \mathrm{d}\|\vec{\boldsymbol{k}}\|$, is given by

$$\frac{(1/8)\cdot 4\pi\|\vec{\boldsymbol{k}}\|^2\mathrm{d}\|\vec{\boldsymbol{k}}\|}{(\pi/L_x)\cdot(\pi/L_y)\cdot(\pi/L_z)}=\frac{V\|\vec{\boldsymbol{k}}\|^2\mathrm{d}\|\vec{\boldsymbol{k}}\|}{2\pi^2}.$$

In terms of frequencies, using the relation $\omega = \|\vec{\boldsymbol{k}}\|c$, and doubling the above expression, due to two directions of polarization (left– and right–circular polarizations), we have that the total number of quantum states of a photon in the range $[\omega, \omega + \mathrm{d}\omega]$ is $V\omega^2\mathrm{d}\omega/\pi^2c^3$. Thus, the number of photons in this frequency range is

$$\mathrm{d}N_\omega = \frac{V}{\pi^2c^3}\cdot\frac{\omega^2\mathrm{d}\omega}{e^{\hbar\omega/kT}-1}. \tag{4.4.2}$$

The contribution of this to the energy is

$$\mathrm{d}E_\omega = \hbar\omega\mathrm{d}N_\omega = \frac{\hbar V}{\pi^2c^3}\cdot\frac{\omega^3\mathrm{d}\omega}{e^{\hbar\omega/kT}-1}. \tag{4.4.3}$$

This expression for the spectrum of black–body radiation is known as *Planck's law*.

Exercise 4.6 Write Planck's law in terms of the wavelength $\mathrm{d}E_\lambda$.

At low frequencies ($\hbar\omega \ll kT$), this gives

$$\mathrm{d}E_\omega \approx \frac{kTV}{\pi^2c^3}\omega^2\mathrm{d}\omega \tag{4.4.4}$$

which is the *Rayleigh–Jeans law*. This is actually the classic limit (see footnote at the Introduction to Chap. 3), obtained from multiplying kT by the "number of waves." In the other extreme of $\hbar\omega \gg kT$, we have

$$\mathrm{d}E_\omega = \hbar\omega\mathrm{d}N_\omega \approx \frac{\hbar V}{\pi^2c^3}\cdot\omega^3e^{-\hbar\omega/kT}\mathrm{d}\omega, \tag{4.4.5}$$

which is *Wien's law*. At low temperatures, this is an excellent approximation over a very wide range of frequencies. The frequency of maximum radiation is (Fig. 4.2)

$$\omega_{\max} = 2.822\cdot\frac{kT}{\hbar}, \tag{4.4.6}$$

namely, linear in temperature. This relation has immediate applications. For example, the sun is known to be a source of radiation, which with a good level of approximation, can be considered a black body. Using a spectrometer, one can measure the frequency $\omega_{\max}$ of maximum radiation (which turns out to be at the lower limit of the visible range), and estimate the sun's surface temperature (from Eq. (4.4.6)), to be $T \approx 5800°$ K. At room temperature, $\omega_{\max}$ falls deep in the infrared range, and thus invisible to the human eye. Hence the name black body.

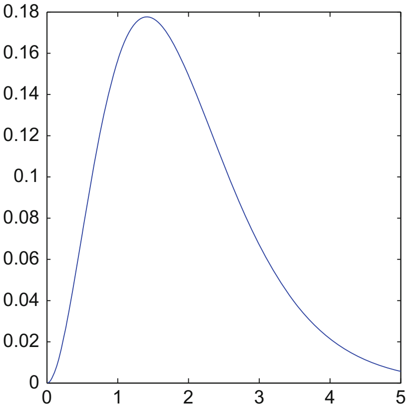


Fig. 4.2 Illustration of Planck's law. The energy density per unit frequency as a function of frequency

Now, the energy density is

$$\frac{E}{V} = \frac{\hbar}{\pi^2 c^3} \int_0^\infty \frac{\omega^3 \mathrm{d}\omega}{e^{\hbar\omega/kT} - 1} = aT^4 \tag{4.4.7}$$

where the second equality is obtained by changing the integration variable to $x = \hbar\omega/kT$ and then

$$a = \frac{\hbar}{\pi^2 c^3} \left(\frac{k}{\hbar}\right)^4 \int_0^\infty \frac{x^3 \mathrm{d}x}{e^x - 1} = \frac{\pi^2 k^4}{15\hbar^3 c^3}. \tag{4.4.8}$$

The relation $E/V = aT^4$ is called the *Stefan–Boltzmann law*. The heat capacity at constant volume, $C_V = (\partial E/\partial T)_V$, is therefore proportional to T^3.

Exercise 4.7 Calculate ρ, the density of photons.

Additional thermodynamic quantities can now be calculated from the logarithm of the grand–canonical partition function

$$\ln \Xi = -\sum_r \ln[1 - e^{-\hbar\omega_r/kT}] = -\frac{V}{\pi^2 c^3} \int_0^\infty \mathrm{d}\omega \cdot \omega^2 \ln[1 - e^{-\hbar\omega/kT}]. \tag{4.4.9}$$

For example, the pressure of the photon gas can be calculated from

$$\begin{aligned} P &= \frac{kT \ln \Xi}{V} \\ &= -\frac{kT}{\pi^2 c^3} \int_0^\infty \mathrm{d}\omega \cdot \omega^2 \ln[1 - e^{-\hbar\omega/kT}] \end{aligned}$$

$$= -\frac{(kT)^4}{\pi^2 c^3 \hbar^3} \int_0^\infty \mathrm{d}x \cdot x^2 \ln(1 - e^{-x})$$
$$= \frac{1}{3} a T^4 = \frac{E}{3V}, \tag{4.4.10}$$

where the integral is calculated using integration by parts.[5] Note that while in the ideal gas P was only linear in T, here it is proportional to the fourth power of T. Note also that here, $PV = E/3$, which is different from the ideal gas, where $PV = 2E/3$.

4.5 Suggestions for Supplementary Reading

The exposition in this chapter is heavily based on those of Mandl [1] and Pathria [2]. Additional relevant textbooks are the same as those that are mentioned also in Sect. 3.6 (as BE statistics and FD statistics are almost always presented on a similar footing).

References

1. F. Mandl, *Statistical Physics* (Wiley, Chichester, 1971)
2. R.K. Pathria, *Statistical Mechanics*, 2nd edn. (Elsevier: Butterworth-Heinemann, Oxford, 1996)

---

[5] **Exercise 4.8** Fill in the details.

Chapter 5
Interacting Particle Systems and Phase Transitions

In this chapter, we discuss systems with interacting particles. As we shall see, when the interactions among the particles are significant, the system exhibits a certain collective behavior that, in the thermodynamic limit, may be subjected to *phase transitions*, i.e., abrupt changes in the behavior of the system in the presence of a gradual change in an external control parameter, like temperature, pressure, or magnetic field. The contents of this chapter has a considerable overlap with Chap. 5 of [1], and it is provided in this book too for the sake of completeness.

5.1 Introduction – Sources of Interaction

So far, we have dealt almost exclusively with systems that have additive Hamiltonians, $\mathcal{E}(\boldsymbol{x}) = \sum_i \mathcal{E}(x_i)$, which means, under the canonical ensemble, that the particles are statistically independent and there are no interactions among them. In Nature, of course, this is seldom really the case. Sometimes this is still a reasonably good approximation, but in other cases, the interactions are appreciably strong and cannot be neglected. Among the different particles there could be many sorts of mutual forces, such as mechanical, electrical, or magnetic forces. There could also be interactions that stem from quantum–mechanical effects: as described earlier, fermions must obey Pauli's exclusion principle. Another type of interaction stems from the fact that the particles are indistinguishable, so permutations between them are not considered as distinct states. For example, referring to BE statistics, had the N particles been statistically independent, the resulting partition function would be

$$\begin{aligned} Z_N(\beta) &= \left[\sum_r e^{-\beta\epsilon_r}\right]^N \\ &= \sum_{N_1, N_2, \ldots} \delta\left(\sum_r N_r = N\right) \frac{N!}{\prod_r N_r!} \cdot \exp\left\{-\beta \sum_r N_r \epsilon_r\right\} \qquad (5.1.1) \end{aligned}$$

N. Merhav, *Statistical Physics for Electrical Engineering*,
DOI 10.1007/978-3-319-62063-3_5

whereas in Eq. (3.2.1), the combinatorial factor, $N!/\prod_r N_r!$, that distinguishes between the various permutations among the particles, is absent. This introduces dependency, which means interaction. Indeed, for the ideal boson gas, we have encountered the effect of Bose–Einstein condensation, which is a phase transition, and phase transitions can occur only in systems of interacting particles, as will be discussed in this chapter.[1]

5.2 Models of Interacting Particles

The simplest forms of deviation from the purely additive Hamiltonian structure are those that consist, in addition to the individual energy terms, $\{\mathcal{E}(x_i)\}$, also terms that depend on pairs, and/or triples, and/or even larger cliques of particles. In the case of purely pairwise interactions, this means a structure like the following:

$$\mathcal{E}(\boldsymbol{x}) = \sum_{i=1}^{N} \mathcal{E}(x_i) + \sum_{(i,j)} \varepsilon(x_i, x_j) \tag{5.2.1}$$

where the summation over pairs can be defined over all pairs $i \neq j$, or over some of the pairs, according to a given rule, e.g., depending on the distance between particle i and particle j, and according to the geometry of the system, or according to a certain graph whose edges connect the relevant pairs of variables (that in turn, are designated as nodes).

For example, in a one–dimensional array (a lattice) of particles, a customary model accounts for interactions between neighboring pairs only, neglecting more remote ones, thus the second term above would be $\sum_i \varepsilon(x_i, x_{i+1})$. A well known special case of this is that of a polymer or a solid with crystal lattice structure, where, in the one–dimensional version of the model, atoms are thought of as a chain of masses connected by springs (see left part of Fig. 5.1), i.e., an array of coupled harmonic oscillators. In this case, $\varepsilon(x_i, x_{i+1}) = \frac{1}{2}K(x_{i+1} - x_i)^2$, where K is a constant and x_i is the displacement of the i-th atom from its equilibrium location, i.e., the potential energies of the springs. In higher dimensional arrays (or lattices), similar interactions apply, there are just more neighbors to each site, from the various directions (see right part of Fig. 5.1). These kinds of models will be discussed in the next chapter in some depth.

In a system where the particles are mobile and hence their locations vary and have no geometrical structure, like in a gas, the interaction terms are also potential energies pertaining to the mutual forces (see Fig. 5.2), and these normally depend solely on the distances $\|\vec{r}_i - \vec{r}_j\|$.

---

[1] Another way to understand the dependence is to observe that occupation numbers $\{N_r\}$ are dependent via the constraint on their sum. This is different from the grand–canonical ensemble, where they are independent.

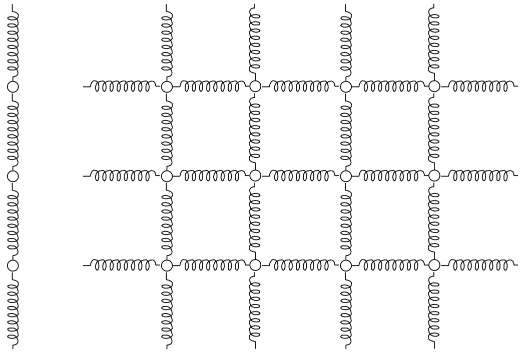


Fig. 5.1 Elastic interaction forces between adjacent atoms in a one–dimensional lattice (*left part* of the figure) and in a two–dimensional lattice (*right part*)

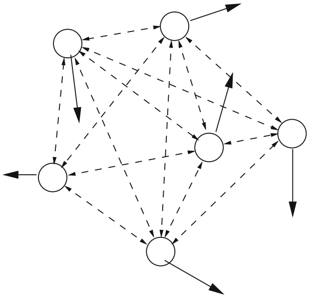


Fig. 5.2 Mobile particles and mutual forces between them

For example, in a non–ideal gas,

$$\mathcal{E}(\boldsymbol{x}) = \sum_{i=1}^{N} \frac{\|\vec{p}_i\|^2}{2m} + \sum_{i \neq j} \phi(\|\vec{r}_i - \vec{r}_j\|). \tag{5.2.2}$$

A simple special case is that of hard spheres (Billiard balls), without any forces, where

$$\phi(\|\vec{r}_i - \vec{r}_j\|) = \begin{cases} \infty & \|\vec{r}_i - \vec{r}_j\| < 2R \\ 0 & \|\vec{r}_i - \vec{r}_j\| \geq 2R \end{cases} \tag{5.2.3}$$

which expresses the simple fact that balls cannot physically overlap. The analysis of this model can be carried out using diagrammatic techniques (the cluster expansion, etc.), but we will not get into details in this book.[2] To demonstrate, however, the effect of interactions on the deviation from the equation of state of the ideal gas, we consider next a simple one–dimensional example.

Example 5.1 (*Non–ideal gas in one dimension*) Consider a one–dimensional object of length L that contains $N+1$ particles, whose locations are $0 \equiv r_0 \leq r_1 \leq \ldots \leq r_{N-1} \leq r_N \equiv L$, namely, the first and the last particles are fixed at the edges. The

---

[2] The reader can find the derivations in any textbook on elementary statistical mechanics, for example, [2, Chap. 9].

order of the particles is fixed, namely, they cannot be swapped. Let the Hamiltonian be given by

$$\mathcal{E}(\boldsymbol{x}) = \sum_{i=1}^{N} \phi(r_i - r_{i-1}) + \sum_{i=1}^{n} \frac{p_i^2}{2m} \tag{5.2.4}$$

where ϕ is a given potential function designating the interaction between two neighboring particles along the line. The partition function, which is an integral of the Boltzmann factor pertaining to this Hamiltonian, should incorporate the fact that the positions $\{r_i\}$ are not independent. It is convenient to change variables to $\xi_i = r_i - r_{i-1}$, $i = 1, 2, \ldots, N$, where it should be kept in mind that $\xi_i \geq 0$ for all i and $\sum_{i=1}^{N} \xi_i = L$. Let us assume that L is an extensive variable, i.e., $L = N\xi_0$ for some constant $\xi_0 > 0$. Thus, the partition function is

$$Z_N(\beta, L) = \frac{1}{h^N} \int \mathrm{d}p_1 \cdots \mathrm{d}p_N \int_{\mathbb{R}_N^+} \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_N e^{-\beta \sum_{i=1}^{N} [\phi(\xi_i) + p_i^2/2m]} \cdot \delta\left(L - \sum_{i=1}^{N} \xi_i\right) \tag{5.2.5}$$

$$= \frac{1}{\lambda^N} \int_{\mathbb{R}_N^+} \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_N e^{-\beta \sum_{i=1}^{N} \phi(\xi_i)} \cdot \delta\left(L - \sum_{i=1}^{N} \xi_i\right), \tag{5.2.6}$$

where $\lambda = h/\sqrt{2\pi mkT}$. The constraint $\sum_{i=1}^{N} \xi_i = L$ makes the analysis of the configurational partition function difficult. Let us pass to the corresponding Gibbs ensemble where instead of fixing the length L, we control it by applying a force f.[3] The corresponding partition function now reads

$$\begin{aligned}
Y_N(\beta, f) &= \lambda^{-N} \int_0^\infty \mathrm{d}L e^{-\beta f L} Z_N(\beta, L) \\
&= \lambda^{-N} \int_0^\infty \mathrm{d}L e^{-\beta f L} \int_{\mathbb{R}_N^+} \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_N e^{-\beta \sum_{i=1}^{N} \phi(\xi_i)} \cdot \delta\left(L - \sum_{i=1}^{N} \xi_i\right) \\
&= \lambda^{-N} \int_{\mathbb{R}_N^+} \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_N \left[\int_0^\infty \mathrm{d}L e^{-\beta f L} \delta\left(L - \sum_{i=1}^{N} \xi_i\right)\right] e^{-\beta \sum_{i=1}^{N} \phi(\xi_i)} \\
&= \lambda^{-N} \int_{\mathbb{R}_N^+} \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_N \exp\left\{-\beta\left[f \sum_{i=1}^{N} \xi_i + \sum_{i=1}^{N} \phi(\xi_i)\right]\right\} \\
&\stackrel{\Delta}{=} \lambda^{-N} \int_{\mathbb{R}_N^+} \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_N \exp\left[-s \sum_{i=1}^{N} \xi_i - \beta \sum_{i=1}^{N} \phi(\xi_i)\right] \\
&= \left\{\frac{1}{\lambda} \int_0^\infty \mathrm{d}\xi \cdot e^{-[s\xi + \beta\phi(\xi)]}\right\}^N
\end{aligned} \tag{5.2.7}$$

---

[3] Here we use the principle of ensemble equivalence.

With a slight abuse of notation, from now on, we will denote the last expression by $Y_N(\beta, s)$. Consider now the following potential function

$$\phi(\xi) = \begin{cases} \infty & 0 \leq \xi \leq d \\ -\epsilon & d < \xi \leq d + \delta \\ 0 & \xi > d + \delta \end{cases} \tag{5.2.8}$$

In words, distances below d are strictly forbidden (e.g., because of the size of the particles), in the range between d and $d+\delta$ there is a negative potential $-\epsilon$, and beyond $d + \delta$ the potential is zero.[4] Now, for this potential function, the one–dimensional integral above is given by

$$I = \int_0^\infty \mathrm{d}\xi e^{-[s\xi + \beta\phi(\xi)]} = \frac{e^{-sd}}{s}[e^{-s\delta}(1 - e^{\beta\epsilon}) + e^{\beta\epsilon}], \tag{5.2.9}$$

and so,

$$\begin{aligned} Y_N(\beta, s) &= \frac{e^{-sdN}}{\lambda^N s^N}[e^{-s\delta}(1 - e^{\beta\epsilon}) + e^{\beta\epsilon}]^N \\ &= \exp\left\{N\left[\ln[e^{-s\delta}(1 - e^{\beta\epsilon}) + e^{\beta\epsilon}] - sd - \ln(\lambda s)\right]\right\} \end{aligned} \tag{5.2.10}$$

Now, the average length of the system is given by

$$\begin{aligned} \langle L \rangle &= -\frac{\partial \ln Y_N(\beta, s)}{\partial s} \\ &= \frac{N\delta e^{-s\delta}(1 - e^{\beta\epsilon})}{e^{-s\delta}(1 - e^{\beta\epsilon}) + e^{\beta\epsilon}} + Nd + \frac{N}{s}, \end{aligned} \tag{5.2.11}$$

or, equivalently, $\langle \Delta L \rangle = \langle L \rangle - Nd$, which is the excess length beyond the possible minimum, is given by

$$\langle \Delta L \rangle = \frac{N\delta e^{-f\delta/kT}(1 - e^{\epsilon/kT})}{e^{-f\delta/kT}(1 - e^{\epsilon/kT}) + e^{\epsilon/kT}} + \frac{NkT}{f}. \tag{5.2.12}$$

Thus,

$$\begin{aligned} f \cdot \langle \Delta L \rangle &= NkT + \frac{Nf\delta e^{-f\delta/kT}(1 - e^{\epsilon/kT})}{e^{-f\delta/kT}(1 - e^{\epsilon/kT}) + e^{\epsilon/kT}} \\ &= N\left[kT - \frac{f\delta}{e^{(\epsilon + f\delta)/kT}/(e^{\epsilon/kT} - 1) - 1}\right] \end{aligned} \tag{5.2.13}$$

---

[4] This is a caricature of the Lennard–Jones potential function $\phi(\xi) \propto [(d/\xi)^{12} - (d/\xi)^6]$, which begins from $+\infty$, decreases down to a negative minimum, and finally increases and tends to zero.

where the last line is obtained after some standard algebraic manipulation. Note that without the potential well of the intermediate range of distances ($\epsilon = 0$ or $\delta = 0$), the second term in the square brackets disappears and we get a one dimensional version of the equation of state of the ideal gas (with the volume being replaced by length and the pressure – replaced by force). The second term is then a correction term due to the interaction. The attractive potential reduces the product $f \cdot \Delta L$. □

Yet another example of a model, or more precisely, a very large class of models with interactions, are those of magnetic materials. These models will closely accompany our discussions from this point onward in this chapter. Although few of these models are solvable, most of them are not. For the purpose of our discussion, a magnetic material is one for which the relevant property of each particle is its *magnetic moment*. As a reminder, the magnetic moment is a vector proportional to the angular momentum of a revolving charged particle (like a rotating electron, or a current loop), or the *spin*, and it designates the intensity of its response to the net magnetic field that this particle 'feels'. This magnetic field is given by the superposition of an externally applied magnetic field and the magnetic fields generated by the neighboring spins.

