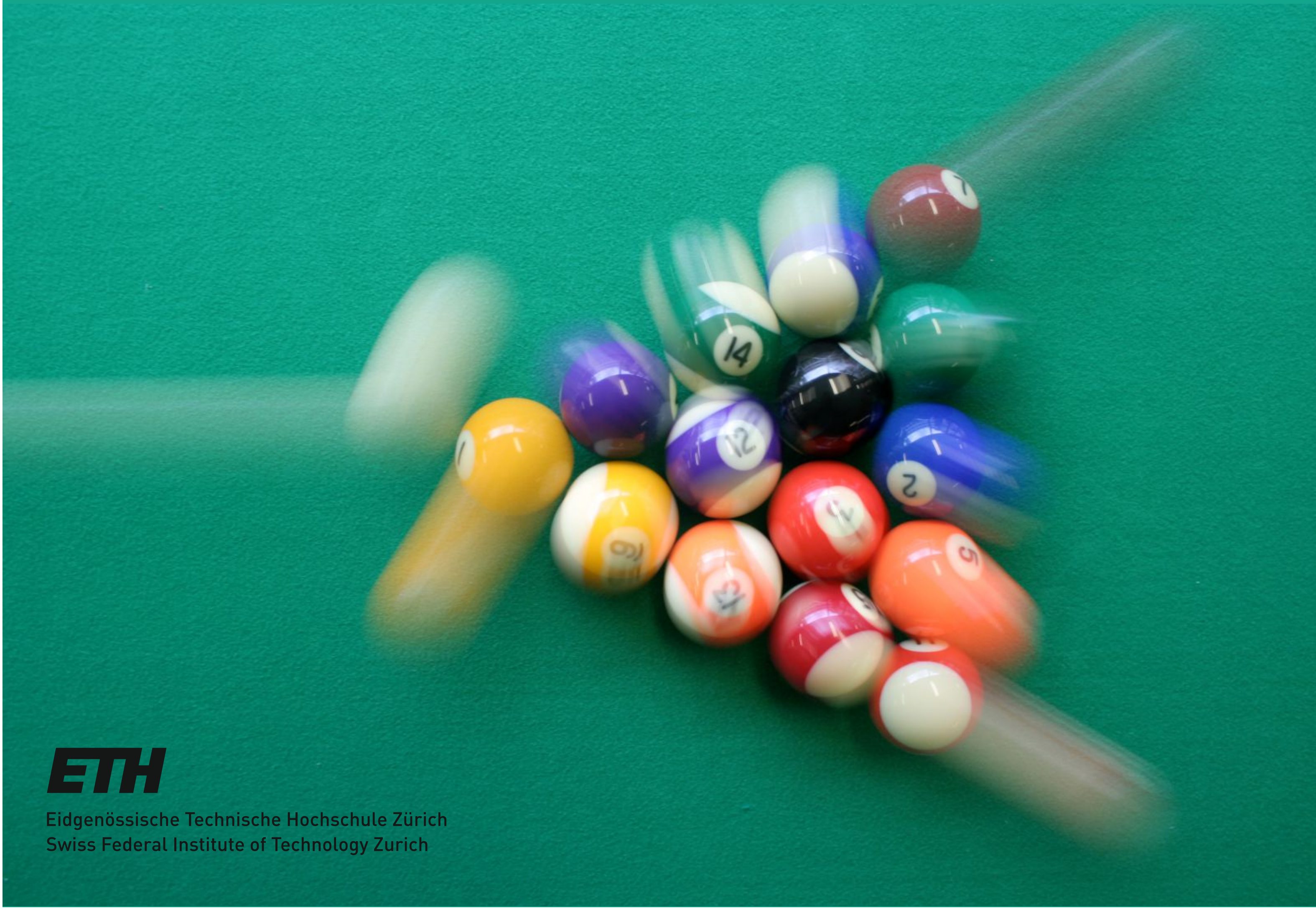
Lecture Notes

Advanced Physical Chemistry

Statistical Thermodynamics

Gunnar Jeschke



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Copyright © 2015 Gunnar Jeschke