Quantum mechanical considerations dictate that each spin, which will be denoted by s_i, is quantized, that is, it may take only one out of finitely many values. In the simplest case to be adopted in our study – two values only. These will be designated by $s_i = +1$ ("spin up") and $s_i = -1$ ("spin down"), corresponding to the same intensity, but in two opposite directions, one parallel to the magnetic field, and the other – anti-parallel (see Fig. 5.3). The Hamiltonian associated with an array of spins $\boldsymbol{s} = (s_1, \ldots, s_N)$ is customarily modeled (up to certain constants that, among other things, accommodate for the physical units) with a structure like this:

$$\mathcal{E}(\boldsymbol{s}) = -B \cdot \sum_{i=1}^{N} s_i - \sum_{(i,j)} J_{ij} s_i s_j, \tag{5.2.14}$$

where B is the externally applied magnetic field and $\{J_{ij}\}$ are the coupling constants that designate the levels of interaction between spin pairs, and they depend on properties of the magnetic material and on the geometry of the system. The first term accounts for the contributions of potential energies of all spins due to the magnetic field, which in general, are given by the inner product $\vec{B} \cdot \vec{s}_i$, but since each $\vec{s}_i$ is either parallel or anti-parallel to $\vec{B}$, as said, these boil down to simple products, where only the sign of each s_i counts. Since $P(\boldsymbol{s})$ is proportional to $e^{-\beta \mathcal{E}(\boldsymbol{s})}$, the spins 'prefer' to be parallel, rather than anti-parallel to the magnetic field. The second term in the above Hamiltonian accounts for the interaction energy. If J_{ij} are all positive, they also prefer to be parallel to each other (the probability for this is larger), which is the case where the material is called *ferromagnetic* (like iron and nickel). If they are all negative, the material is *antiferromagnetic*. In the mixed case, it is called a *spin glass*. In the latter, the behavior is rather complicated.

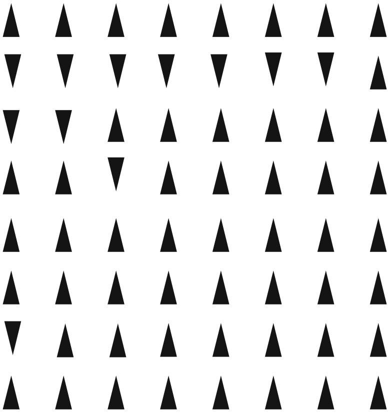


Fig. 5.3 Illustration of a spin array on a square lattice

The case where all J_{ij} are equal and the double summation over $\{(i, j)\}$ is over nearest neighbors only is called the *Ising model*. A more general version of it is called the $O(n)$ model, according to which each spin is an n–dimensional unit vector $\vec{s}_i = (s_i^1, \ldots, s_i^n)$ (and so is the magnetic field), where n is not necessarily related to the dimension d of the lattice in which the spins reside. The case $n = 1$ is then the Ising model. The case $n = 2$ is called the *XY model*, and the case $n = 3$ is called the *Heisenberg model*.

Of course, the above models for the Hamiltonian can (and, in fact, are being) generalized to include interactions formed also, by triples, quadruples, or any fixed size p (that does not grow with N) of spin–cliques.

We next discuss a very important effect that exists in some systems with strong interactions (both in magnetic materials and in other models): the effect of *phase transitions*.

5.3 A Qualitative Discussion on Phase Transitions

As was mentioned in the introductory paragraph of this chapter, a phase transition means an abrupt change in the collective behavior of a physical system, as we change gradually one of the externally controlled parameters, like the temperature, pressure, or magnetic field. The most common example of a phase transition in our everyday life is the water that we boil in the kettle when we make coffee, or when it turns into ice as we put it in the freezer.

What exactly these phase transitions are? In physics, phase transitions can occur only if the system has interactions. Consider, the above example of an array of spins with $B = 0$, and let us suppose that all $J_{ij} > 0$ are equal, and thus will be denoted commonly by J (like in the $O(n)$ model). Then,

$$P(\boldsymbol{s}) = \frac{\exp\left\{\beta J \sum_{(i,j)} s_i s_j\right\}}{Z(\beta)} \tag{5.3.1}$$

and, as mentioned earlier, this is a ferromagnetic model, where all spins 'like' to be in the same direction, especially when βJ is large. In other words, the interactions, in this case, tend to introduce *order* into the system. On the other hand, the second law talks about maximum entropy, which tends to increase the *disorder*. So there are two conflicting effects here. Which one of them prevails?

The answer turns out to depend on temperature. Recall that in the canonical ensemble, equilibrium is attained at the point of minimum free energy $f = \epsilon - Ts(\epsilon)$. Now, T plays the role of a weighting factor for the entropy. At low temperatures, the weight of the second term of f is small, and minimizing f is approximately equivalent to minimizing ϵ, which is obtained by states with a high level of order, as $\mathcal{E}(\boldsymbol{s}) = -J \sum_{(i,j)} s_i s_j$, in this example. As T grows, however, the weight of the term $-Ts(\epsilon)$ increases, and min f, becomes more and more equivalent to max $s(\epsilon)$, which is achieved by states with a high level of disorder (see Fig. 5.4). Thus, the order–disorder characteristics depend primarily on temperature. It turns out that for some magnetic systems of this kind, this transition between order and disorder may be abrupt, in which case, we call it a *phase transition*. At a certain critical temperature, called the *Curie temperature*, there is a sudden transition between order and disorder. In the ordered phase, a considerable fraction of the spins align in the same direction, which means that the system is spontaneously magnetized (even without an external magnetic field), whereas in the disordered phase, about half of the spins are in either direction, and then the net magnetization vanishes. This happens if the interactions, or more precisely, their dimension in some sense, is strong enough.

What is the mathematical significance of a phase transition? If we look at the partition function, $Z_N(\beta)$, which is the key to all physical quantities of interest, then for every finite N, this is simply the sum of finitely many exponentials in β and therefore it is continuous and differentiable infinitely many times. So what kind of abrupt changes could there possibly be in the behavior of this function? It turns out that while this is true for all finite N, it is no longer necessarily true if we look at the thermodynamic limit, i.e., if we look at the behavior of

$$\phi(\beta) = \lim_{N\to\infty} \frac{\ln Z_N(\beta)}{N}. \tag{5.3.2}$$

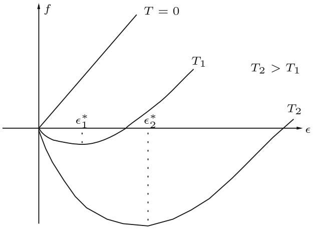


Fig. 5.4 Qualitative graphs of $f(\epsilon)$ at various temperatures. The minimizing ϵ increases with T

While $\phi(\beta)$ must be continuous for all $\beta > 0$ (since it is convex), it need not necessarily have continuous derivatives. Thus, a phase transition, if exists, is fundamentally an asymptotic property, it may exist in the thermodynamic limit only. While a physical system is, after all finite, it is nevertheless well approximated by the thermodynamic limit when it is very large.

The above discussion explains also why a system without interactions, where all $\{x_i\}$ are i.i.d., cannot have phase transitions. In this case, $Z_N(\beta) = [Z_1(\beta)]^N$, and so, $\phi(\beta) = \ln Z_1(\beta)$, which is always a smooth function without any irregularities. For a phase transition to occur, the particles must behave in some collective manner, which is the case only if interactions take place.

There is a distinction between two types of phase transitions:

- If $\phi(\beta)$ has a discontinuous first order derivative, then this is called a *first order phase transition*.
- If $\phi(\beta)$ has a continuous first order derivative, but a discontinuous second order derivative then this is called a *second order phase transition*, or a *continuous phase transition*.

We can talk, of course, about phase transitions w.r.t. additional parameters other than temperature. In the above magnetic example, if we introduce back the magnetic field B into the picture, then Z, and hence also ϕ, become functions of B too. If we then look at derivative of

$$\begin{aligned}\phi(\beta, B) &= \lim_{N\to\infty} \frac{\ln Z_N(\beta, B)}{N} \\ &= \lim_{N\to\infty} \frac{1}{N} \ln \left[\sum_{s} \exp\left\{\beta B \sum_{i=1}^{N} s_i + \beta J \sum_{(i,j)} s_i s_j\right\}\right]\end{aligned} \tag{5.3.3}$$

w.r.t. the product (βB), which multiplies the magnetization, $\sum_i s_i$, at the exponent, this would give exactly the average magnetization per spin

$$m(\beta, B) = \left\langle \frac{1}{N} \sum_{i=1}^{N} S_i \right\rangle, \tag{5.3.4}$$

and this quantity might not always be continuous. Indeed, as mentioned earlier, below the Curie temperature there might be a spontaneous magnetization. If $B \downarrow 0$, then this magnetization is positive, and if $B \uparrow 0$, it is negative, so there is a discontinuity at $B = 0$. We shall see this more concretely later on.

5.4 The One–Dimensional Ising Model

The most familiar model of a magnetic system with interactions is the one–dimensional Ising model, according to which

$$\mathcal{E}(\boldsymbol{s}) = -B\sum_{i=1}^{N} s_i - J\sum_{i=1}^{N} s_i s_{i+1} \tag{5.4.1}$$

with the periodic boundary condition $s_{N+1} = s_1$. Thus,

$$\begin{aligned} Z_N(\beta, B) &= \sum_{\boldsymbol{s}} \exp\left\{\beta B\sum_{i=1}^{N} s_i + \beta J\sum_{i=1}^{N} s_i s_{i+1}\right\} \\ &= \sum_{\boldsymbol{s}} \exp\left\{h\sum_{i=1}^{N} s_i + K\sum_{i=1}^{N} s_i s_{i+1}\right\} \qquad h \stackrel{\Delta}{=} \beta B, \quad K \stackrel{\Delta}{=} \beta J \\ &= \sum_{\boldsymbol{s}} \exp\left\{\frac{h}{2}\sum_{i=1}^{N}(s_i + s_{i+1}) + K\sum_{i=1}^{N} s_i s_{i+1}\right\}. \end{aligned} \tag{5.4.2}$$

Consider now the 2×2 matrix P whose entries are $\exp\{\frac{h}{2}(s + s') + Kss'\}$, $s, s' \in \{-1, +1\}$, i.e.,

$$P = \begin{pmatrix} e^{K+h} & e^{-K} \\ e^{-K} & e^{K-h} \end{pmatrix}. \tag{5.4.3}$$

Also, $s_i = +1$ will be represented by the column vector $\sigma_i = (1, 0)^T$ and $s_i = -1$ will be represented by $\sigma_i = (0, 1)^T$. Thus,

$$\begin{aligned} Z(\beta, B) &= \sum_{\sigma_1} \cdots \sum_{\sigma_N} (\sigma_1^T P \sigma_2) \cdot (\sigma_2^T P \sigma_2) \cdots (\sigma_N^T P \sigma_1) \\ &= \sum_{\sigma_1} \sigma_1^T P \left(\sum_{\sigma_2} \sigma_2 \sigma_2^T\right) P \left(\sum_{\sigma_3} \sigma_3 \sigma_3^T\right) P \cdots P \left(\sum_{\sigma_N} \sigma_N \sigma_N^T\right) P \sigma_1 \\ &= \sum_{\sigma_1} \sigma_1^T P \cdot I \cdot P \cdot I \cdots I \cdot P \sigma_1 \\ &= \sum_{\sigma_1} \sigma_1^T P^N \sigma_1 \\ &= \operatorname{tr}\{P^N\} \\ &= \lambda_1^N + \lambda_2^N \end{aligned} \tag{5.4.4}$$

where λ_1 and λ_2 are the eigenvalues of P, which are

$$\lambda_{1,2} = e^K \cosh(h) \pm \sqrt{e^{-2K} + e^{2K}\sinh^2(h)}. \tag{5.4.5}$$

Letting λ_1 denote the larger (the dominant) eigenvalue, i.e.,

$$\lambda_1 = e^K \cosh(h) + \sqrt{e^{-2K} + e^{2K}\sinh^2(h)}, \tag{5.4.6}$$

then clearly,

$$\phi(h, K) = \lim_{N\to\infty} \frac{\ln Z_N(h, K)}{N} = \ln \lambda_1. \tag{5.4.7}$$

The average magnetization is

$$\begin{aligned} M(h, K) &= \left\langle \sum_{i=1}^{N} S_i \right\rangle \\ &= \frac{\sum_s (\sum_{i=1}^N s_i) \exp\{h \sum_{i=1}^N s_i + K \sum_{i=1}^N s_i s_{i+1}\}}{\sum_s \exp\{h \sum_{i=1}^N s_i + K \sum_{i=1}^N s_i s_{i+1}\}} \\ &= \frac{\partial \ln Z(h, K)}{\partial h} \end{aligned} \tag{5.4.8}$$

and so, the per–spin magnetization is:

$$m(h, K) \stackrel{\Delta}{=} \lim_{N\to\infty} \frac{M(h, K)}{N} = \frac{\partial \phi(h, K)}{\partial h} = \frac{\sinh(h)}{\sqrt{e^{-4K} + \sinh^2(h)}} \tag{5.4.9}$$

or, returning to the original parametrization:

$$m(\beta, B) = \frac{\sinh(\beta B)}{\sqrt{e^{-4\beta J} + \sinh^2(\beta B)}}. \tag{5.4.10}$$

For $\beta > 0$ and $B > 0$ this is a smooth function, and so, there are no phase transitions and no spontaneous magnetization at any finite temperature.[5] However, at the absolute zero ($\beta \to \infty$), we get

$$\lim_{B\downarrow 0} \lim_{\beta\to\infty} m(\beta, B) = +1; \quad \lim_{B\uparrow 0} \lim_{\beta\to\infty} m(\beta, B) = -1, \tag{5.4.11}$$

---

[5] Note, in particular, that for $J = 0$ (i.i.d. spins) we get paramagnetic characteristics $m(\beta, B) = \tanh(\beta B)$, in agreement with the result pointed out in the example of two–level systems, in the comment that follows Example 2.3.

thus m is discontinuous w.r.t. B at $\beta \to \infty$, which means that there is a phase transition at $T = 0$. In other words, the Curie temperature is $T_c = 0$ independent of J.

We see then that one–dimensional Ising model is easy to handle, but it is not very interesting in the sense that there is actually no phase transition. The extension to the two–dimensional Ising model on the square lattice is surprisingly more difficult. It is still solvable, but only without an external magnetic field. It was first solved in 1944 by Onsager [3], who has shown that it exhibits a phase transition with Curie temperature given by

$$T_c = \frac{2J}{k\ln(\sqrt{2}+1)}. \tag{5.4.12}$$

For lattice dimension larger than two, the problem is still open.

It turns out then that what counts for the existence of phase transitions, is not only the intensity of the interactions (designated by the magnitude of J), but more importantly, the "dimensionality" of the structure of the pairwise interactions. If we denote by n_ℓ the number of ℓ–th order neighbors of every given site, namely, the number of sites that can be reached within ℓ steps from the given site, then what counts is how fast does the sequence $\{n_\ell\}$ grow, or more precisely, what is the value of $d \triangleq \lim_{\ell\to\infty} \frac{\ln n_\ell}{\ln \ell}$, which is exactly the ordinary dimensionality for hyper-cubic lattices. Loosely speaking, this dimension must be sufficiently large for a phase transition to exist.

To demonstrate this point, we next discuss an extreme case of a model where this dimensionality is actually infinite. In this model "everybody is a neighbor of everybody else" and to the same extent, so it definitely has the highest connectivity possible. This is not quite a physically realistic model, but it is pleasing that it is easy to solve and that it exhibits a phase transition that is fairly similar to those that exist in real systems. It is also intimately related to a very popular approximation method in statistical mechanics, called the *mean field approximation*. Hence it is sometimes called the *mean field model*. It is also known as the *Curie–Weiss model* or the *infinite range model*.

Finally, we should comment that there are other "infinite–dimensional" Ising models, like the one defined on the Bethe lattice (an infinite tree without a root and without leaves), which is also easily solvable (by recursion) and it also exhibits phase transitions [4], but we will not discuss it here.

5.5 The Curie–Weiss Model

According to the Curie–Weiss (C–W) model,

$$\mathcal{E}(\boldsymbol{s}) = -B\sum_{i=1}^{N} s_i - \frac{J}{2N}\sum_{i\neq j} s_i s_j. \tag{5.5.1}$$

Here, all pairs $\{(s_i, s_j)\}$ communicate to the same extent, and without any geometry. The $1/N$ factor here is responsible for keeping the energy of the system extensive (linear in N), as the number of interaction terms is quadratic in N. The factor $1/2$ compensates for the fact that the summation over $i \neq j$ counts each pair twice. The first observation is the trivial fact that

$$\left(\sum_i s_i\right)^2 = \sum_i s_i^2 + \sum_{i \neq j} s_i s_j = N + \sum_{i \neq j} s_i s_j \tag{5.5.2}$$

where the second equality holds since $s_i^2 \equiv 1$. It follows then, that our Hamiltonian is, up to a(n immaterial) constant, equivalent to

$$\begin{aligned}\mathcal{E}(\boldsymbol{s}) &= -B\sum_{i=1}^N s_i - \frac{J}{2N}\left(\sum_{i=1}^N s_i\right)^2 \\ &= -N\left[B \cdot \left(\frac{1}{N}\sum_{i=1}^N s_i\right) + \frac{J}{2}\left(\frac{1}{N}\sum_{i=1}^N s_i\right)^2\right],\end{aligned} \tag{5.5.3}$$

thus $\mathcal{E}(\boldsymbol{s})$ depends on $\boldsymbol{s}$ only via the magnetization $m(\boldsymbol{s}) = \frac{1}{N}\sum_i s_i$. This fact makes the C–W model very easy to handle:

$$\begin{aligned}Z_N(\beta, B) &= \sum_{\boldsymbol{s}} \exp\left\{N\beta\left[B \cdot m(\boldsymbol{s}) + \frac{J}{2}m^2(\boldsymbol{s})\right]\right\} \\ &= \sum_{m=-1}^{+1} \Omega(m) \cdot e^{N\beta(Bm + Jm^2/2)} \\ &\doteq \sum_{m=-1}^{+1} e^{Nh_2((1+m)/2)} \cdot e^{N\beta(Bm+Jm^2/2)} \\ &\doteq \exp\left\{N \cdot \max_{|m|\le 1}\left[h_2\left(\frac{1+m}{2}\right) + \beta Bm + \frac{\beta m^2 J}{2}\right]\right\}\end{aligned} \tag{5.5.4}$$

and so,

$$\phi(\beta, B) = \max_{|m|\le 1}\left[h_2\left(\frac{1+m}{2}\right) + \beta Bm + \frac{\beta m^2 J}{2}\right]. \tag{5.5.5}$$

The maximum is found by equating the derivative to zero, i.e.,

$$0 = \frac{1}{2}\ln\left(\frac{1-m}{1+m}\right) + \beta B + \beta Jm \equiv -\tanh^{-1}(m) + \beta B + \beta Jm \tag{5.5.6}$$

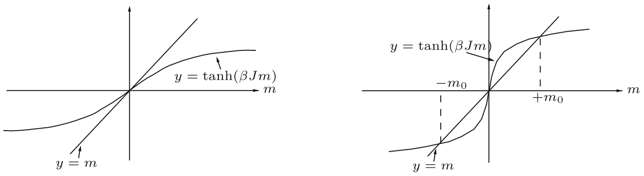


Fig. 5.5 Graphical solutions of equation $m = \tanh(\beta J m)$: The *left part* corresponds to the case $\beta J < 1$, where there is one solution only, $m^* = 0$. The *right part* corresponds to the case $\beta J > 1$, where in addition to the zero solution, there are two non–zero solutions $m^* = \pm m_0$

or equivalently, the maximizing (and hence the dominant) m is a solution m^* to the equation[6]

$$m = \tanh(\beta B + \beta J m).$$

Consider first the case $B = 0$, where the equation boils down to

$$m = \tanh(\beta J m). \tag{5.5.7}$$

It is instructive to look at this equation graphically. Referring to Fig. 5.5, we have to make a distinction between two cases: If $\beta J < 1$, namely, $T > T_c \triangleq J/k$, the slope of the function $y = \tanh(\beta J m)$ at the origin, βJ, is smaller than the slope of the linear function $y = m$, which is 1, thus these two graphs intersect only at the origin. It is easy to check that in this case, the second derivative of

$$\psi(m) \triangleq h_2\left(\frac{1+m}{2}\right) + \frac{\beta J m^2}{2} \tag{5.5.8}$$

at $m = 0$ is negative, and therefore it is indeed the maximum (see Fig. 5.6, left part). Thus, the dominant magnetization is $m^* = 0$, which means disorder and hence no spontaneous magnetization for $T > T_c$. On the other hand, when $\beta J > 1$, which means temperatures lower than T_c, the initial slope of the tanh function is larger than that of the linear function, but since the tanh function cannot take values outside the interval $(-1, +1)$, the two functions must intersect also at two additional, symmetric, non–zero points, which we denote by $+m_0$ and $-m_0$ (see Fig. 5.5, right part). In this case, it can readily be shown that the second derivative of $\psi(m)$ is positive at the origin (i.e., there is a local minimum at $m = 0$) and negative at $m = \pm m_0$, which means that there are maxima at these two points (see Fig. 5.6, right part). Thus, the dominant magnetizations are $\pm m_0$, each capturing about half of the probability.

Consider now the case $\beta J > 1$, where the magnetic field B is brought back into the picture. This will break the symmetry of the right graph of Fig. 5.6 and the corresponding graphs of $\psi(m)$ would be as in Fig. 5.7, where now the higher local

---

[6] Once again, for $J = 0$, we are back to non–interacting spins and then this equation gives the paramagnetic behavior $m = \tanh(\beta B)$.

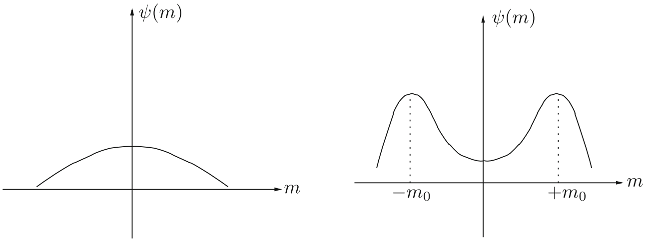


Fig. 5.6 The function $\psi(m) = h_2((1+m)/2) + \beta J m^2/2$ has a unique maximum at $m = 0$ when $\beta J < 1$ (*left graph*) and two local maxima at $\pm m_0$, in addition to a local minimum at $m = 0$, when $\beta J > 1$ (*right graph*)

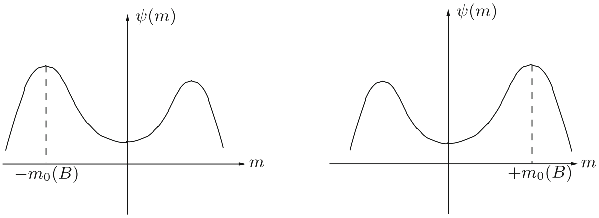


Fig. 5.7 The case $\beta J > 1$ with a magnetic field B. The *left graph* corresponds to $B < 0$ and the *right graph* – to $B > 0$

maximum (which is also the global one) is at $m_0(B)$ whose sign is as that of B. But as $B \to 0$, $m_0(B) \to m_0$ of Fig. 5.6. Thus, we see the spontaneous magnetization here. Even after removing the magnetic field, the system remains magnetized to the level of m_0, depending on the direction (the sign) of B before its removal. Obviously, the magnetization $m(\beta, B)$ has a discontinuity at $B = 0$ for $T < T_c$, which is a first order phase transition w.r.t. B (see Fig. 5.8). We note that the point $T = T_c$ is the

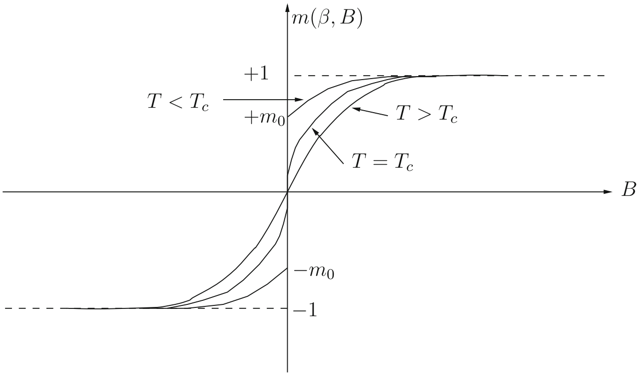


Fig. 5.8 Magnetization versus magnetic field: For $T < T_c$ there is spontaneous magnetization: $\lim_{B\downarrow 0} m(\beta, B) = +m_0$ and $\lim_{B\uparrow 0} m(\beta, B) = -m_0$, and so there is a discontinuity at $B = 0$

boundary between the region of existence and the region of non–existence of a phase transition w.r.t. B. Such a point is called a *critical point*. The phase transition w.r.t. β is of the second order.

Finally, we should mention here an alternative technique that can be used to analyze this model, which is based on the so called Hubbard–Stratonovich transform. Specifically, we have the following chain of equalities:

$$
\begin{aligned}
Z(h,K) &= \sum_{s} \exp\left\{ h\sum_{i=1}^{N} s_i + \frac{K}{2N}\left(\sum_{i=1}^{N} s_i\right)^2 \right\} \qquad h \triangleq \beta B,\ K \triangleq \beta J \\
&= \sum_{s} \exp\left\{ h\sum_{i=1}^{N} s_i \right\} \cdot \exp\left\{ \frac{K}{2N}\left(\sum_{i=1}^{N} s_i\right)^2 \right\} \\
&= \sum_{s} \exp\left\{ h\sum_{i=1}^{N} s_i \right\} \cdot \sqrt{\frac{N}{2\pi K}} \int_{\mathbb{R}} \mathrm{d}z \exp\left\{ -\frac{Nz^2}{2K} + z\cdot \sum_{i=1}^{N} s_i \right\} \\
&= \sqrt{\frac{N}{2\pi K}} \int_{\mathbb{R}} \mathrm{d}z e^{-Nz^2/(2K)} \sum_{s} \exp\left\{ (h+z)\sum_{i=1}^{N} s_i \right\} \\
&= \sqrt{\frac{N}{2\pi K}} \int_{\mathbb{R}} \mathrm{d}z e^{-Nz^2/(2K)} \left[\sum_{s=-1}^{1} e^{(h+z)s} \right]^N \\
&= \sqrt{\frac{N}{2\pi K}} \int_{\mathbb{R}} \mathrm{d}z e^{-Nz^2/(2K)} [2\cosh(h+z)]^N \\
&= 2^N \cdot \sqrt{\frac{N}{2\pi K}} \int_{\mathbb{R}} \mathrm{d}z \exp\{N[\ln\cosh(h+z) - z^2/(2K)]\}, \qquad (5.5.9)
\end{aligned}
$$

where the passage from the second to the third line follows the use of the characteristic function of a Gaussian random variable: If $X \sim \mathcal{N}(0,\sigma^2)$, then $\langle e^{\alpha X}\rangle = e^{\alpha^2\sigma^2/2}$ (in our case, $\sigma^2 = K/N$ and $\alpha = \sum_i s_i$).

The integral in the last line can be shown (see, e.g., [1, Chap. 4]) to be dominated by e to N times the maximum of the function in the square brackets at the exponent of the integrand, or equivalently, the minimum of the function

$$\gamma(z) = \frac{z^2}{2K} - \ln\cosh(h+z). \qquad (5.5.10)$$

by equating its derivative to zero, we get the very same equation as $m = \tanh(\beta B + \beta J m)$ by setting $z = \beta J m$. The function $\gamma(z)$ is different from the function ψ that we maximized earlier, but the extremum is the same. This function is called the *Landau free energy*.

5.6 Spin Glasses*

So far we discussed only models where the non–zero coupling coefficients, $\boldsymbol{J} = \{J_{ij}\}$ are equal, thus they are either all positive (ferromagnetic models) or all negative (antiferromagnetic models). As mentioned earlier, there are also models where the signs of these coefficients are mixed, which are called *spin glass* models.

Spin glass models have a much more complicated and more interesting behavior than ferromagnets, because there might be meta-stable states, due to the fact that not all spin pairs $\{(s_i, s_j)\}$ can necessarily be in their preferred mutual polarization. It might be the case that some of these pairs are "frustrated." In order to model situations of amorphism and disorder in such systems, it is customary to model the coupling coefficients as random variables. This model with random parameters means that there are now two levels of randomness:

- Randomness of the coupling coefficients $\boldsymbol{J}$.
- Randomness of the spin configuration $\boldsymbol{s}$ given $\boldsymbol{J}$, according to the Boltzmann distribution, i.e.,

$$P(\boldsymbol{s}|\boldsymbol{J}) = \frac{\exp\left\{\beta\left[B\sum_{i=1}^{N} s_i + \sum_{(i,j)} J_{ij}s_is_j\right]\right\}}{Z(\beta, B|\boldsymbol{J})}. \tag{5.6.1}$$

However, these two sets of random variables have a rather different stature. The underlying setting is normally such that $\boldsymbol{J}$ is considered to be randomly drawn once and for all, and then remain fixed, whereas $\boldsymbol{s}$ keeps varying all the time (according to the dynamics of the system). At any rate, the time scale along which $\boldsymbol{s}$ varies is much smaller than that of $\boldsymbol{J}$. Another difference is that $\boldsymbol{J}$ is normally not assumed to depend on temperature, whereas $\boldsymbol{s}$ does. In the terminology of physicists, $\boldsymbol{s}$ is considered an *annealed* random variable, whereas $\boldsymbol{J}$ is considered a *quenched* random variable.[7] Accordingly, there is a corresponding distinction between *annealed averages* and *quenched averages*.

Let us see what is exactly the difference between the quenched averaging and the annealed one. If we examine, for instance, the free energy, or the log–partition function, $\ln Z(\beta|\boldsymbol{J})$, this is now a random variable because it depends on the random $\boldsymbol{J}$. If we denote by $\langle\cdot\rangle_{\boldsymbol{J}}$ the expectation w.r.t. the randomness of $\boldsymbol{J}$, then quenched averaging means $\langle\ln Z(\beta|\boldsymbol{J})\rangle_{\boldsymbol{J}}$, whereas annealed averaging means $\ln\langle Z(\beta|\boldsymbol{J})\rangle_{\boldsymbol{J}}$. Normally, the relevant average is the quenched one, because the random variable $\frac{1}{N}\ln Z(\beta|\boldsymbol{J})$ typically converges to the same limit as its expectation $\frac{1}{N}\langle\ln Z(\beta|\boldsymbol{J})\rangle_{\boldsymbol{J}}$ (the so called self–averaging property), but more often than not, it is also much harder to calculate. Clearly, the annealed average is never smaller than the quenched one because of Jensen's inequality, but they sometimes coincide at high temperatures. The difference between them is that in quenched averaging, the dominant realizations

---

[7] In a nutshell, annealing means slow cooling, whereas quenching means fast cooling, that causes the material to freeze without enough time to settle in an ordered structure. The result is then a disordered structure, modeled by frozen (fixed) random parameters, $\boldsymbol{J}$.

of $\boldsymbol{J}$ are the typical ones, whereas in annealed averaging, this is not necessarily the case. This follows from the following sketchy consideration. As for the annealed average, we have:

$$\begin{aligned}
\langle Z(\beta|\boldsymbol{J})\rangle_{\boldsymbol{J}} &= \sum_{\boldsymbol{J}} P(\boldsymbol{J})Z(\beta|\boldsymbol{J}) \\
&\approx \sum_{\alpha} \Pr\{\boldsymbol{J} :\ Z(\beta|\boldsymbol{J}) \doteq e^{N\alpha}\} \cdot e^{N\alpha} \\
&\approx \sum_{\alpha} e^{-NE(\alpha)} \cdot e^{N\alpha} \qquad \text{(assuming exponential probabilities)} \\
&\doteq e^{N \max_\alpha [\alpha - E(\alpha)]} \qquad (5.6.2)
\end{aligned}$$

which means that the annealed average is dominated by realizations of the system with

$$\frac{\ln Z(\beta|\boldsymbol{J})}{N} \approx \alpha^* \stackrel{\Delta}{=} \arg\max_{\alpha}[\alpha - E(\alpha)], \qquad (5.6.3)$$

which may differ from the typical value of α, which is

$$\alpha = \phi(\beta) \equiv \lim_{N\to\infty} \frac{1}{N} \langle \ln Z(\beta|\boldsymbol{J})\rangle . \qquad (5.6.4)$$

On the other hand, when it comes to quenched averaging, the random variable $\ln Z(\beta|\boldsymbol{J})$ behaves linearly in N, and concentrates strongly around the typical value $N\phi(\beta)$, whereas other values are weighted by (exponentially) decaying probabilities.

The literature on spin glasses includes many models for the randomness of the coupling coefficients. We end this part by listing just a few.

- The *Edwards–Anderson* (E–A) model, where $\{J_{ij}\}$ are non–zero for nearest–neighbor pairs only (e.g., $j = i \pm 1$ in one–dimensional model). According to this model, $\{J_{ij}\}$ are i.i.d. random variables, which are normally modeled to have a zero–mean Gaussian pdf, or binary symmetric with levels $\pm J_0$. It is customary to work with a zero–mean distribution if we have a pure spin glass in mind. If the mean is nonzero, the model has either a ferromagnetic or an anti-ferromagnetic bias, according to the sign of the mean.
- The *Sherrington–Kirkpatrick* (S–K) model, which is similar to the E–A model, except that the support of $\{J_{ij}\}$ is extended to include all $N(N-1)/2$ pairs, and not only nearest–neighbor pairs. This can be thought of as a stochastic version of the C–W model in the sense that here too, there is no geometry, and every spin interacts with every other spin to the same extent, but here the coefficients are random, as said.
- The *p–spin* model, which is similar to the S–K model, but now the interaction term consists, not only of pairs, but also of triples, quadruples, and so on, up to cliques of size p, i.e., products $s_{i_1}s_{i_2}\cdots s_{i_p}$, where $(i_1, \ldots, i_p)$ exhaust all possible subsets of p spins out of N. Each such term has a Gaussian coefficient $J_{i_1,\ldots,i_p}$ with an appropriate variance.

Considering the p–spin model, it turns out that if we look at the extreme case of $p \to \infty$ (taken after the thermodynamic limit $N \to \infty$), the resulting behavior turns out to be extremely erratic: all energy levels $\{\mathcal{E}(\boldsymbol{s})\}_{\boldsymbol{s}\in\{-1,+1\}^N}$ become i.i.d. Gaussian random variables. This is, of course, a toy model, which has very little to do with reality (if any), but it is surprisingly interesting and easy to work with. It is called the *random energy model* (REM).

5.7 Suggestions for Supplementary Reading

As mentioned earlier, part of the presentation in this chapter is similar to Chap. 5 of [1]. The topic of interacting particles and phase transitions is covered in many textbooks, including: Huang [5, Part C], Kardar [6, Chap. 5], Landau and Lifshitz [7, Chap. VI and onward], Pathria [2, Chaps. 10–13], and Reif [8, Chap. 10].

References

1. N. Merhav, Statistical physics and information theory. Found. Trends Commun. Inf. Theory **6**(1–2), 1–212 (2009)
2. R.K. Pathria, *Statistical Mechanics*, 2nd edn. (Elsevier: Butterworth–Heinemann, Oxford, 1996)
3. L. Onsager, Crystal statistics. I. A two-dimensional model with an order-disorder transition. Phys. Rev. **65**(3–4), 117–149 (1944)
4. R.J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, London, 1982)
5. K. Huang, *Statistical Mechanics*, 2nd edn. (Wiley, New York, 1987)
6. M. Kardar, *Statistical Physics of Particles* (Cambridge University Press, Cambridge, 2007)
7. L.D. Landau, E.M. Lifshitz, *Course of Theoretical Physics – Volume 5: Statistical Physics, Part 1*, 3rd edn. (Elsevier: Butterworth–Heinemann, New York, 1980)
8. F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1965)

Chapter 6
Vibrations in a Solid – Phonons and Heat Capacity*

6.1 Introduction

In analogy to black–body radiation, discussed earlier, there is a similar issue related to vibrational modes of a solid. As in black–body radiation, the analysis of vibrational modes in a solid can be viewed either by regarding the system as *interacting* harmonic oscillators or as a gas of particles, called *phonons* – the analogue of photons, but in the context of sound waves, rather than electromagnetic waves.

In this chapter, we shall use this point of view and apply statistical mechanical methods to calculate the heat capacity pertaining to the lattice vibrations of crystalline solids.[1] There are two basic experimental facts which any reasonable theory must be able to explain. The first is that in room temperature the heat capacity of most solids is about 3 k for each atom.[2] This is the Dulong and Petit law (1819), but this is only an approximation. The second fact is that at low temperatures, the heat capacity at constant volume, C_V, decreases, and actually vanishes at $T = 0$. Experimentally, it was observed that the low–temperature dependence is of the form $C_V = \alpha T^3 + \gamma T$, where α and γ are constants that depend on the material and the volume. For certain insulators, like potassium chloride, $\gamma = 0$, namely, C_V is proportional to T^3. For metals (like copper), the linear term is present, but it is contributed by the conduction electrons. A good theory of the vibrational contribution to heat capacity should therefore predict the T^3 behavior at low temperatures. In classical statistical mechanics, the equipartition theorem suggests a constant heat capacity *at all temperatures*, in contradiction with both experiments and with the third law of

---

[1] In general, there are additional contributions to the heat capacity (e.g., from orientational ordering in paramagnetic salts, or from conduction electrons in metals, etc.), but here we shall consider only the vibrational heat capacity.

[2] Each atom has 6 degrees of freedom (3 of position + 3 of momentum). Classically, each one of them contributes one quadratic term to the Hamiltonian, whose mean is $kT/2$, thus a total mean energy of 3 kT, which means specific heat of 3 k per atom.

N. Merhav, *Statistical Physics for Electrical Engineering*,
DOI 10.1007/978-3-319-62063-3_6

thermodynamics that asserts that as $T \to 0$, the entropy S tends to zero (whereas a constant heat capacity would yield $S \propto \ln T$ for small T). A fundamental contribution in resolving this contradiction between theory and experiment was due to Einstein (1907), who considered the lattice vibrations in the quantum regime. Einstein's derivations reproduce the desired features (observed experimentally) at least qualitatively. However, he used a simplified model and did not expect full agreement with experiment, but he pointed out the kind of modifications which the model requires. Einstein's theory was later improved by Debye (1912), who considered a more realistic model.

6.2 Formulation

Consider a Hamiltonian of a classical solid composed of N atoms whose positions in space are specified by the coordinates $\boldsymbol{x} = (x_1, \dots, x_{3N})$. In the state of lowest energy (the ground state), these coordinates are denoted by $\bar{\boldsymbol{x}} = (\bar{x}_1, \dots, \bar{x}_{3N})$, which are normally points of a lattice in the three–dimensional space, if the solid in question is a crystal. Let $\xi_i = x_i - \bar{x}_i$, $i = 1, \dots, 3N$, denote the displacements. The kinetic energy of the system is clearly

$$K = \frac{1}{2} m \sum_{i=1}^{3N} \dot{x}_i^2 = \frac{1}{2} m \sum_{i=1}^{3N} \dot{\xi}_i^2 \tag{6.2.1}$$

and the potential energy is

$$\Phi(\boldsymbol{x}) = \Phi(\bar{\boldsymbol{x}}) + \sum_i \left(\frac{\partial \Phi}{\partial x_i} \right)_{\boldsymbol{x}=\bar{\boldsymbol{x}}} \xi_i + \sum_{i,j} \frac{1}{2} \left(\frac{\partial^2 \Phi}{\partial x_i \partial x_j} \right)_{\boldsymbol{x}=\bar{\boldsymbol{x}}} \xi_i \xi_j + \dots \tag{6.2.2}$$

The first term in this expansion represents the minimum energy when all atoms are at rest in their mean positions $\bar{x}_i$. We henceforth denote this energy by Φ_0. The second term is identically zero because $\Phi(\boldsymbol{x})$ is minimized at $\boldsymbol{x} = \bar{\boldsymbol{x}}$. The second order terms of this expansion represent the *harmonic component* of the vibrations. If we assume that the overall amplitude of the vibrations is reasonably small, we can safely neglect all successive terms and then we are working with the so called *harmonic approximation*. Thus, we may write

$$\mathcal{E}(\boldsymbol{x}) = \Phi_0 + \frac{1}{2} m \sum_{i=1}^{3N} \dot{\xi}_i^2 + \sum_{i,j} \alpha_{ij} \xi_i \xi_j \tag{6.2.3}$$

where we have denoted

$$\alpha_{ij} = \frac{1}{2} \cdot \frac{\partial^2 \Phi}{\partial x_i \partial x_j}\bigg|_{x=\bar{x}}. \tag{6.2.4}$$

This Hamiltonian corresponds to harmonic oscillators that are coupled to one another, as discussed in Sects. 2.2.2 and 5.2, where the off–diagonal terms of the matrix $A = \{\alpha_{ij}\}$ designate the pairwise interactions. This Hamiltonian obeys the general form of Eq. (5.2.1)

While Einstein neglected the off–diagonal terms of A in the first place, Debye did not. In the following, we present the latter approach, which is more general (and more realistic), whereas the former will essentially be a special case.