Title image: „Billard“ von No-w-ay in collaboration with H. Caps - Eigenes Werk.
Lizenziert unter GFDL über Wikimedia Commons -
`https://commons.wikimedia.org/wiki/File:Billard.JPG`
Chapter 2 Word cloud: `http://de.123rf.com/profile_radiantskies`
Chapter 3 Dice image: `http://de.123rf.com/profile_whitetag`
Chapter 4 Matryoshka image: `http://de.123rf.com/profile_mikewaters`
Chapter 5 Word cloud: `http://de.123rf.com/profile_radiantskies`
Chapter 7 Dumbbell image: `http://de.123rf.com/profile_filipobr`
Chapter 8 Spaghetti image: `http://de.123rf.com/profile_winterbee`

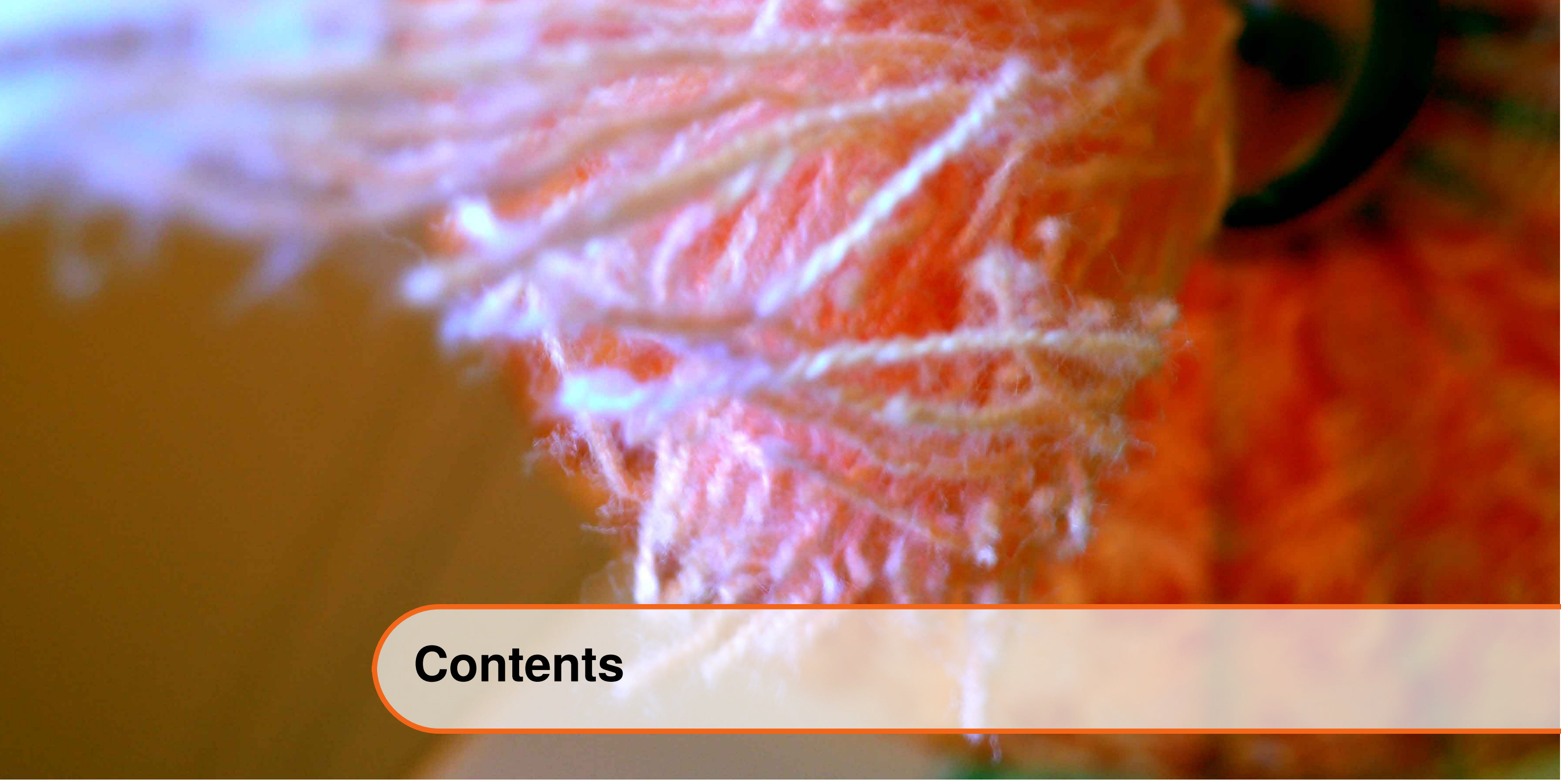
PUBLISHED BY GUNNAR JESCHKE

`http://www.epr.ethz.ch`

Licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License (the “License”). You may not use this file except in compliance with the License. You may obtain a copy of the License at `http://creativecommons.org/licenses/by-nc/3.0`.

Design and layout of the lecture notes are based on the Legrand Orange Book available at `http://latextemplates.com/template/the-legrand-orange-book`.

First printing, September 2015



Contents

1 Introduction . . . 7

1.1 General Remarks 7

1.2 Suggested Reading 8

1.3 Acknowledgment 9

2 Basics of Statistical Mechanics . . . 11

2.1 Basic Assumptions of Statistical Thermodynamics 11

2.1.1 Thermodynamics Based on Statistical Mechanics . . . 11
2.1.2 The Markovian Postulate . . . 12

2.2 Phase space 13

2.2.1 Hamiltonian Equations of Motion . . . 13
2.2.2 The Liouville Equation . . . 14
2.2.3 Quantum Systems . . . 14

2.3 Statistical Mechanics Based on Postulates 15

2.3.1 The Penrose Postulates . . . 15
2.3.2 Implications of the Penrose Postulates . . . 15

3 Probability Theory . . . 17

3.1 Discrete Probability Theory 17