6.3 Heat Capacity Analysis

The first idea of the analysis is to transform the coordinates into a new domain where the components are all decoupled. This means diagonalizing the matrix A. Since A is a symmetric non–negative definite matrix, it is clearly diagonalizable by a unitary matrix formed by its eigenvectors, and the diagonal elements of the diagonalized matrix (which are the eigenvalues of A) are non–negative. Let us denote the new coordinates of the system by q_i, $i = 1, \ldots, 3N$, and the eigenvalues – by $\frac{1}{2}m\omega_i^2$. By linearity of the differentiation operation, the same transformation take us from the vector of velocities $\{\dot{\xi}_i\}$ (of the kinetic component of the Hamiltonian) to the vector of derivatives of $\{q_i\}$, which will be denoted $\{\dot{q}_i\}$. Fortunately enough, since the transformation is unitary it leaves the components $\{\dot{q}_i\}$ decoupled. In other words, by the Parseval theorem, the norm of $\{\dot{\xi}_i\}$ is equal to the norm of $\{\dot{q}_i\}$. Thus, in the transformed domain, the Hamiltonian reads

$$\mathcal{E}(\boldsymbol{q}) = \Phi_0 + \frac{1}{2}m\sum_i (\dot{q}_i^2 + \omega_i^2 q_i^2). \tag{6.3.1}$$

which can be viewed as $3N$ *decoupled* harmonic oscillators, each one oscillating in its individual normal mode ω_i. The parameters $\{\omega_i\}$ are called *characteristic frequencies* or *normal modes*.

Example 6.1 (*One–dimensional ring of springs*) If the system has translational symmetry and if, in addition, there are periodic boundary conditions, then the matrix A is circulant, which means that it is always diagonalized by the discrete Fourier transform (DFT). In this case, q_i are the corresponding spatial frequency variables, conjugate to the location displacement variables ξ_i. The simplest example of this is a ring of N one–dimensional springs, as discussed in Sect. 5.2 (see left part of Fig. 5.1), where the Hamiltonian (in the current notation) is

$$\mathcal{E}(\boldsymbol{x}) = \Phi_0 + \frac{1}{2}m\sum_i \dot{\xi}_i^2 + \frac{1}{2}K\sum_i (\xi_{i+1} - \xi_i)^2. \tag{6.3.2}$$

In this case, the matrix A is given by

$$A = K \cdot \begin{pmatrix} 1 & -\frac{1}{2} & 0 & \dots & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & \dots & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & 0 & \dots & 0 & -\frac{1}{2} & 1 \end{pmatrix} \tag{6.3.3}$$

The eigenvalues of A are $\lambda_i = K[1 - \cos(2\pi i/N)]$, which are simply the DFT coefficients of the N–sequence formed by any row of A (removing the complex exponential of the phase factor). This means that the normal modes are $\omega_i = \sqrt{2K[1 - \cos(2\pi i/N)]/m}$. □

Classically, each of the $3N$ normal modes of vibration corresponds to a wave of distortion of the lattice. Quantum–mechanically, these modes give rise to quanta called *phonons*, in analogy to the fact that vibrational modes of electromagnetic waves give rise to photons. There is one important difference, however: while the number of normal modes in the case of an electromagnetic wave is infinite, here the number of modes (or the number of phonon energy levels) is finite – there are exactly $3N$ of them. This gives rise to a few differences in the physical behavior, but at low temperatures, where the high–frequency modes of the solid become unlikely to be excited, these differences become insignificant.

The Hamiltonian is then

$$E(n_1, n_2, \dots) = \Phi_0 + \sum_i \left(n_i + \frac{1}{2}\right)\hbar\omega_i, \tag{6.3.4}$$

where the non-negative integers $\{n_i\}$ denote the 'states of excitation' of the various oscillators, or equally well, the occupation numbers of the various phonon levels in the system. The internal energy is then

$$\begin{aligned} \langle E \rangle &= -\frac{\partial}{\partial\beta} \ln Z_{3N}(\beta) \\ &= -\frac{\partial}{\partial\beta} \ln\left(\sum_{n_1}\sum_{n_2}\dots\exp\left\{-\beta\left[\Phi_0 + \sum_i \left(n_i + \frac{1}{2}\right)\hbar\omega_i\right]\right\}\right) \\ &= -\frac{\partial}{\partial\beta} \ln\left[e^{-\beta\Phi_0}\prod_i \frac{e^{-\beta\hbar\omega_i/2}}{1 - e^{-\beta\hbar\omega_i}}\right] \end{aligned}$$

$$= \Phi_0 + \sum_i \frac{1}{2}\hbar\omega_i + \sum_i \frac{\partial}{\partial\beta}\ln(1 - e^{-\beta\hbar\omega_i})$$

$$= \Phi_0 + \sum_i \frac{1}{2}\hbar\omega_i + \sum_i \frac{\hbar\omega_i}{1 - e^{-\beta\hbar\omega_i}}. \tag{6.3.5}$$

Only the last term of the last expression depends on T. Thus, the heat capacity at constant volume[3] is:

$$C_V = \frac{\partial\langle E\rangle}{\partial T} = k\sum_i \frac{(\hbar\omega_i/kT)^2 e^{h\omega_i/kT}}{(e^{h\omega_i/kT} - 1)^2}. \tag{6.3.6}$$

To proceed from here, one has to know (or assume something) about the form of the density $g(\omega)$ of $\{\omega_i\}$ and then pass from summation to integration. It is this point where the difference between Einstein's approach and Debye's approach starts to show up.

6.3.1 Einstein's Theory

For Einstein, who assumed that the oscillators do not interact in the original, ξ–domain, all the normal modes are equal $\omega_i = \omega_E$ for all i, because then (assuming translational symmetry) A is proportional to the identity matrix and then all its eigenvalues are the same. Thus, in Einstein's model $g(\omega) = 3N\delta(\omega - \omega_E)$, and the result is

$$C_V = 3Nk\boldsymbol{E}(x) \tag{6.3.7}$$

where $\boldsymbol{E}(x)$ is the so–called the *Einstein function*:

$$\boldsymbol{E}(x) = \frac{x^2 e^x}{(e^x - 1)^2} \tag{6.3.8}$$

with

$$x = \frac{\hbar\omega_E}{kT} \stackrel{\Delta}{=} \frac{\Theta_E}{T}. \tag{6.3.9}$$

where $\Theta_E = \hbar\omega_E/k$ is called the *Einstein temperature*. At high temperatures $T \gg \Theta_E$, where $x \ll 1$ and then $\boldsymbol{E}(x) \approx 1$, we readily see that $C_V(T) \approx 3Nk$, in agreement with classical physics. For low temperatures, $C_V(T)$ falls exponentially fast as $T \to 0$. This theoretical rate of decay, however, is way too fast compared to

---

[3] **Exercise 6.1** Why is this the heat capacity at constant volume? Where is the assumption of constant volume being used here?.

the observed rate, which is cubic, as described earlier. But at least, Einstein's theory predicts the qualitative behavior correctly.

6.3.2 Debye's Theory

Debye (1912), on the other hand, assumed a continuous density $g(\omega)$. He assumed some cutoff frequency ω_D, so that

$$\int_0^{\omega_D} g(\omega)\mathrm{d}\omega = 3N. \tag{6.3.10}$$

For $g(\omega)$ in the range $0 \le \omega \le \omega_D$, Debye adopted a Rayleigh expression in the spirit of the one we saw in black–body radiation, but with a distinction between the longitudinal mode and the two independent transverse modes associated with the propagation of each wave at a given frequency. Letting v_L and v_T denote the corresponding velocities of these modes, this amounts to

$$g(\omega)\mathrm{d}\omega = V\left(\frac{\omega^2\mathrm{d}\omega}{2\pi^2 v_L^3} + \frac{\omega^2\mathrm{d}\omega}{\pi^2 v_T^3}\right). \tag{6.3.11}$$

This, together with the previous equation, determines the cutoff frequency to be

$$\omega_D = \left[\frac{18\pi^2\rho}{1/v_L^3 + 2/v_T^3}\right]^{1/3} \tag{6.3.12}$$

where $\rho = N/V$ is the density of the atoms. Accordingly,

$$g(\omega) = \begin{cases} \frac{9N}{\omega_D^3}\omega^2 & \omega \le \omega_D \\ 0 & \text{elsewhere} \end{cases} \tag{6.3.13}$$

The Debye formula for the heat capacity is now

$$C_V = 3Nk\boldsymbol{D}(x_0) \tag{6.3.14}$$

where $\boldsymbol{D}(\cdot)$ is called the *Debye function*

$$\boldsymbol{D}(x_0) = \frac{3}{x_0^3}\int_0^{x_0} \frac{x^4 e^x \mathrm{d}x}{(e^x - 1)^2} \tag{6.3.15}$$

with

$$x_0 = \frac{\hbar\omega_D}{kT} \triangleq \frac{\Theta_D}{T}, \tag{6.3.16}$$

where $\Theta_D = \hbar\omega_D/k$ is called the *Debye temperature*. Integrating by parts, the Debye function can also be written as

$$\boldsymbol{D}(x_0) = -\frac{3x_0}{e^{x_0}-1} + \frac{12}{x_0^3}\int_0^{x_0} \frac{x^3\mathrm{d}x}{e^x-1}. \tag{6.3.17}$$

Now, for $T \gg \Theta_D$, which means $x_0 \ll 1$, $D(x_0)$ can be approximated by a Taylor series expansion:

$$\boldsymbol{D}(x_0) = 1 - \frac{x_0^2}{20} + \ldots \tag{6.3.18}$$

Thus, for high temperatures, we again recover the classical result $C_V = 3Nk$. On the other hand, for $T \ll \Theta_D$, which is $x_0 \gg 1$, the dominant term in the integration by parts is the second one, which gives the approximation

$$\boldsymbol{D}(x_0) \approx \frac{12}{x_0^3}\int_0^{\infty} \frac{x^3\mathrm{d}x}{e^x-1} = \frac{4\pi^4}{5x_0^3} = \frac{4\pi^4}{5}\left(\frac{T}{\Theta_D}\right)^3. \tag{6.3.19}$$

Therefore, at low temperatures, the heat capacity is

$$C_V \approx \frac{12\pi^4}{5} Nk \left(\frac{T}{\Theta_D}\right)^3. \tag{6.3.20}$$

In other words, Debye's theory indeed recovers the T^3 behavior at low temperatures, in agreement with experimental evidence. Moreover, the match to experimental results is very good, not only near $T = 0$, but across a rather wide range of temperatures. In some textbooks, like [1, p. 164, Fig. 6.7], or [2, p. 177, Fig. 7.10], there are plots of C_V as a function of T for certain materials, which show impressive proximity between theory and measurements.

Exercise 6.2 Extend Debye's analysis to allow two different cutoff frequencies, ω_L and ω_T – for the longitudinal and the transverse modes, respectively.

Exercise 6.3 Calculate the density $g(\omega)$ for a ring of springs as described in Example 6.1. Write an expression for C_V as an integral and try to simplify it as much as you can.

6.4 Suggestions for Supplementary Reading

The exposition in this chapter is based largely on the books by Mandl [1] and Pathria [2]. Additional material appears also in Kardar [3, Sect. 6.2].

References

1. F. Mandl, *Statistical Physics* (Wiley, Chichester, 1971)
2. R.K. Pathria, *Statistical Mechanics*, 2nd edn. (Elsevier: Butterworth–Heinemann, Oxford, 1996)
3. M. Kardar, *Statistical Physics of Particles* (Cambridge University Press, Cambridge, 2007)

Chapter 7
Fluctuations, Stochastic Dynamics and Noise

So far we have discussed mostly systems in equilibrium. Extensive quantities like volume, energy, etc., have been calculated as means of certain ensembles, and these were not only means, but moreover, the values of high probability in the thermodynamic limit. In this chapter, we investigate the statistical fluctuations around these means, as well as dynamical issues, like the rate of approach to equilibrium when a system is initially away from equilibrium. We will also discuss noise generation mechanisms as well as their implications on electric circuits and other systems.

Historically, the theory of fluctuations has been interesting and useful because it made several experimental effects explicable. This was refreshing, considering the fact that late–nineteenth–century classical physicists were not able to explain these effects rigorously. One such phenomenon is Brownian motion – the irregular, random motion of very light particles suspended in a drop of liquid, which is observed using a microscope. Another phenomenon is electrical noise, such as thermal noise and shot noise, as mentioned in the previous paragraph.

Classical thermodynamics cannot explain fluctuations and, in fact, even denies their existence, because a fluctuation into a less probable state leads to a decrease of entropy, which is seemingly contradictory to the ideas of the consistent increase of entropy. This contradiction is resolved by the statistical–mechanical viewpoint, according to which the increase of entropy holds true only on the average (or with high probability), not deterministically. Apart from their theoretical interest, fluctuations are important to understand in order to make accurate measurements of physical properties and at the same time, to realize that the precision is limited by the fluctuations.

N. Merhav, *Statistical Physics for Electrical Engineering*,
DOI 10.1007/978-3-319-62063-3_7

7.1 Elements of Fluctuation Theory

So far, we have established probability distributions for various physical situations and have taken for granted the most likely value (or the mean value) as *the value* of the physical quantity of interest. For example, the internal energy in the canonical ensemble was taken to be $\langle E\rangle$, which is also the most likely value, with a very sharp peak as N grows.

The first question is what is the probabilistic characterization of the departure from the mean. One of the most natural measures of this departure is the variance, in the above example of the energy, $\text{Var}\{E\} = \left\langle E^2\right\rangle - \langle E\rangle^2$, or the relative standard deviation $\sqrt{\text{Var}\{E\}}/\langle E\rangle$. When several physical quantities are involved, then the covariances between them are also measures of fluctuation. There are two possible routes to assess fluctuations in this second order sense. The first is directly from the relevant ensemble, and the second is by a Gaussian approximation. It is emphasized that when it comes to fluctuations, the principle of ensemble equivalence no longer holds. For example, in the microcanonical ensemble, $\text{Var}\{E\} = 0$ (since E is fixed), whereas in the canonical ensemble, it is normally extensive, as we shall see shortly. Only $\sqrt{\text{Var}\{E\}}/\langle E\rangle$, which is proportional to $1/\sqrt{N}$ and hence tends to 0, can be considered asymptotically equivalent (in a very rough sense) to that of the microcanonical ensemble.

Consider a system in the canonical ensemble, and let us calculate the probability of energy level $\mathcal{E}(\boldsymbol{x}) = E$, which fluctuates from the mean E^*. Then,

$$\begin{aligned} P(E) &= \frac{\Omega(E)e^{-\beta E}}{Z(\beta)} \\ &\approx \frac{e^{-\beta[E-TS(E)]}}{e^{-\beta F}} \\ &= \frac{e^{-\beta[E-TS(E)]}}{e^{-\beta[E^*-TS(E^*)]}} \\ &= e^{-\beta[\Delta E-T\Delta S]}, \qquad \Delta E = E - E^*;\ \Delta S = S(E) - S(E^*). \end{aligned} \tag{7.1.1}$$

Now,

$$\begin{aligned} \Delta S &= \frac{\partial S}{\partial E}\Delta E + \frac{1}{2}\cdot\frac{\partial^2 S}{\partial E^2}\cdot(\Delta E)^2 + \ldots \\ &= \frac{1}{T}\cdot\Delta E + \frac{1}{2}\cdot\frac{\partial^2 S}{\partial E^2}\cdot(\Delta E)^2 + \ldots \end{aligned} \tag{7.1.2}$$

and so,

$$P(E) \approx e^{-\beta[\Delta E - T\Delta S]} \approx \exp\left\{\frac{\beta T}{2}\cdot\frac{\partial^2 S}{\partial E^2}\cdot(\Delta E)^2\right\}$$
$$= \exp\left\{\frac{1}{2k}\cdot\frac{\partial^2 S}{\partial E^2}\cdot(E - E^*)^2\right\}. \tag{7.1.3}$$

One should keep in mind that since $S(E)$ is concave, its second derivative is negative, so in the vicinity of E^*, the random variable $\mathcal{E}(X)$ is nearly *Gaussian* with mean E^* and variance $k/|\partial^2 S/\partial E^2|$. How does this variance scale with N? Note that since S and E are both extensive (proportional to N), the first derivative is intensive, and the second derivative is proportional to $1/N$, so the variance of E is proportional to N, which means that the standard deviation of the energy fluctuations scales like $\sqrt{N}$, and so the relative variance $\sqrt{\text{Var}\{E\}}/E$ scales like $1/\sqrt{N}$ (cf. the additive case, where $\mathcal{E}(X) = \sum_i \mathcal{E}(X_i)$ is the sum of N i.i.d. random variables). This asymptotic Gaussianity should not be a surprise, as we have approximated $F(E)$ by a second order Taylor series expansion around its minimum, so $e^{-\beta F(E)}$ is approximated by an exponentiated quadratic expression which is Gaussian. The same idea can be used for additional quantities that fluctuate. For example, in the Gibbsian ensemble, where both E and V fluctuate, the Gibbs free energy is nearly quadratic in $(\Delta E, \Delta V)$ around its equilibrium value, and so, this random vector is Gaussian with a covariance matrix that is proportional to the inverse of the Hessian of S w.r.t. E and V.

Example 7.1 (*Ideal gas*) In the case of the ideal gas, Eq. (2.2.6) gives

$$E(S, V) = \frac{3N^{5/3}h^2}{4\pi e^{5/3} m V^{2/3}} e^{2S/(3Nk)}, \tag{7.1.4}$$

whose Hessian is

$$\nabla^2 E = \frac{2E}{9}\begin{pmatrix} 2/(Nk)^2 & -2/(NkV) \\ -2/(NkV) & 5/V^2 \end{pmatrix}. \tag{7.1.5}$$

Thus, the covariance matrix of $(\Delta V, \Delta S)$ is

$$\Lambda = kT\cdot(\nabla^2 E)^{-1} = \frac{9kT}{2E}\cdot\frac{(NkV)^2}{6}\cdot\begin{pmatrix} 5/V^2 & 2/(NkV) \\ 2/(NkV) & 2/(Nk)^2 \end{pmatrix} \tag{7.1.6}$$

or, using the relation $E = 3NkT/2$,

$$\Lambda = \begin{pmatrix} 5Nk^2/2 & kV \\ kV & V^2/N \end{pmatrix}. \tag{7.1.7}$$

Thus, $\text{Var}\{\Delta S\} = 5Nk^2/2$, $\text{Var}\{\Delta V\} = V^2/N$, and $\langle \Delta S \Delta V\rangle = kV$, which are all extensive. □

7.2 Brownian Motion and the Langevin Equation

The term "Brownian motion" is after the botanist Robert Brown, who, in 1828, had been observing tiny pollen grains in a liquid under a microscope and saw that they moved in a random fashion, and that this motion was not triggered by currents or other processes in the liquid, like evaporation, etc. The movement was caused by frequent collisions with the particles of the liquid. Einstein (1905) was the first to provide a sound theoretical analysis of Brownian motion on the basis of the "random walk problem." Here, we introduce the topic using a formulation due to the French physicist Paul Langevin (1872–1946), which makes the derivation extremely simple. Langevin focuses on the motion of a relatively large particle of mass m, located at $x(t)$ at time t, whose velocity is $v(t) = \dot{x}(t)$. The particle is subjected to the influence of a force, composed of two components: one is a slowly varying macroscopic force and the other is varying rapidly and randomly. The latter has zero mean, but it fluctuates. In the one–dimensional case, it obeys the differential equation

$$m\ddot{x}(t) + \gamma\dot{x}(t) = F + F_{\mathrm{r}}(t), \tag{7.2.1}$$

where γ is a frictional (dissipative) coefficient and $F_{\mathrm{r}}(t)$ is the random component of the force. While this differential equation is nothing but Newton's law and hence obvious in macroscopic physics, it should not be taken for granted in the microscopic regime. In elementary Gibbsian statistical mechanics, all processes are time reversible in the microscopic level, since energy is conserved in collisions as the effect of dissipation in binary collisions is traditionally neglected. A reasonable theory, however, should incorporate the dissipative term.

The response of $x(t)$ to $F + F_{\mathrm{r}}(t)$ is, clearly, the superposition of the individual responses to F and to $F_{\mathrm{r}}(t)$ separately. The former is the solution to a simple (deterministic) differential equation, which is not the center of our interest here. Considering the response to $F_{\mathrm{r}}(t)$ only, multiply Eq. (7.2.1) by $x(t)$, to get

$$mx(t)\ddot{x}(t) \equiv m\left[\frac{\mathrm{d}(x(t)\dot{x}(t))}{\mathrm{d}t} - \dot{x}^2(t)\right] = -\gamma x(t)\dot{x}(t) + x(t)F_{\mathrm{r}}(t). \tag{7.2.2}$$

Taking the expectation, while assuming that, due to the randomness of $\{F_{\mathrm{r}}(t)\}$, $x(t)$ and $F_{\mathrm{r}}(t)$ at time t, are independent, we have, $\langle x(t)F_{\mathrm{r}}(t)\rangle = \langle x(t)\rangle \langle F_{\mathrm{r}}(t)\rangle = 0$. Also, note that $m\left\langle \dot{x}^2(t)\right\rangle = kT$ by the energy equipartition theorem (which applies here since we are assuming the classical regime), and so, we end up with

$$m\frac{\mathrm{d}\langle x(t)\dot{x}(t)\rangle}{\mathrm{d}t} = kT - \gamma \langle x(t)\dot{x}(t)\rangle, \tag{7.2.3}$$

a simple first order differential equation, whose solution is

$$\langle x(t)\dot{x}(t)\rangle = \frac{kT}{\gamma} + Ce^{-\gamma t/m}, \tag{7.2.4}$$

where C is a constant of integration. Imposing the condition that $x(0) = 0$, this gives $C = -kT/\gamma$, and so

$$\frac{1}{2}\frac{\mathrm{d}\langle x^2(t)\rangle}{\mathrm{d}t} \equiv \langle x(t)\dot{x}(t)\rangle = \frac{kT}{\gamma}\left(1 - e^{-\gamma t/m}\right), \tag{7.2.5}$$

which yields

$$\langle x^2(t)\rangle = \frac{2kT}{\gamma}\left[t - \frac{m}{\gamma}(1 - e^{-\gamma t/m})\right]. \tag{7.2.6}$$

The last equation gives the mean square deviation of a particle away from its origin, at time t. The time constant of the dynamics, a.k.a. the *relaxation time*, is $\theta = m/\gamma$. For short times $(t \ll \theta)$, $\langle x^2(t)\rangle \approx kTt^2/m$, which means that it looks like the particle is moving at constant velocity of $\sqrt{kT/m}$. For $t \gg \theta$, however,

$$\langle x^2(t)\rangle \approx \frac{2kT}{\gamma}\cdot t. \tag{7.2.7}$$

It should now be pointed out that this linear growth rate of $\langle x^2(t)\rangle$ is a characteristic of Brownian motion. Here it is only an approximation for $t \gg \theta$, as for $m > 0$, $\{x(t)\}$ is not a pure Brownian motion. Pure Brownian motion corresponds to the case $m = 0$ (hence $\theta = 0$), namely, the term $m\ddot{x}(t)$ in the Langevin equation can be neglected, and then $x(t)$ is simply proportional to $\int_0^t F_\mathrm{r}(\tau)\mathrm{d}\tau$ where $\{F_\mathrm{r}(t)\}$ is white noise. Figure 7.1 illustrates a few realizations of a Brownian motion in one dimension and in two dimensions.

We may visualize each collision on the pollen grain as that of an impulse, because the duration of each collision is extremely short. In other words, the position of the particle $x(t)$ is responding to a sequence of (positive and negative) impulses at random times. Let

$$R_v(\tau) = \langle v(t)v(t+\tau)\rangle = \langle \dot{x}(t)\dot{x}(t+\tau)\rangle \tag{7.2.8}$$

denote the autocorrelation of the random process $v(t) = \dot{x}(t)$ and let $S_v(\omega) = \mathcal{F}\{R_v(\tau)\}$ be the power spectral density.[1]

Clearly, by the Langevin equation $\{v(t)\}$ is the response of a linear, time–invariant linear system

$$H(s) = \frac{1}{ms+\gamma} = \frac{1}{m(s+1/\theta)}; \qquad h(t) = \frac{1}{m}e^{-t/\theta}u(t) \tag{7.2.9}$$

---

[1] To avoid confusion, one should keep in mind that although $S_v(\omega)$ is expressed as a function of the radial frequency ω, which is measured in radians per second, the physical units of the spectral density function itself here are Volt^2/Hz and not Volt^2/[radian per second]. To pass to the latter, one should divide by 2π. Thus, to calculate power, one must use $\int_{-\infty}^{+\infty} S_v(2\pi f)\mathrm{d}f$.

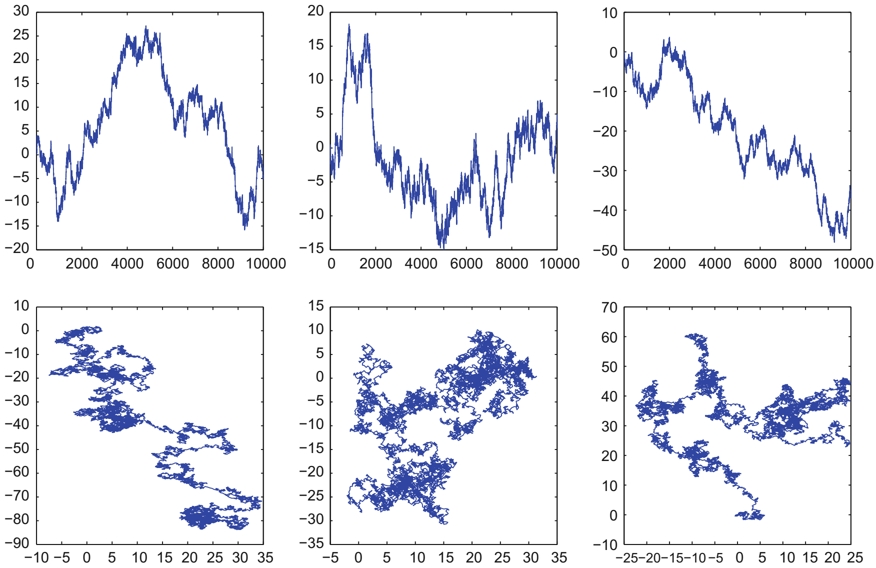


Fig. 7.1 Illustration of a Brownian motion. *Upper* figures: one–dimensional Brownian motion – three realizations of $x(t)$ as a function of t. *Lower* figures: two–dimensional Brownian motion – three realizations of $\vec{r}(t) = [x(t), y(t)]$. All realizations start at the origin

to the random input process $\{F_\mathrm{r}(t)\}$. Assuming that the impulse process $\{F_\mathrm{r}(t)\}$ is white noise, then

$$R_v(\tau) = \text{const} \cdot h(\tau) * h(-\tau) = \text{const} \cdot e^{-|\tau|/\theta} = R_v(0)e^{-|\tau|/\theta} = \frac{kT}{m} \cdot e^{-|\tau|/\theta} \tag{7.2.10}$$

and

$$S_v(\omega) = \frac{2kT}{m} \cdot \frac{\omega_0}{\omega^2 + \omega_0^2}, \qquad \omega_0 = \frac{1}{\theta} = \frac{\gamma}{m} \tag{7.2.11}$$

that is, a Lorentzian spectrum. We see that the relaxation time θ is indeed a measure of the "memory" of the particle and $\omega_0 = 1/\theta$ plays the role of 3 dB cutoff frequency of the spectrum of $\{v(t)\}$. What is the spectral density of the driving input white noise process?

$$S_{F_\mathrm{r}}(\omega) = \frac{S_v(\omega)}{|H(i\omega)|^2} = \frac{2kT\omega_0}{m(\omega^2 + \omega_0^2)} \cdot m^2(\omega^2 + \omega_0^2) = 2kTm\omega_0 = 2kT\gamma. \tag{7.2.12}$$

This result is very important. The spectral density of the white noise is $2kT$ times the dissipative coefficient of the system, γ. In other words, the dissipative element of the system is ‘responsible’ for the noise. At first glance, it may seem surprising: why should the intensity of the (external) driving force $F_{\mathrm{r}}(t)$ be related to the dissipative coefficient γ? The answer is that they are related via energy balance considerations, since we are assuming thermal equilibrium. Because the energy waste (dissipation) is proportional to γ, the energy supply from $F_{\mathrm{r}}(t)$ must also be proportional to γ in order to balance it.

Example 7.2 (*Energy balance for the Brownian particle*) The friction force $F_{\mathrm{friction}}(t) = -\gamma v(t)$ causes the particle to loose kinetic energy at the rate of

$$P_{\mathrm{loss}} = \langle F_{\mathrm{friction}}(t) v(t) \rangle = -\gamma \left\langle v^2(t) \right\rangle = -\gamma \cdot \frac{kT}{m} = -\frac{kT}{\theta}.$$

On the other hand, the driving force $F_{\mathrm{r}}(t)$ injects kinetic energy at the rate of

$$\begin{aligned} P_{\mathrm{injected}} &= \langle F_{\mathrm{r}}(t) v(t) \rangle && (7.2.13) \\ &= \left\langle F_{\mathrm{r}}(t) \int_0^\infty \mathrm{d}\tau h(\tau) F_{\mathrm{r}}(t-\tau) \right\rangle && (7.2.14) \\ &= 2kT\gamma \int_0^\infty \mathrm{d}\tau h(\tau)\delta(\tau) && (7.2.15) \\ &= kT\gamma h(0) = \frac{kT\gamma}{m} = \frac{kT}{\theta}, && (7.2.16) \end{aligned}$$

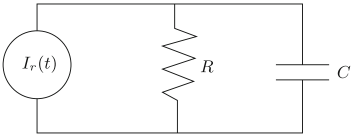
which exactly balances the loss. Here, we used that fact that $\int_0^\infty \mathrm{d}\tau h(\tau)\delta(\tau) = h(0)/2$ since only “half” of the delta function is “alive” where $h(\tau) > 0$.[2] **Exercise 7.1** What happens if $\gamma = 0$, yet $F_{\mathrm{r}}(t)$ has spectral density N_0? Calculate the rate of kinetic energy increase in two different ways: (i) Show that $\frac{m}{2}\left\langle v^2(t) \right\rangle$ is linear in t and find the constant of proportionality. (ii) Calculate $\langle F_{\mathrm{r}}(t) v(t) \rangle$ for this case.

These principles apply not only to a Brownian particle in a liquid, but to any linear system that obeys a first order stochastic differential equation with a white noise input, provided that the energy equipartition theorem applies. An obvious electrical analogue of this is a simple electric circuit where a resistor R and a capacitor C are connected to each other (see Fig. 7.2). The thermal noise generated by the resistor (due to the thermal random motion of the colliding free electrons in the conductor with extremely short mean time between collisions), a.k.a. the *Johnson–Nyquist noise*, is modeled as a current source connected in parallel to the resistor (or as an equivalent voltage source connected in series to the resistor), which generates a white noise current process $I_{\mathrm{r}}(t)$. The differential equation pertaining to Kirchoff’s current law is

---

[2] More rigorously, think of the delta function here as the limit of narrow (symmetric) autocorrelation functions which all integrate to unity.

Fig. 7.2 An *R–C* circuit



$$C\dot{V}(t) + \frac{V(t)}{R} = I_{\rm r}(t) \tag{7.2.17}$$

where $V(t)$ is the voltage across the resistor as well as the parallel capacitor. Now, this is exactly the same differential equation as before, where $I_{\rm r}(t)$ plays the role of the driving force, $V(t)$ is replacing $\dot{x}(t)$, C substitutes m, and $1/R$ is the dissipative coefficient instead of γ. Thus, the spectral density of the current is

$$S_{I_{\rm r}}(\omega) = \frac{2kT}{R}. \tag{7.2.18}$$

Alternatively, if one adopts the equivalent serial voltage source model then $V_{\rm r}(t) = RI_{\rm r}(t)$ and so

$$S_{V_{\rm r}}(\omega) = \frac{2kT}{R} \cdot R^2 = 2kTR. \tag{7.2.19}$$

This result is studied in every elementary course on random processes.

Finally, note that here we have something similar to the ultraviolet catastrophe: white noise has infinite power, which is nonphysical. Once again, this happens because we have not addressed quantum effects pertaining to high frequencies ($\hbar\omega \gg kT$), which as in black–body radiation, cause an exponential decay in the spectrum beyond a frequency of about $kT/\hbar$. We will get back to this later on in Sect. 7.5.

7.3 Diffusion and the Fokker–Planck Equation

In this subsection, we consider the temporal evolution of the probability density function of $x(t)$ (and not only its second order statistics, as in the previous subsection), under quite general conditions. The first successful treatment of Brownian motion was due to Einstein, who as mentioned earlier, reduced the problem to one of diffusion. Einstein's argument can be summarized as follows: assume that all particles move independently. The relaxation time is short compared to the observation time, but long enough for the motions of a particle in two consecutive intervals of θ to be independent. Let the number of suspended grains be N and let the x coordinate change by Δ in one relaxation time. Δ is a random variable, symmetrically distributed around 0. The number of particles $\mathrm{d}N$ displaced by more than Δ but less than $\Delta + \mathrm{d}\Delta$ is $\mathrm{d}N = Np(\Delta)\mathrm{d}\Delta$ where $p(\Delta)$ is the pdf of Δ. Since only small displacements are likely to occur, $p(\Delta)$ is sharply peaked at the origin. Let $\rho(x, t)$ denote the density of particles at position x, at time t. The number of particles in the interval $[x, x + \mathrm{d}x]$ at time $t + \delta$ (δ small) is

$$\rho(x, t + \delta)\mathrm{d}x = \mathrm{d}x \int_{-\infty}^{+\infty} \rho(x - \Delta, t)p(\Delta)\mathrm{d}\Delta \tag{7.3.1}$$

This equation tells that the probability of finding the particle around x at time $t + \delta$ is the probability of finding it in $x - \Delta$ (for any Δ) at time t, and then moving by Δ within duration δ to arrive at x at time $t + \delta$. Here we assume independence between the location $x - \Delta$ at time t and the probability distribution of Δ, as $p(\Delta)$ is independent of $x - \Delta$. Since δ is small, we use the Taylor series expansion

$$\rho(x, t + \delta) \approx \rho(x, t) + \delta \cdot \frac{\partial \rho(x, t)}{\partial t}. \tag{7.3.2}$$

Also, for small Δ, we approximate $\rho(x - \Delta, t)$, this time to the second order:

$$\rho(x - \Delta, t) \approx \rho(x, t) - \Delta \cdot \frac{\partial \rho(x, t)}{\partial x} + \frac{\Delta^2}{2} \cdot \frac{\partial^2 \rho(x, t)}{\partial x^2}. \tag{7.3.3}$$

Putting these in Eq. (7.3.1), we get

$$\begin{aligned} \rho(x, t) + \delta \cdot \frac{\partial \rho(x, t)}{\partial t} = \rho(x, t) \int_{-\infty}^{+\infty} p(\Delta)\mathrm{d}\Delta - \frac{\partial \rho(x, t)}{\partial x} \int_{-\infty}^{+\infty} \Delta p(\Delta)\mathrm{d}\Delta + \\ \frac{1}{2} \cdot \frac{\partial^2 \rho(x, t)}{\partial x^2} \int_{-\infty}^{+\infty} \Delta^2 p(\Delta)\mathrm{d}\Delta \end{aligned} \tag{7.3.4}$$

or

$$\frac{\partial \rho(x, t)}{\partial t} = \frac{1}{2\delta} \cdot \frac{\partial^2 \rho(x, t)}{\partial x^2} \int_{-\infty}^{+\infty} \Delta^2 p(\Delta)\mathrm{d}\Delta \tag{7.3.5}$$

which is the diffusion equation

$$\frac{\partial \rho(x, t)}{\partial t} = D \cdot \frac{\partial^2 \rho(x, t)}{\partial x^2} \tag{7.3.6}$$

with the diffusion coefficient being

$$D = \lim_{\delta \to 0} \frac{\langle \Delta^2 \rangle}{2\delta} = \lim_{\delta \to 0} \frac{\langle [x(t + \delta) - x(t)]^2 \rangle}{2\delta}. \tag{7.3.7}$$

To solve the diffusion equation, define $\varrho(\kappa, t)$ as the Fourier transform of $\rho(x, t)$ w.r.t. the variable x, i.e.,

$$\varrho(\kappa, t) = \int_{-\infty}^{+\infty} \mathrm{d}x \cdot e^{-i\kappa x} \rho(x, t). \tag{7.3.8}$$

Then, the diffusion equation becomes an ordinary differential equation w.r.t. t:

$$\frac{\varrho(\kappa, t)}{\partial t} = D(i\kappa)^2 \varrho(\kappa, t) \equiv -D\kappa^2 \varrho(\kappa, t) \tag{7.3.9}$$

whose solution is easily found to be $\varrho(\kappa, t) = C(\kappa)e^{-D\kappa^2 t}$. Assuming $\rho(x, 0) = \delta(x)$, this means $C(\kappa) = \varrho(\kappa, 0) = 1$ for all κ, and so $\varrho(\kappa, t) = e^{-D\kappa^2 t}$. The density $\rho(x, t)$ is obtained by the inverse Fourier transform, which is

$$\rho(x, t) = \frac{e^{-x^2/(4Dt)}}{\sqrt{4\pi Dt}}, \tag{7.3.10}$$

and so $x(t)$ is zero–mean Gaussian with variance $\left\langle x^2(t) \right\rangle = 2Dt$.[3] Of course, any other initial location x_0 would yield a Gaussian with the same variance $2Dt$, but the mean would be x_0. Comparing the variance $2Dt$ with (7.2.7), we have $D = kT/\gamma$, which is known as the *Einstein relation*, widely used in semiconductor physics.

The analysis thus far assumed that $\langle \Delta \rangle = 0$, namely, there is no drift to either the left or the right direction. We next drop this assumption. In this case, the diffusion equation generalizes to

$$\frac{\partial \rho(x, t)}{\partial t} = -v \cdot \frac{\partial \rho(x, t)}{\partial x} + D \cdot \frac{\partial^2 \rho(x, t)}{\partial x^2} \tag{7.3.11}$$

where

$$v = \lim_{\delta \to 0} \frac{\langle \Delta \rangle}{\delta} = \lim_{\delta \to 0} \frac{\langle x(t + \delta) - x(t) \rangle}{\delta} = \frac{\mathrm{d}}{\mathrm{d}t} \langle x(t) \rangle = \langle \dot{x}(t) \rangle \tag{7.3.12}$$

has the obvious meaning of the average velocity. Equation (7.3.11) is well known as the *Fokker–Planck equation*. The diffusion equation and the Fokker–Planck equation are very central in physics. As mentioned already in Chap. 1, they are fundamental in semiconductor physics, describing processes of propagation of concentrations of electrons and holes in semiconductor materials.

Exercise 7.2 Solve the Fokker–Planck equation and show that the solution is $\rho(x, t) = \mathcal{N}(vt, 2Dt)$. Explain the intuition.