3.1.1 Discrete Random Variables . . . 17
3.1.2 Multiple Discrete Random Variables . . . 18
3.1.3 Functions of Discrete Random Variables . . . 20
3.1.4 Discrete Probability Distributions . . . 20
3.1.5 Probability Distribution of a Sum of Random Numbers . . . 21
3.1.6 Binomial Distribution . . . 22
3.1.7 Stirling's Formula . . . 24

3.2 Continuous Probability Theory 24
3.2.1 Probability Density 24
3.2.2 Selective Integration of Probability Densities 27
3.2.3 Sum of Two Continuous Random Numbers 28

4 Classical Ensembles 31

4.1 Statistical Ensembles 31
4.1.1 Concept of an Ensemble 31
4.1.2 Ergodicity 31

4.2 Microcanonical Ensemble 32

4.3 Canonical Ensemble 33
4.3.1 Boltzmann Distribution 33
4.3.2 Equipartition Theorem 37
4.3.3 Internal Energy and Heat Capacity of the Canonical Ensemble 38

4.4 Grand Canonical Ensemble 39

5 Entropy 41

5.1 Swendsen’s Postulates of Thermodynamics 41
5.1.1 Cautionary Remarks on Entropy 41
5.1.2 Swendsen’s Postulates 41
5.1.3 Entropy in Phenomenological Thermodynamics 42
5.1.4 Boltzmann’s Entropy Definition 43

5.2 The Relation of State Functions to the Partition Function 44
5.2.1 Entropy and the Partition Function 44
5.2.2 Helmholtz Free Energy 46
5.2.3 Gibbs Free Energy, Enthalpy, and Pressure 46

5.3 Irreversibility 47
5.3.1 Historical Discussion 47
5.3.2 Irreversibility as an Approximation 48

5.4 Entropy and Information 48
5.4.1 Gibbs Entropy 48
5.4.2 Von Neumann Entropy 49
5.4.3 Shannon Entropy 50

6 Quantum Ensembles 51

6.1 Quantum Canonical Ensemble 51
6.1.1 Density Matrix 51
6.1.2 Quantum Partition Function 53

6.2 Quantum and Classical Statistics 54
6.2.1 Types of Permutation Symmetry 54
6.2.2 Bose-Einstein Statistics 55
6.2.3 Fermi-Dirac Statistics 56
6.2.4 Maxwell-Boltzmann Statistics 57