It is possible to further extend the Fokker–Planck equation so as to allow the pdf of Δ to be location–dependent, that is, $p(\Delta)$ would be replaced by $p_x(\Delta)$, but the

---

[3] A point to think about: what is the intuition behind the resultant Gaussianity? We have not assumed any specific distribution in advance.

important point to retain is that given the present location $x(t) = x$, Δ would be independent of the earlier history of $\{x(t'), t' < t\}$, which means that $\{x(t)\}$ should be a *Markov process*. Consider then a general continuous–time Markov process defined by the transition probability density function $W_\delta(x'|x)$, which denotes the pdf of $x(t+\delta)$ at x' given that $x(t) = x$. A straightforward extension of the earlier derivation would lead to the following more general form[4]

$$\frac{\partial \rho(x,t)}{\partial t} = -\frac{\partial}{\partial x}[v(x)\rho(x,t)] + \frac{\partial^2}{\partial x^2}[D(x)\rho(x,t)], \tag{7.3.13}$$

where

$$v(x) = \lim_{\delta \to 0} \frac{1}{\delta} \int_{-\infty}^{+\infty} (x' - x) W_\delta(x'|x) \mathrm{d}x' = \boldsymbol{E}[\dot{x}(t)|x(t) = x] \tag{7.3.14}$$

is the local average velocity and

$$D(x) = \lim_{\delta \to 0} \frac{1}{2\delta} \int_{-\infty}^{+\infty} (x' - x)^2 W_\delta(x'|x) \mathrm{d}x' = \lim_{\delta \to 0} \frac{1}{2\delta} \boldsymbol{E}\{[x(t+\delta) - x(t)]^2 | x(t) = x\} \tag{7.3.15}$$

is the local diffusion coefficient.

Example 7.3 Consider the stochastic differential equation

$$\dot{x}(t) = -ax(t) + n(t),$$

where $n(t)$ is a Gaussian white noise with spectral density $N_0/2$. From the solution of this differential equation, it is easy to see that

$$x(t+\delta) = x(t)e^{-a\delta} + e^{-a(t+\delta)} \int_t^{t+\delta} \mathrm{d}\tau n(\tau) e^{a\tau}.$$

This relation, between $x(t)$ and $x(t+\delta)$, can be used to derive the first and the second moments of $[x(t+\delta) - x(t)]$ for small δ, and to find that $v(x) = -ax$ and $D(x) = N_0/4$ (**Exercise 7.4** Show this). Thus, the Fokker–Planck equation, in this example, reads

$$\frac{\partial \rho(x,t)}{\partial t} = a \cdot \frac{\partial}{\partial x}[x \cdot \rho(x,t)] + \frac{N_0}{4} \cdot \frac{\partial^2 \rho(x,t)}{\partial x^2}.$$

It is easy to check that the r.h.s. vanishes for $\rho(x,t) \propto e^{-2ax^2/N_0}$ (independent of t), which means that in equilibrium, $x(t)$ is Gaussian with zero mean and variance

---

[4] **Exercise 7.3** Prove it.

$N_0/4a$. This is in agreement with the fact that, as $x(t)$ is the response of the linear system $H(s) = 1/(s+a)$ (or in the time domain, $h(t) = e^{-at}u(t)$) to $n(t)$, its variance is indeed (as we know from courses on random processes):

$$\frac{N_0}{2}\int_0^\infty h^2(t)\mathrm{d}t = \frac{N_0}{2}\int_0^\infty e^{-2at}\mathrm{d}t = \frac{N_0}{4a}.$$

Note that if $x(t)$ is the voltage across the capacitor in a simple R–C network, then $a = 1/RC$, and since $\mathcal{E}(x) = Cx^2/2$, then in equilibrium we have the Boltzmann weight $\rho(x) \propto \exp(-Cx^2/2kT)$, which is again, a zero–mean Gaussian. Comparing the exponents, we immediately obtain $N_0/2 = 2kT/RC^2$.

Exercise 7.5 Find the solution $\rho(x,t)$ for all x and t subject to the initial condition $\rho(x,0) = \delta(x)$. □

A slightly different representation of the Fokker–Planck equation is the following:

$$\frac{\partial\rho(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left\{v(x)\rho(x,t) - \frac{\partial}{\partial x}[D(x)\rho(x,t)]\right\}. \tag{7.3.16}$$

Now, $v(x)\rho(x,t)$ has the obvious interpretation of the *drift current density* $J_{\text{drift}}(x,t)$ of a 'mass' whose density is $\rho(x,t)$ (in this case, it is a probability mass), whereas

$$J_{\text{diffusion}}(x,t) = -\frac{\partial}{\partial x}[D(x)\rho(x,t)]$$

is the *diffusion current density*.[5] While the drift current is related to the overall motion of the object, the diffusion current is associated with the tendency to equalize the density ρ (which is why it is proportional to the negative density gradient). Thus,

$$\frac{\partial\rho(x,t)}{\partial t} = -\frac{\partial J_{\text{total}}(x,t)}{\partial x} \tag{7.3.17}$$

where

$$J_{\text{total}}(x,t) = J_{\text{drift}}(x,t) + J_{\text{diffusion}}(x,t).$$

Equation (7.3.17) is the *equation of continuity*, which we saw in Chap. 1. In steady state, when $\rho(x,t)$ is time–invariant, the total current may vanish. The drift current and the diffusion current balance each other, or at least the net current is homogeneous (independent of x), so no mass accumulates anywhere.

---

[5] This generalizes Fick's law that we have seen in Chap. 1. There, D was fixed (independent of x), and so the diffusion current was proportional to the negative gradient of the density.

Comment It is interesting to relate the diffusion constant D to the mobility of the electrons in the context of electric conductivity. The mobility μ is defined according to $v = -\mu E$, where E is the electric field, and the minus sign is because electrons are accelerated in a direction opposite to that of the electric field (due to their negative charge). According to Fick's law, the diffusion current density is proportional to the negative gradient of the concentration, and D is defined as the constant of proportionality, i.e., $J_{\text{diff. elec.}} = -D\partial\rho/\partial x$, which is $J_{\text{diffusion}} = Dq_e\partial\rho/\partial x$, with the sign change, again, due to the negative charge of the electron. If one sets up a field E in an open circuit, the diffusion current cancels the drift current, that is

$$J = \rho q_e \mu E + Dq_e\frac{\partial\rho}{\partial x} = 0. \tag{7.3.18}$$

This gives $\rho(x) \propto e^{-\mu Ex/D}$. On the other hand, under thermal equilibrium, with potential energy $V(x) = q_e Ex$, we also have $\rho(x) \propto e^{-V/kT} = e^{-q_e Ex/kT}$. Upon comparing the exponents, we readily obtain the Einstein relation, $D = kT\mu/q_e$. Note that $\mu/q_e = |v|/q_e|E|$ is related to the admittance (the dissipative coefficient) since $|v|$ is proportional to the current and $|E|$ is proportional to the voltage.

7.4 The Fluctuation–Dissipation Theorem

We next take another point of view on stochastic dynamics of a physical system: suppose that a system (not necessarily a single particle as in the previous subsections) is initially in equilibrium of the canonical ensemble, but at a certain time instant, it is subjected to an abrupt, yet small, change in a certain parameter that controls it (say, a certain force, like pressure, magnetic field, etc.). Right after this abrupt change in the parameter, the system is, of course, no longer in equilibrium, but it is not far, since the change is assumed small. How fast does the system re–equilibrate and what is its dynamical behavior in the course of the passage to the new equilibrium? Also, since the change was abrupt and not quasi–static, how is energy dissipated? Quite remarkably, it turns out that the answers to both questions are related to the *equilibrium fluctuations* of the system. Accordingly, the principle that quantifies and characterizes this relationship is called the *fluctuation–dissipation theorem* and this is the subject of this subsection. We shall also relate it to the derivations of the previous subsection.

Consider a physical system, which in the absence of any applied external field, has an Hamiltonian $\mathcal{E}(x)$, where here x denotes the microstate, and so, its equilibrium distribution is the Boltzmann distribution with a partition function given by:

$$Z(\beta) = \sum_x e^{-\beta\mathcal{E}(x)}. \tag{7.4.1}$$

Now, let $w(x)$ be an observable (a measurable physical quantity that depends on the microstate), which has a conjugate force F, so that when F is applied, the change in the Hamiltonian is $\Delta\mathcal{E}(x) = -F \cdot w(x)$. Next, suppose that the external force is time–varying according to a certain waveform $\{F(t), \ -\infty < t < \infty\}$. As in the previous subsection, it should be kept in mind that the overall effective force can be thought of as a superposition of two contributions, a deterministic contribution, which is the above mentioned $F(t)$ – the external field that the experimentalist applies on purpose and fully controls, and a random part $F_{\mathrm{r}}(t)$, which pertains to interaction with the environment (or the heat bath at temperature T). The former is deterministic and the latter symbolizes the random, spontaneous thermal fluctuations.[6] The random component $F_{\mathrm{r}}(t)$ is responsible for the randomness of the microstate x and hence also the randomness of the observable. We shall denote the random variable corresponding to the observable at time t by $W(t)$. Thus, $W(t)$ is random variable, which takes values in the set $\{w(x), \ x \in \mathcal{X}\}$, where $\mathcal{X}$ is the space of microstates. When the external deterministic field is kept fixed ($F(t) \equiv$ const.), the system is expected to converge into equilibrium and eventually obey the Boltzmann law. While in the section on Brownian motion, we focused only on the contribution of the random part, $F_{\mathrm{r}}(t)$, now let us refer only to the deterministic part, $F(t)$. We will get back to the random part later on.

Let us assume first that $F(t)$ was switched on to a small level ϵ at time $-\infty$, and then switched off at time $t = 0$, in other words, $F(t) = \epsilon U(-t)$, where $U(\cdot)$ is the unit step function (a.k.a. the Heaviside function). We are interested in the behavior of the mean of the observable $W(t)$ at time t, which we shall denote by $\langle W(t)\rangle$, for $t > 0$. Also, $\langle W(\infty)\rangle$ will denote the limit of $\langle W(t)\rangle$ as $t \to \infty$, namely, the equilibrium mean of the observable in the absence of an external field. Define now the (negative) step response function as

$$\zeta(t) = \lim_{\epsilon \to 0} \frac{\langle W(t)\rangle - \langle W(\infty)\rangle}{\epsilon} \tag{7.4.2}$$

and the auto–covariance function pertaining to the final equilibrium as

$$R_W(\tau) \stackrel{\Delta}{=} \lim_{t \to \infty} \langle W(t)W(t+\tau)\rangle - \langle W(\infty)\rangle^2, \tag{7.4.3}$$

Then, the *fluctuation–dissipation theorem* (FDT) asserts that

$$R_W(\tau) = kT \cdot \zeta(\tau). \tag{7.4.4}$$

The FDT then relates between the linear transient response of the system to a small excitation (after it has been removed) and the autocovariance of the observable in

---

[6] The random part of the force $F_{\mathrm{r}}(t)$ does not necessarily exist physically, but it is a way to refer the random thermal fluctuations of the system to the 'input' F from a pure signals–and–systems perspective. For example, think again of the example a Brownian particle colliding with other particles. The other particles can be thought of as the environment in this case.

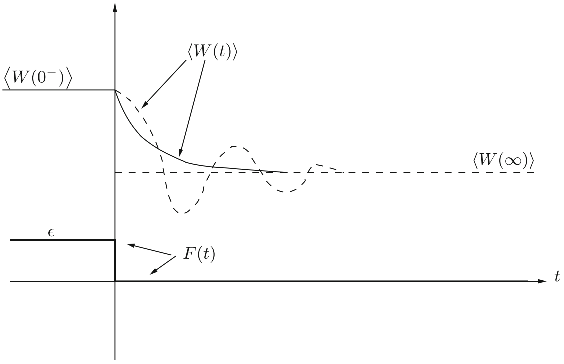


Fig. 7.3 Illustration of the response of $\langle W(t)\rangle$ to a step function at the input force $F(t) = \epsilon U(-t)$. According to the FDT, the response (on *top* of the asymptotic level $\langle W(\infty)\rangle$) is proportional to the equilibrium autocorrelation function $R_W(t)$, which in turn may decay either monotonically (*solid curve*) or in an oscillatory manner (*dashed curve*)

equilibrium. The transient response, that fades away is the dissipation, whereas the autocovariance is the fluctuation. Normally, $R_W(\tau)$ decays for large τ and so $\langle W(t)\rangle$ converges to $\langle W(\infty)\rangle$ at the same rate (see Fig. 7.3).

To prove this result, we proceed as follows: first, we have by definition:

$$\langle W(\infty)\rangle = \frac{\sum_x w(x)e^{-\beta\mathcal{E}(x)}}{\sum_x e^{-\beta\mathcal{E}(x)}}. \tag{7.4.5}$$

Now, for $t < 0$, we have

$$P(x) = \frac{e^{-\beta\mathcal{E}(x)-\beta\Delta\mathcal{E}(x)}}{\sum_x e^{-\beta\mathcal{E}(x)-\beta\Delta\mathcal{E}(x)}}. \tag{7.4.6}$$

Thus, for all negative times, and for $t = 0^-$ in particular, we have

$$\left\langle W(0^-)\right\rangle = \frac{\sum_x w(x)e^{-\beta\mathcal{E}(x)-\beta\Delta\mathcal{E}(x)}}{\sum_x e^{-\beta\mathcal{E}(x)-\beta\Delta\mathcal{E}(x)}}. \tag{7.4.7}$$

Let $P_t(x'|x)$ denote the probability that the system will be at state x' at time t ($t > 0$) given that it was at state x at time $t = 0^-$. This probability depends on the dynamical properties of the system (in the absence of the perturbing force). Let us define $\langle W(t)\rangle_x = \sum_{x'} w(x')P_t(x'|x)$, which is the expectation of $W(t)$ ($t > 0$) given that the system was at state x at $t = 0^-$. Now,

$$\begin{aligned}\langle W(t)\rangle &= \frac{\sum_x \langle W(t)\rangle_x\, e^{-\beta\mathcal{E}(x)-\beta\Delta\mathcal{E}(x)}}{\sum_x e^{-\beta\mathcal{E}(x)-\beta\Delta\mathcal{E}(x)}}\\ &= \frac{\sum_x \langle W(t)\rangle_x\, e^{-\beta\mathcal{E}(x)+\beta\epsilon w(x)}}{\sum_x e^{-\beta\mathcal{E}(x)+\beta\epsilon w(x)}}\end{aligned} \tag{7.4.8}$$

and $\langle W(\infty)\rangle$ can be seen as a special case of this quantity for $\epsilon = 0$ (no perturbation at all). Thus, $\zeta(t)$ is, by definition, nothing but the derivative of $\langle W(t)\rangle$ w.r.t. ϵ, computed at $\epsilon = 0$. I.e.,

$$\begin{aligned}\zeta(\tau) &= \frac{\partial}{\partial\epsilon}\left[\frac{\sum_x \langle W(\tau)\rangle_x\, e^{-\beta\mathcal{E}(x)+\beta\epsilon w(x)}}{\sum_x e^{-\beta\mathcal{E}(x)+\beta\epsilon w(x)}}\right]_{\epsilon=0}\\ &= \beta\cdot\frac{\sum_x \langle W(\tau)\rangle_x\, w(x)e^{-\beta\mathcal{E}(x)}}{\sum_x e^{-\beta\mathcal{E}(x)}} - \beta\cdot\frac{\sum_x \langle W(\tau)\rangle_x\, e^{-\beta\mathcal{E}(x)}}{\sum_x e^{-\beta\mathcal{E}(x)}}\cdot\frac{\sum_x w(x)e^{-\beta\mathcal{E}(x)}}{\sum_x e^{-\beta\mathcal{E}(x)}}\\ &= \beta\left[\lim_{t\to\infty}\langle W(t)W(t+\tau)\rangle - \langle W(\infty)\rangle^2\right]\\ &= \beta R_W(\tau),\end{aligned} \tag{7.4.9}$$

where we have used the fact that the dynamics of $\{P_t(x'|x)\}$ preserve the equilibrium distribution.

Exercise 7.6 Extend the FDT to account for a situation where the force $F(t)$ is not conjugate to $w(x)$, but to another physical quantity $v(x)$.

While $\zeta(t)$ is essentially the response of the system to a (negative) step function in $F(t)$, then obviously,

$$h(t) = \begin{cases} 0 & t < 0 \\ -\dot{\zeta}(t) & t \ge 0 \end{cases} = \begin{cases} 0 & t < 0 \\ -\beta\dot{R}_W(t) & t \ge 0 \end{cases} \tag{7.4.10}$$

would be the impulse response. Thus, we have characterized the “linear” system that describes the transient response of $\langle W(t)\rangle - \langle W(\infty)\rangle$ to a small input signal in $F(t)$. It is directly related to the equilibrium autocovariance function of $\{W(t)\}$. We can now express the response of $\langle W(t)\rangle$ to a general small signal $F(t)$ that vanishes for $t \ge 0$ to be

$$\begin{aligned}\langle W(t)\rangle - \langle W(\infty)\rangle &\approx -\beta\int_{-\infty}^{0}\dot{R}_W(t-\tau)F(\tau)\mathrm{d}\tau\\ &= -\beta\int_{-\infty}^{0}R_W(t-\tau)\dot{F}(\tau)\mathrm{d}\tau\\ &= -\beta R_W \otimes \dot{F},\end{aligned} \tag{7.4.11}$$

where the second passage is from integration by parts and where $\otimes$ denotes convolution. Indeed, in our first example, $\dot{F}(t) = -\epsilon\delta(t)$ and we are back with the result $\langle W(t)\rangle - \langle W(\infty)\rangle = \beta\epsilon R_W(t)$.

It is instructive to look at these relations also in the frequency domain. Applying the one sided Fourier transform on both sides of the relation $h(t) = -\beta\dot{R}_W(t)$ and taking the complex conjugate (i.e., multiplying by $e^{i\omega t}$ and integrating over $t > 0$), we get

$$H(-i\omega) \equiv \int_0^\infty h(t)e^{i\omega t}\mathrm{d}t = -\beta\int_0^\infty \dot{R}_W(t)e^{i\omega t}\mathrm{d}t = \beta i\omega\int_0^\infty R_W(t)e^{i\omega t}\mathrm{d}t + \beta R_W(0), \tag{7.4.12}$$

where the last step is due to integration by parts. Upon taking the imaginary parts of both sides, we get:

$$\mathrm{Im}\{H(-i\omega)\} = \beta\omega\int_0^\infty R_W(t)\cos(\omega t)\mathrm{d}t = \frac{1}{2}\beta\omega S_W(\omega), \tag{7.4.13}$$

where $S_W(\omega)$ is the power spectrum of $\{W(t)\}$ in equilibrium, that is, the Fourier transform of $R_W(\tau)$. Equivalently, we have:

$$S_W(\omega) = 2kT\cdot\frac{\mathrm{Im}\{H(-i\omega)\}}{\omega} = -2kT\cdot\frac{\mathrm{Im}\{H(i\omega)\}}{\omega} \tag{7.4.14}$$

Example 7.4 (*An electric circuit*) Consider the circuit in Fig. 7.4. The driving force is the voltage source $V(t)$ and the conjugate variable is $Q(t)$ the electric charge of the capacitor. The resistors are considered part of thermal environment. The voltage waveform is $V(t) = \epsilon U(-t)$. At time $t = 0^-$, the voltage across the capacitor is $\epsilon/2$ and the energy is $\frac{1}{2}C(V_\mathrm{r} + \frac{\epsilon}{2})^2$, whereas for $t \to \infty$, it is $\frac{1}{2}CV_\mathrm{r}^2$, so the difference is $\Delta\mathcal{E} = \frac{1}{2}CV_\mathrm{r}\epsilon = \frac{1}{2}Q_\mathrm{r}\epsilon$, neglecting the $O(\epsilon^2)$ term. According to the FDT then, $\zeta(t) = \frac{1}{2}\beta R_Q(t)$, where the factor of 1/2 follows the one in $\Delta\mathcal{E}$. This then gives

$$S_Q(\omega) = 4kT\cdot\frac{\mathrm{Im}\{H(-i\omega)\}}{\omega}. \tag{7.4.15}$$

In this case,

$$H(i\omega) = \frac{(R\|1/[i\omega C])\cdot C}{R + (R\|1/[i\omega C])} = \frac{C}{2 + i\omega RC} \tag{7.4.16}$$

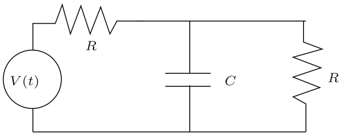


Fig. 7.4 Electric circuit for Example 7.4

for which

$$\mathrm{Im}\{H(-i\omega)\} = \frac{\omega RC^2}{4 + (\omega RC)^2} \tag{7.4.17}$$

and finally,

$$S_Q(\omega) = \frac{4kTRC^2}{4 + (\omega RC)^2}. \tag{7.4.18}$$

Thus, the thermal noise voltage across the capacitor is $4kTR/[4 + (\omega RC)^2]$. The same result can be obtained, of course, using the method studied in courses on random processes, where the voltage noise spectrum across a certain pair of points in the circuit is given by $2kT$ times the real part of the input impedance seen from these points, which in this case, is given by

$$2kT \cdot \mathrm{Re}\left\{R\|R\|\frac{1}{i\omega C}\right\} = \frac{4kTR}{4 + (\omega RC)^2}. \tag{7.4.19}$$

This concludes Example 7.4. □

Earlier, we have seen that $\langle W(t)\rangle - \langle W(\infty)\rangle$ responds to a deterministic (small) waveform $F(t)$ via a linear (or actually, linearized) system with an impulse response $h(t)$. By the same token, we can think of the random part around the mean, $W(t) - \langle W(t)\rangle$, as the response of the same system to a random input $F_\mathrm{r}(t)$ (thus, the total response is the superposition). In other words, we are decomposing the total "output signal" as

$$W(t) - \langle W(\infty)\rangle = [\langle W(t)\rangle - \langle W(\infty)\rangle] + [W(t) - \langle W(t)\rangle],$$

viewing the first bracketed term as the deterministic part, responding to the deterministic signal $F(t)$, and the second bracketed term as the random fluctuation, responding to a random input $F_\mathrm{r}(t)$. If we wish to think of our physical system in equilibrium as a linear(ized) system with input $F_\mathrm{r}(t)$ and output $W(t) - \langle W(t)\rangle$, then what should the spectrum of the input process $\{F_\mathrm{r}(t)\}$ be in order to comply with the last result? Denoting by $S_{F_\mathrm{r}}(\omega)$ the spectrum of the input process, we know from the basic of random processes that

$$S_W(\omega) = S_{F_\mathrm{r}}(\omega) \cdot |H(i\omega)|^2 \tag{7.4.20}$$

and so comparing with (7.4.14), we have

$$S_{F_\mathrm{r}}(\omega) = 2kT \cdot \frac{\mathrm{Im}\{H(-i\omega)\}}{\omega \cdot |H(i\omega)|^2}. \tag{7.4.21}$$

This extends our earlier result concerning the spectrum of the driving white noise in the case of the Brownian particle, where we obtained a spectral density of $2kT\gamma$.

Example 7.5 (*Second order linear system*) For a second order linear system (e.g., a damped harmonic oscillator),

$$m\ddot{W}(t) + \gamma\dot{W}(t) + KW(t) = F_{\rm r}(t) \qquad (7.4.22)$$

the force $F_{\rm r}(t)$ is indeed conjugate to the variable $W(t)$, which is the location, as required by the FDT. Here, we have

$$H(i\omega) = \frac{1}{m(i\omega)^2 + \gamma i\omega + K} = \frac{1}{K - m\omega^2 + \gamma i\omega}. \qquad (7.4.23)$$

In this case,

$$\text{Im}\{H(-i\omega)\} = \frac{\gamma\omega}{(K - m\omega^2)^2 + \gamma^2\omega^2} = \gamma\omega|H(i\omega)|^2 \qquad (7.4.24)$$

and so, we readily obtain

$$S_{F_{\rm r}}(\omega) = 2kT\gamma, \qquad (7.4.25)$$

recovering the principle that the spectral density of the noise process is $2kT$ times the dissipative coefficient γ of the system, which is responsible to the irreversible component. The difference between this and the earlier derivation is that earlier, we *assumed in advance* that the input noise process is white and we only computed its spectral level, whereas now, we have actually shown that at least for a second order linear system like this, it must be white noise (as far as the classical approximation holds). □

From Eq. (7.4.21), we see that the thermal interaction with the environment, when *referred to the input* of the system, has a spectral density of the form that we can calculate. In general, it does *not* necessarily have to be a flat spectrum. Consider for example, an arbitrary electric network consisting of one voltage source (in the role of $F(t)$) and several resistors and capacitors. Suppose that our observable $W(t)$ is the voltage across one of the capacitors. Then, there is a certain transfer function $H(i\omega)$ from the voltage source to $W(t)$. The thermal noise process stemming from all resistors is calculated by considering equivalent noise sources (parallel current sources or serial voltage sources) attached to each resistor. However, in order to refer the contribution of these noise sources to the input $F(t)$, we must calculate equivalent noise sources which are in series with the given voltage source $F(t)$. These equivalent noise sources will no longer generate white noise processes, in general. For example, in the circuit of Fig. 7.4, if an extra capacitor C would be connected in series to one of the resistors, then, the contribution of the right resistor referred to the left one is not white noise.[7]

---

[7] **Exercise 7.7** Calculate its spectrum.

Finally, it should be pointed out that this concept of referring the randomness in the system to the input is not always feasible, as in general, there is no apparent guarantee that the r.h.s. of Eq. (7.4.21) is a legitimate spectrum density function, i.e., that it is non–negative everywhere. In the absence of this condition, the idea of referring the noise to the input should simply be abandoned.

7.5 Johnson–Nyquist Noise in the Quantum–Mechanical Regime

As promised at the end of Sect. 7.3, we now return to the problematics of the formula $S_{V_{\mathrm{r}}}(\omega) = 2kTR$ when it comes to very high frequencies, namely, the electrical analogue to the ultraviolet catastrophe. Very high frequencies means very short waves, much shorter than the physical sizes of the electric circuit.

The remedy to the unreasonable classical results in the high frequency range, is to view the motion of electrons in a resistor as an instance of black–body radiation, but instead of the three–dimensional case that we studied earlier, this time we are talking about the one–dimensional case. The difference is mainly the calculation of the density of states. Consider a long transmission line with characteristic impedance R of length L, terminating at both ends by resistances R (see Fig. 7.5), so that the impedances are matched at both ends. Then any voltage wave propagating along the transmission line is fully absorbed by the terminating resistor without reflection. The system resides in thermal equilibrium at temperature T. The resistor then can be thought of as a black–body radiator in one dimension. A voltage wave of the form $V(x,t) = V_0 \exp[i(\kappa x - \omega t)]$ propagates along the transmission line with velocity $v = \omega/\kappa$, which depends on the capacitance and the inductance of the transmission line per unit length. To assess the number of modes, let us impose the periodic boundary condition $V(0,t) = V(L,t)$. Then $\kappa L = 2\pi n$ for any positive integer n. Thus, there are $\Delta n = L\Delta\kappa/2\pi = L\Delta\omega/2\pi v$ such modes in the frequency range between $\omega = v\kappa$ and $\omega + \Delta\omega = v(\kappa + \Delta\kappa)$. The mean energy of such a mode is given by

$$\epsilon(\omega) = \frac{\hbar\omega}{e^{\hbar\omega/kT} - 1}. \tag{7.5.1}$$

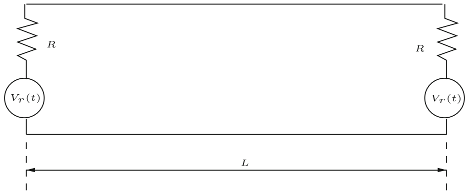


Fig. 7.5 Transmission line of length L, terminated by resistances R at both ends

Since there are $\Delta n = L\Delta\omega/(2\pi v)$ propagating modes in this frequency range, the mean energy per unit time (i.e., the power) incident upon a resistor in this frequency range is

$$P = \frac{1}{L/v} \cdot \frac{L\Delta\omega}{2\pi v} \cdot \epsilon(\omega) = \frac{1}{2\pi} \cdot \frac{\hbar\omega\Delta\omega}{e^{\hbar\omega/kT} - 1}, \tag{7.5.2}$$

where L/v at the denominator is the travel time of the wave along the transmission line. This is the radiation power absorbed by the resistor, which must be equal to the power emitted by the resistor in this frequency range. Let the thermal voltage generated by the resistor in the frequency range $[\omega, \omega + \Delta\omega]$ be denoted by $V_{\mathrm{r}}(t)[\omega, \omega + \Delta\omega]$. This voltage sets up a current of $V_{\mathrm{r}}(t)[\omega, \omega + \Delta\omega]/2R$ and hence an average power of $\left\langle V_{\mathrm{r}}^2(t)[\omega, \omega + \Delta\omega]\right\rangle/4R$. Thus, the balance between the absorbed and the emitted power gives

$$\frac{\left\langle V_{\mathrm{r}}^2(t)[\omega, \omega + \Delta\omega]\right\rangle}{4R} = \frac{1}{2\pi} \cdot \frac{\hbar\omega \cdot \Delta\omega}{e^{\hbar\omega/kT} - 1}, \tag{7.5.3}$$

which is

$$\frac{\left\langle V_{\mathrm{r}}^2(t)[\omega, \omega + \Delta\omega]\right\rangle}{\Delta\omega} = \frac{4R}{2\pi} \cdot \frac{\hbar\omega}{e^{\hbar\omega/kT} - 1} \tag{7.5.4}$$

or

$$\frac{\left\langle V_{\mathrm{r}}^2(t)[f, f + \Delta f]\right\rangle}{\Delta f} = 4R \cdot \frac{hf}{e^{hf/kT} - 1}. \tag{7.5.5}$$

Taking the limit $\Delta f \to 0$, the left–hand side becomes the one–sided spectral density of the thermal noise, and so (returning to the angular frequency domain), the two–sided spectral density is

$$S_{V_{\mathrm{r}}}(\omega) = 2R \cdot \frac{\hbar\omega}{e^{\hbar\omega/kT} - 1}. \tag{7.5.6}$$

We see that when quantum–mechanical considerations are incorporated, the noise spectrum is no longer flat. As long as $\hbar\omega \ll kT$, the denominator is very well approximated by $\hbar\omega/kT$, and we recover the formula $2kTR$, but for frequencies of the order of magnitude of $\omega_c \stackrel{\Delta}{=} kT/\hbar$, the spectrum decays exponentially rapidly. So the quantum–mechanical correction is the substitution:

$$kT \Longrightarrow \frac{\hbar\omega}{e^{\hbar\omega/kT} - 1}. \tag{7.5.7}$$

At $T = 300\,^\circ\mathrm{K}$, the cutoff frequency is $f_c = \omega_c/2\pi \approx 6.2\,\mathrm{THz}$,[8] so the spectrum can be safely considered $2kTR$ flat over any frequency range of practical interest.

What is the total RMS noise voltage generated by a resistor R at temperature T? The total mean square noise voltage is

$$\begin{aligned}\left\langle V_\mathrm{r}^2(t)\right\rangle &= 4R\int_0^\infty \frac{\hbar\omega\mathrm{d}\omega}{e^{\hbar\omega/kT}-1}\\ &= \frac{4R(kT)^2}{\hbar}\int_0^\infty \frac{x\mathrm{d}x}{e^x-1}\\ &= \frac{4R(kT)^2}{\hbar}\int_0^\infty \frac{xe^{-x}\mathrm{d}x}{1-e^{-x}}\\ &= \frac{4R(kT)^2}{\hbar}\sum_{n=1}^\infty\int_0^\infty xe^{-nx}\mathrm{d}x\\ &= \frac{4R(kT)^2}{\hbar}\sum_{n=1}^\infty\frac{1}{n^2}\\ &= \frac{2R(\pi kT)^2}{3\hbar},\end{aligned} \tag{7.5.8}$$

which is quadratic in T since both the (low frequency) spectral density and the effective bandwidth are linear in T. The RMS is then

$$V_\mathrm{RMS} = \sqrt{\left\langle V_\mathrm{r}^2(t)\right\rangle} = \sqrt{\frac{2R}{3\hbar}}\cdot\pi kT, \tag{7.5.9}$$

namely, proportional to temperature and to the square root of the resistance. To assess the order of magnitude, a resistor of $100\,\Omega$ at $T = 300\,^\circ\mathrm{K}$ generates an RMS thermal noise of about 10 mV when it stands alone (without a circuit that limits the bandwidth much more drastically than ω_c). The equivalent noise bandwidth is

$$B_\mathrm{eq} = \frac{2R(\pi kT)^2/3h}{2kTR} = \frac{\pi^2 kT}{3h} = \frac{\pi^2}{3}\cdot f_c. \tag{7.5.10}$$

Exercise 7.8 Derive an expression for the autocorrelation function of the Johnson–Nyquist noise in the quantum mechanical regime.

---

[8] Recall that $1\,\mathrm{THz} = 10^{12}\,\mathrm{Hz}$.

7.6 Other Noise Sources

In addition to thermal noise, there are other physical mechanisms that generate noise in Nature in general, and in electronic circuits, in particular. We will only provide short descriptions here. The interested reader is referred to the course notes “Fundamentals of Noise Processes” by Y. Yamamoto in the following link: `http://www.nii.ac.jp/qis/first-quantum/e/forStudents/lecture/index.html` These notes contain a very detailed and comprehensive account of many more topics that evolve around the physics of noise processes in electronic circuitry and other systems.

Flicker Noise

Flicker noise, also known as $1/f$ noise, is a random process with a spectrum that falls off steadily into the higher frequencies. It occurs in almost all electronic devices, and results from a variety of effects, though always related to a direct current. According to the underlying theory, there are fluctuations in the conductivity due to the superposition of many independent thermal processes of alternate excitation and relaxation of certain defects (e.g., dopant atoms or vacant lattice sites). This means that every once in a while, a certain lattice site or a dopant atom gets excited and it moves into a state of higher energy for some time, and then it relaxes back to the lower energy state until the next excitation. Each one of these excitation/relaxation processes can be modeled as a random telegraph signal (RTS) with a different time constant θ (due to different physical/geometric characteristics) and hence contributes a Lorentzian spectrum parametrized by θ. The superposition of these processes, whose spectrum is given by the integral of the Lorentzian function over a range of values of θ (with a certain weight), gives rise to the $1/f$ behavior over a wide range of frequencies. To see this more concretely in the mathematical language, an RTS $X(t)$ is given by $X(t) = (-1)^{N(t)}$, where $N(t)$ is a Poisson process of rate λ. It is a binary signal where the level $+1$ can symbolize excitation and the level -1 designates relaxation. Here the dwell times between jumps are exponentially distributed. The autocorrelation function is given by

$$\begin{aligned}\langle X(t)X(t+\tau)\rangle &= \left\langle (-1)^{N(t)+N(t+\tau)} \right\rangle \\ &= \left\langle (-1)^{N(t+\tau)-N(t)} \right\rangle \\ &= \left\langle (-1)^{N(\tau)} \right\rangle \\ &= e^{-\lambda\tau} \sum_{k=0}^{\infty} \frac{(\lambda\tau)^k}{k!} \cdot (-1)^k \\ &= e^{-\lambda\tau} \sum_{k=0}^{\infty} \frac{(-\lambda\tau)^k}{k!} \\ &= e^{-2\lambda\tau} \qquad (7.6.1)\end{aligned}$$

and so the spectrum is Lorentzian:

$$S_X(\omega) = \mathcal{F}\{e^{-2\lambda|\tau|}\} = \frac{4\lambda}{\omega^2 + 4\lambda^2} = \frac{2\theta}{1 + (\omega\theta)^2}, \tag{7.6.2}$$

where the time constant is $\theta = 1/2\lambda$ and the cutoff frequency is $\omega_c = 2\lambda$. Now, calculating the integral

$$\int_{\theta_{\min}}^{\theta_{\max}} \mathrm{d}\theta \cdot g(\theta) \cdot \frac{2\theta}{1 + (\omega\theta)^2}$$

with $g(\theta) = 1/\theta$, yields a composite spectrum that is proportional to

$$\frac{1}{\omega} \tan^{-1}(\omega\theta_{\max}) - \frac{1}{\omega} \tan^{-1}(\omega\theta_{\min}).$$

For $\omega \ll 1/\theta_{\max}$, using the approximation $\tan^{-1}(x) \approx x$ ($|x| \ll 1$), this is approximately a constant. For $\omega \gg 1/\theta_{\min}$, using the approximation $\tan^{-1}(x) \approx \frac{\pi}{2} - \frac{1}{x}$ ($|x| \gg 1$), this is approximately proportional to $1/\omega^2$. In between, in the range $1/\theta_{\max} \ll \omega \ll 1/\theta_{\min}$ (assuming that $1/\theta_{\max} \ll 1/\theta_{\min}$), the behavior is according to

$$\frac{1}{\omega}\left(\frac{\pi}{2} - \frac{1}{\omega\theta_{\max}}\right) - \theta_{\min} = \frac{1}{\omega}\left(\frac{\pi}{2} - \frac{1}{\omega\theta_{\max}} - \omega\theta_{\min}\right) \approx \frac{\pi}{2\omega},$$

which is the $1/f$ behavior in this wide range of frequencies. There are several theories why $g(\theta)$ should be inversely proportional to θ, but the truth is that they are not perfect, and the issue of $1/f$ noise is not yet perfectly (and universally) understood.

Shot Noise

Shot noise in electronic devices consists of unavoidable random statistical fluctuations of the electric current in an electrical conductor. Random fluctuations are inherent when current flows, as the current is a flow of discrete charges (electrons).

First, some background on Poisson processes: a Poisson process $\{N(t)\}_{t\geq 0}$ is a continuous–time counting process, starting from $N(0) = 0$ and incremented by 1 at random time instants $T_1, T_2, \ldots$. The number of events $N(t)$, counted up to time t, is distributed according to

$$\Pr\{N(t) = k\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad k = 0, 1, 2, \ldots \tag{7.6.3}$$

and events counted at non–overlapping time intervals are statistically independent. Thus, over a total time interval of t_0 seconds, the joint probability of $N(t_0) = k$ together with counting event times within $[\tau_1, \tau_1 + \mathrm{d}\tau_1] \times \ldots \times [\tau_k, \tau_k + \mathrm{d}\tau_k]$ is given by

$$\Pr\{T_1 \in [\tau_1, \tau_1 + \mathrm{d}\tau_1), \ldots, T_k \in [\tau_k, \tau_k + \mathrm{d}\tau_k),\ N(t_0) = k\} \tag{7.6.4}$$

$$\stackrel{\Delta}{=} f(\tau_1, \ldots, \tau_k,\ N(t_0) = k)\mathrm{d}\tau_1 \cdots \mathrm{d}\tau_k \tag{7.6.5}$$

$$= e^{-\lambda\tau_1} \cdot \lambda\mathrm{d}\tau_1 \cdot e^{-\lambda(\tau_2-\tau_1)} \cdot \lambda\mathrm{d}\tau_2 \cdots \lambda\mathrm{d}\tau_k e^{-\lambda(t_0-\tau_k)}$$

$$= e^{-\lambda t_0}\lambda^k \cdot \mathrm{d}\tau_1 \cdots \mathrm{d}\tau_k. \tag{7.6.6}$$

Therefore, by the Bayes theorem

$$f(\tau_1, \ldots, \tau_k | N(t_0) = k) = \frac{e^{-\lambda t_0}\lambda^k}{e^{-\lambda t_0}(\lambda t_0)^k/k!} \tag{7.6.7}$$

$$= k! \cdot \left(\frac{1}{t_0}\right)^k. \tag{7.6.8}$$

Consider k independent random variables, $\Theta_1, \ldots, \Theta_k$, all uniformly distributed within the interval $[0, t_0]$. Their joint density is, of course $(1/t_0)^k$, which is similar to the above except the factor $k!$. But $T_1, T_2, \ldots, T_k$ are ordered in increasing order, whereas $\Theta_1, \Theta_2, \ldots, \Theta_k$ are not necessarily so. One can think of $T_1, T_2, \ldots, T_k$ as a result of *ordering* $\Theta_1, \Theta_2, \ldots, \Theta_k$ in increasing order, and since there are $k!$ possible orderings, this gives rise to the factor $k!$ in $f(\tau_1, \ldots, \tau_k|N(t_0) = k)$. This means that one can simulate a Poisson process $\{N(t),\ 0 \le t < t_0\}$ as follows: (i) First, randomly select k according to the Poisson distribution (7.6.3). (ii) Draw $\Theta_1, \ldots, \Theta_k$ independently and uniformly at random over $[0, t_0]$. (iii) Sort $\Theta_1, \Theta_2, \ldots, \Theta_k$ in an increasing order to obtain $T_1, T_2, \ldots, T_k$. (iv) Let $N(t) = \sum_{i=1}^k u(t - T_i) = \sum_{i=1}^k u(t - \Theta_i)$ for $t \in [0, t_0)$.