6.3 Simple Quantum Systems 59
6.3.1 Spin $S = 1/2$ 59
6.3.2 Harmonic Oscillator 60
6.3.3 Einstein and Debye Models of a Crystal 62

7 Partition Functions of Gases 65

7.1 Separation of contributions 65
7.1.1 Collective Degrees of Freedom 65
7.1.2 Factorization of Energy Modes 66

7.2 Translational Partition Function 67
7.2.1 Density of States of an Ideal Gas 67
7.2.2 Partition Function and Accessible States 68

7.3 Nuclear Spin Partition Function 69
7.3.1 High-Temperature Limit 69
7.3.2 Symmetry Requirements 70

7.4 Rotational Partition Function 71
7.4.1 Rigid Rotor Assumption and Rotamers 71
7.4.2 Accessible States and Symmetry 72

7.5 Vibrational Partition Function 74
7.5.1 The Harmonic Oscillator Extended 74
7.5.2 Vibrational Contributions to U, C_V, and S 75

7.6 Electronic Partition Function 75

7.7 Equilibrium Constant for Gas Reactions 77

8 Macromolecules 81

8.1 Thermodynamics of Mixing 81
8.1.1 Entropy of Binary Mixing 81
8.1.2 Energy of Binary Mixing 83

8.2 Entropic Elasticity 85
8.2.1 Ideal Chain Model 85
8.2.2 Random Walk 87
8.2.3 Conformational Entropy and Free Energy 91