Now, consider a DC current in a device, which shoots electrons according to a Poisson process (e.g., a p-n junction of a diode), i.e.,

$$I(t) = \sum_i i_e(t - T_i), \tag{7.6.9}$$

where $i_e(\cdot)$ is the (very short) current pulse generated by the passage of a single electron.[9] The DC current is simply the average of this, which is $\lambda t_0 q_e/t_0 = \lambda q_e \stackrel{\Delta}{=} I_0$. The noise, which is associated with the fluctuations around this average, is given by the second order statistics. Neglecting edge effects, we have:

$$R(s) \stackrel{\Delta}{=} \boldsymbol{E}\{I(t)I(t+s)\} = \boldsymbol{E}\{\boldsymbol{E}\{I(t)I(t+s)|K\}\} \tag{7.6.10}$$

$$= \boldsymbol{E}\left\{\sum_{i=1}^K i_e(t - \Theta_i) \sum_{j=1}^K i_e(t + s - \Theta_j)\right\} \tag{7.6.11}$$

---

[9] Note that $i_e(t)$ integrates to q_e, and since it is a very short pulse, it is nearly $q_e\delta(t)$ for a passage at time $t = 0$.

$$= \boldsymbol{E}\left\{\sum_{i=1}^{K} \frac{1}{t_0} \int_0^{t_0} i_e(t+s-\theta) i_e(t-\theta) \mathrm{d}\theta\right\} + \tag{7.6.12}$$

$$\boldsymbol{E}\left\{\sum_{i \neq j} \frac{1}{t_0^2} \int_0^{t_0} i_e(t+s-\theta) \mathrm{d}\theta \cdot \int_0^{t_0} i_e(t-\theta) \mathrm{d}\theta\right\} \tag{7.6.13}$$

$$= \frac{\boldsymbol{E}\{K\}}{t_0} R_e(s) + \frac{\boldsymbol{E}\{K^2 - K\}}{t_0^2} \cdot q_e^2 \tag{7.6.14}$$

$$= \frac{\lambda t_0}{t_0} \cdot R_e(s) + \frac{\lambda^2 t_0^2}{t_0^2} q_e^2 \tag{7.6.15}$$

$$= \frac{I_0}{q_e} \cdot R_e(s) + I_0^2, \tag{7.6.16}$$

where

$$R_e(s) = \int_{-\infty}^{+\infty} i_e(t) i_e(t+s) \mathrm{d}t. \tag{7.6.17}$$

Now, the second term, I_0^2, is the contribution of the pure DC component, i.e., the (stationary) average current. The first term is the fluctuation noise. Note that for $i_e(t) = q_e \delta(t)$, we have $R_e(s) = q_e^2 \delta(s)$, and so, the (flat) spectrum of the noisy part is

$$S_{\text{shot}}(\omega) = I_0 q_e, \tag{7.6.18}$$

or $S(\omega) = 2 I_0 q_e$ for the single–sided spectrum. A few comments are in order:

1. By measuring the noise intensity in a diode, one can find q_e experimentally.
2. In the derivation above, we assumed that $i_e(t)$ is proportional to the Dirac delta function, which is an idealization. For a general pulse shape, $S_{\text{shot}}(\omega)$ would become proportional to the Fourier transform of $R_e(s)$ as defined in (7.6.17). Equivalently, this can be thought of as letting the white noise process derived above undergo a linear filter whose impulse response is $i_e(t)$.
3. The result applies as long as I_0 is not too large. For a strong DC current I_0, there is another effect that kicks in, namely, the *spatial charge effect*: if a large bulk of electrons cross at the same time, they create a spatial charge that interferes with the emission of additional electrons, and this causes the shot noise spectral level to be smaller than predicted by the above derivation.

Burst Noise

Burst noise consists of sudden step–like transitions between two or more levels (non-Gaussian), as high as several hundred micro-volts, at random and unpredictable times. Each shift in offset voltage or current lasts for several milliseconds, and the intervals between pulses tend to be in the audio range (less than 100 Hz), leading to the term *popcorn noise* for the popping or crackling sounds it produces in audio circuits. Burst noise is customarily modeled as an RTS and therefore, another synonym for burst noise is *RTS noise*. Accordingly, it has a Lorentzian spectrum, similarly as in (7.6.2):

$$S_{\text{burst}}(\omega) \propto \frac{1}{1+(\omega/\omega_0)^2}, \tag{7.6.19}$$

which means that the spectrum is nearly flat at low frequencies (compared to the cutoff frequency ω_0) and nearly proportional to $1/\omega^2$ for high frequencies.

Avalanche Noise

Avalanche noise is the noise produced when a junction diode is operated at the onset of avalanche breakdown, a semiconductor junction phenomenon in which carriers in a high voltage gradient develop sufficient energy to dislodge additional carriers through physical impact, creating ragged current flows.

7.7 Suggestions for Supplementary Reading

Parts of the material in this chapter are based on Beck [1, Chaps. 6 and 9] and Reif [2, Chap. 15]. For additional recommended reading, the reader is referred to van Kampen [3], Risken [4], and Sethna [5, Chap. 10].

References

1. A.H.W. Beck, *Statistical Mechanics, Fluctuations and Noise* (Edward Arnold Publishers, London, 1976)
2. F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1965)
3. N.G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North Holland, Amsterdam, 1992)
4. H. Risken, *The Fokker–Planck Equation - Methods of Solution and Applications*, 2nd edn. (Springer, Berlin, 1989)
5. J.P. Sethna, *Statistical Mechanics: Entropy, Order Parameters, and Complexity* (Oxford University Press, Oxford, 2006)

Chapter 8
A Brief Touch on Information Theory*

8.1 Introduction – What Is Information Theory About?

Our last topic in this book consists of a very brief description on the relation between statistical physics and information theory, a research field pioneered by Claude Elwood Shannon (1916–2001), whose seminal paper "A Mathematical Theory of Communications" (1948) has established the corner–stone of this field.

In a nutshell, information theory is a science that focuses on the *fundamental limits*, on the one hand, and the *achievable performance*, on the other hand, concerning various information processing tasks, including most notably:

1. Data compression (lossless/lossy).
2. Error correction coding (coding for protection against errors due to channel noise).
3. Encryption.

There are also additional tasks of information processing that are considered to belong under the umbrella of information theory, like: signal detection, estimation (parameter estimation, filtering/smoothing, prediction), information embedding, process simulation, extraction of random bits, information relaying, and more.

Core information theory, which is called *Shannon theory* in the jargon of the professionals, is about *coding theorems*. It is associated with the development of computable formulas that characterize the best performance that can possibly be achieved in these information processing tasks under some (usually simple) assumptions on the probabilistic models that govern the data, the channel noise, the side information, the jammers if applicable, etc. While in most cases, this theory does not suggest constructive communication systems, it certainly provides insights concerning the features that an optimal (or nearly optimal) communication system must have. Shannon theory serves, first and foremost, as the theoretical basis for modern digital communication engineering. That being said, much of the modern research activity in information theory evolves, not only around Shannon theory, but also on the never-ending efforts to develop methodologies (mostly, specific code structures and

N. Merhav, *Statistical Physics for Electrical Engineering*,
DOI 10.1007/978-3-319-62063-3_8

algorithms) for designing very efficient communication systems, which hopefully come close to the bounds and the fundamental performance limits.

But the scope of information theory it is not limited merely to communication engineering: it plays a role also in computer science, and many other disciplines, one of them is thermodynamics and statistical mechanics, which is the focus of this last chapter. Often, information–theoretic problems are well approached from a statistical–mechanical point of view. We will taste this very briefly in two examples of problems.

In this book, we will not delve into information theory too deeply, but our purpose is merely to touch upon the interface of these two fields. A more advanced exposition, that goes much deeper than our scope here, is provided in [1].

8.2 Entropy in Information Theory and Statistical Physics

Perhaps the first relation that crosses one's mind is that in both fields there is a fundamental notion of *entropy*. Actually, in information theory, the term entropy was coined in the footsteps of the thermodynamic/statistical–mechanical entropy. Along this book, we have seen already three (seemingly) different forms of the entropy: the first is the thermodynamic entropy defined, in its differential form as

$$\delta S = \delta Q/T, \tag{8.2.1}$$

which was first introduced by Clausius in 1850. The second is the statistical entropy

$$S = k \ln \Omega, \tag{8.2.2}$$

which was defined by Boltzmann in 1872. The third is yet another formula for the entropy – the Gibbs formula for the entropy of the canonical ensemble:

$$S = -k \sum_x P(x) \ln P(x) = -k \langle \ln P(x) \rangle, \tag{8.2.3}$$

which we encountered in Chap. 2.

It is virtually impossible to miss the functional resemblance between the last form above and the information–theoretic entropy, a.k.a. the *Shannon entropy*, which is simply

$$H = -\sum_x P(x) \log_2 P(x) \tag{8.2.4}$$

namely, the same expression as above exactly, just without the factor k and with the basis of the logarithm being 2 rather than e. Indeed, this clear analogy was recognized already by Shannon and von Neumann. According to a well–known anecdote, von

Neumann advised Shannon to adopt this term because it would provide him with *"... a great edge in debates because nobody really knows what entropy is anyway."*

What is the information–theoretic meaning of entropy? It turns out that it has many information–theoretic meanings, but the most fundamental one concerns *optimum compressibility* of data. Suppose that we have a string of N i.i.d. random variables, $x_1, x_2, \ldots, x_N$, taking values in a discrete set, say, the components of the microstate in a quantum system of non–interacting particles, and we want to represent the microstate information digitally (in bits) as compactly as possible, without losing any information – in other words, we require the ability to fully reconstruct the data from the compressed binary representation. How short can this binary representation be?

Let us look at the following example. Suppose that each x_i takes values in the set $\{A, B, C, D\}$, independently with probabilities

$$P(A) = 1/2; \quad P(B) = 1/4; \quad P(C) = 1/8; \quad P(D) = 1/8.$$

Clearly, when translating the letters into bits, the naive approach would be to say the following: we have 4 letters, so it takes 2 bits to distinguish between them, by mapping, say lexicographically, as follows:

$$A \to 00; \quad B \to 01; \quad C \to 10; \quad D \to 11.$$

This would mean representing the list of x's using 2 bits per–symbol. This is very simple. But is this the best thing one can do?

It turns out that the answer is negative. Intuitively, if we can assign variable–length code-words to the various letters, using shorter code-words for more probable symbols and longer code-words for the less frequent ones, we might be able to gain something. In our example, A is most probable, while C and D are the least probable, so how about the following solution:

$$A \to 0; \quad B \to 10; \quad C \to 110; \quad D \to 111.$$

Now the average number of bits per symbol is:

$$\frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 = 1.75.$$

We have improved the average bit rate by 12.5%. This is fine, but is this the best one can do or can we improve even further?

It turns out that this time, the answer is affirmative. Note that in this solution, each letter has a probability of the form $2^{-\ell}$ (ℓ – positive integer) and the length of the assigned code-word is exactly ℓ (for A, $\ell = 1$, for B – $\ell = 2$, and for C and D, $\ell = 3$). In other words, the length of the code-word for each letter is the negative logarithm of its probability, so the average number of bits per symbol is $\sum_{x \in \{A,B,C,D\}} P(x)[-\log_2 P(x)]$, which is exactly the entropy H of the information source. One of the basic coding theorems of information theory tells us that we cannot compress to any coding rate below the entropy and still expect to be able to reconstruct the x's perfectly. But why is this true?

We will not get into a rigorous proof of this statement, but we will make an attempt to give a statistical–mechanical insight into it. Consider the microstate $\boldsymbol{x} = (x_1, \ldots, x_N)$ and let us think of the probability function

$$P(x_1, \ldots, x_N) = \prod_{i=1}^{N} P(x_i) = \exp\left\{-(\ln 2)\sum_{i=1}^{N} \log_2[1/P(x_i)]\right\} \tag{8.2.5}$$

as an instance of the canonical ensemble at inverse temperature $\beta = \ln 2$, where Hamiltonian is additive, namely, $\mathcal{E}(x_1, \ldots, x_N) = \sum_{i=1}^{N} \epsilon(x_i)$, with $\epsilon(x_i) = -\log_2 P(x_i)$. Obviously, $Z(\beta) = Z(\ln 2) = 1$, so the free energy is exactly zero here. Now, by the weak law of large numbers, for most realizations of the microstate $\boldsymbol{x}$,

$$\frac{1}{N}\sum_{i=1}^{N} \epsilon(x_i) \approx \langle \epsilon(x_i)\rangle = \langle -\log_2 P(x_i)\rangle = H, \tag{8.2.6}$$

so the average 'internal energy' is NH. It is safe to consider instead, the corresponding *microcanonical ensemble*, which is equivalent as far as macroscopic averages go. In the microcanonical ensemble, we would then have:

$$\frac{1}{N}\sum_{i=1}^{N} \epsilon(x_i) = H \tag{8.2.7}$$

for *every* realization of $\boldsymbol{x}$. How many bits would it take us to represent $\boldsymbol{x}$ in this microcanonical ensemble? Since all $\boldsymbol{x}$'s are equiprobable in the microcanonical ensemble, we assign to all $\boldsymbol{x}$'s binary code-words of the same length, call it L. In order to have a one–to–one mapping between the set of accessible $\boldsymbol{x}$'s and binary strings of representation, 2^L, which is the number of binary strings of length L, should be no less than the number of microstates $\{\boldsymbol{x}\}$ of the microcanonical ensemble. Thus,

$$L \ge \log_2\left|\left\{\boldsymbol{x} : \sum_{i=1}^{N} \epsilon(x_i) = NH\right\}\right| \stackrel{\Delta}{=} \log_2 \Omega(NH), \tag{8.2.8}$$

but the r.h.s. is exactly related (up to a constant factor) to Boltzmann's entropy associated with 'internal energy' at the level of NH. Now, observe that the free energy of the original, canonical ensemble, which is zero, is related to the entropy $\ln \Omega(NH)$ via the Legendre relation $\ln Z \approx \ln \Omega - \beta E$, which is

$$0 = \ln Z(\ln 2) \approx \ln \Omega(NH) - NH \ln 2 \tag{8.2.9}$$

and so,

$$\ln \Omega(NH) \approx NH \ln 2 \tag{8.2.10}$$

or

$$\log_2 \Omega(NH) \approx NH, \tag{8.2.11}$$

and therefore, by (8.2.8):

$$L \geq \log_2 \Omega(NH) \approx NH, \tag{8.2.12}$$

which means that the length of the binary representation essentially cannot be less than NH, namely, a compression rate of H bits per component of $\boldsymbol{x}$. So, we have seen that the entropy has a very concrete information–theoretic meaning, and in fact, it is not the only one, but we will not delve into this any further here.

8.3 Statistical Physics of Optimum Message Distributions

We next study another, very simple paradigm of a communication system, studied by Reiss [2] and Reiss and Huang [3]. The analogy and the parallelism to the basic concepts of statistical mechanics, that were introduced earlier, will be quite evident from the choice of the notation, which is deliberately chosen to correspond to that of analogous physical quantities.

Consider a continuous–time communication system that includes a noiseless channel, with capacity

$$C = \lim_{E \to \infty} \frac{\log_2 M(E)}{E}, \tag{8.3.1}$$

where $M(E)$ is the number of distinct messages (and $\log_2$ of this is the number of bits) that can be transmitted over a time interval of E seconds. Over a duration of E seconds, L information symbols are conveyed, so that the average transmission time per symbol is $\sigma = E/L$ seconds per symbol. In the absence of any constraints on the structure of the encoded messages, $M(E) = r^L = r^{E/\sigma}$, where r is the channel input–output alphabet size. Thus, $C = (\log r)/\sigma$ bits per second.

Consider now the thermodynamic limit of $L \to \infty$. Suppose that the L symbols of duration E form N *words*, where by 'word', we mean a certain variable–length string of channel symbols. The average transmission time per word is then $\epsilon = E/N$. Suppose further, that the code defines a certain set of word transmission times: word number i takes ϵ_i seconds to transmit. What is the optimum allocation of word probabilities $\{P_i\}$ that would support full utilization of the channel capacity? Equivalently, given the probabilities $\{P_i\}$, what are the optimum transmission times $\{\epsilon_i\}$? For simplicity, we will assume that $\{\epsilon_i\}$ are all distinct. Suppose that each word appears N_i times in the entire message. Denoting $\boldsymbol{N} = (N_1, N_2, \ldots)$, $P_i = N_i/N$, and $\boldsymbol{P} = (P_1, P_2, \ldots)$, the total number of messages pertaining to a given $\boldsymbol{N}$ is

$$\Omega(\boldsymbol{N}) = \frac{N!}{\prod_i N_i!} \doteq \exp\{N \cdot H(\boldsymbol{P})\} \tag{8.3.2}$$

where $H(\boldsymbol{P})$ is the Shannon entropy pertaining to the probability distribution $\boldsymbol{P}$. Now,

$$M(E) = \sum_{\boldsymbol{N}:\ \sum_i N_i \epsilon_i = E} \Omega(\boldsymbol{N}). \tag{8.3.3}$$

This sum is dominated by the maximum term, namely, the maximum–entropy assignment of relative frequencies

$$P_i = \frac{e^{-\beta\epsilon_i}}{Z(\beta)} \tag{8.3.4}$$

where $\beta > 0$ is a Lagrange multiplier chosen such that $\sum_i P_i \epsilon_i = \epsilon$, which gives

$$\epsilon_i = -\frac{\ln[P_i Z(\beta)]}{\beta}. \tag{8.3.5}$$

For $\beta = 1$, this is in agreement with our earlier observation that the optimum message length assignment in variable–length lossless data compression is according to the negative logarithm of the probability.

Suppose now that $\{\epsilon_i\}$ are kept fixed and consider a small perturbation in P_i, denoted $\mathrm{d}P_i$. Then

$$\begin{aligned}
\mathrm{d}\epsilon &= \sum_i \epsilon_i \mathrm{d}P_i \\
&= -\frac{1}{\beta}\sum_i (\mathrm{d}P_i)\ln[P_i Z(\beta)] \\
&= -\frac{1}{\beta}\sum_i (\mathrm{d}P_i)\ln P_i - \frac{1}{\beta}\sum_i (\mathrm{d}P_i)\ln Z(\beta) \\
&= -\frac{1}{\beta}\sum_i (\mathrm{d}P_i)\ln P_i \\
&= \frac{1}{k\beta}\mathrm{d}\left(-k\sum_i P_i \ln P_i\right) \\
&\stackrel{\Delta}{=} T\mathrm{d}s,
\end{aligned} \tag{8.3.6}$$

where we have defined $T = 1/(k\beta)$ and $s = -k \sum_i P_i \ln P_i$. The free energy per particle is given by

$$f = \epsilon - Ts = -kT \ln Z, \tag{8.3.7}$$

which is related to the redundancy of the code. In [2], there is also an extension of this setting to the case where N is not fixed, with correspondence to the grand—canonical ensemble.

8.4 Suggestions for Supplementary Reading

In addition to the references mentioned in this chapter, the reader is referred to [1, 4], as well as many references therein, for a much deeper exposition of the relation to information theory. We should also mention that Sect. 8.4 above is similar to Sect. 3.1 of [1].

References

1. N. Merhav, Statistical physics and information theory. Found. Trends Commun. Inf. Theory **6**(1–2), 1–212 (2009)
2. H. Reiss, Thermodynamic-like transformations in information theory. J. Stat. Phys. **1**(1), 107–131 (1969)
3. H. Reiss, C. Huang, Statistical thermodynamic formalism in the solution of information theory problems. J. Stat. Phys. **3**(2), 191–211 (1971)
4. M. Mézard, A. Montanari, *Information, Physics, and Computation* (Oxford University Press, Oxford, 2009)