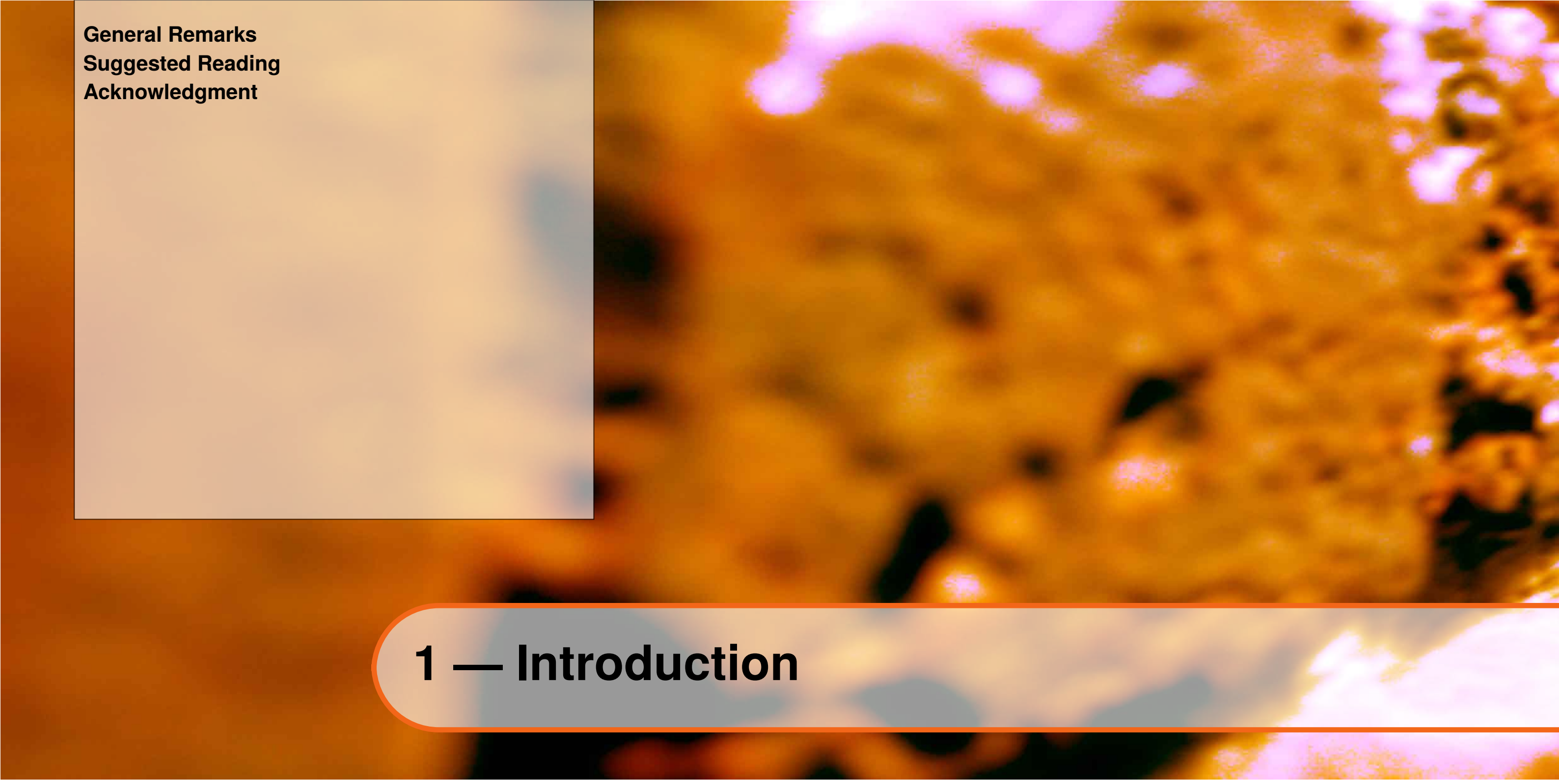
Bibliography 93

Books 93

Articles 94

Web Pages 94

Index 95



General Remarks
Suggested Reading
Acknowledgment

1 — Introduction

„Today", Mr. K. complained, „Scores of people claim in public that they can type sizeable books entirely on their own, and this is universally accepted. The Chinese philosopher Chuan-Tzu, when already at his prime age, brushed a tome of one hundred thousand words with nine out of ten being citations. Such books cannot be written anymore, since wit is missing. Hence, only by the tools of a single man thoughts are produced, whereas he feels lazy if their number is low. Indeed, no thought can be reused and no expression be cited. How little they all need for their doings! A pen and some paper is all they have to show! And without any help, with only the puny material that a single man can carry on his arms, they set up their huts! They don't know of larger buildings than a loner can raise."

Freely translated from: Bertolt Brecht, Geschichten vom Herrn Keuner

1.1 General Remarks

The field of Statistical Thermodynamics is probably the branch of physical chemistry whose coverage in textbooks is most diverse. A canonical way of teaching this subject still appears to be missing, which is partially due to the fact that practitioners have not completely agreed on interpretation of the probabilistic character, on the postulates, and on the way how the mathematical apparatus is derived from a set of postulates. While this may appear troublesome, actually there exists rarely any problem in *applying* statistical thermodynamics[1]. Accordingly, textbooks usually ignore the more problematic aspects and try to give reasons why the interpretation and formulation used by the respective author should be the preferred one. This being an advanced lecture course, we shall not do so, but we shall still present an apparatus that is ready-made for application.

The basic idea of statistical thermodynamics is simple: On the one hand we have Newtonian and quantum mechanics and we know that molecules should adhere to it, and on the other hand we know that systems consisting of many molecules can be adequately described by phenomenological (or classical) thermodynamics. Now let's try to derive the latter theory from the former one. Some care will have to be taken for systems that are subject to quantum statistics, but we might expect that straightforward application of probability theory will provide the required connection. Chapter 2 will discuss this basic idea in some more detail and will present a set of postulates due to Oliver Penrose [Pen70]. The discussion of these postulates clarifies what the remaining mathematical problem is and how we avoid it in applications.

---

[1] The problematic cases are mostly those where the number of particles is too small for the approximations made in the statistical approach.

In this course we do not assume that students are already familiar with probability theory, rather we will introduce its most important concepts in Chapter 3. We do assume that the concepts of phenomenological thermodynamics are known, although we shall shortly explain them on first use in these lecture notes. The most important new concept in this course is the one of an ensemble description, which will be introduced in Chapter 4 first only for classical particles. This will set the stage for discussing the concepts of irreversibility and entropy in Chapter 5. We will complete the foundations part with a discussion of quantum ensembles in Chapter 6. This Chapter will also make the transition to applications, by treating first the harmonic oscillator and second the Einstein model of a crystal with the apparatus that we command at that point.

We shall then illustrate the relation to phenomenological thermodynamics by discussing the partition functions of gases and by computing thermodynamical state functions from these partition functions in Chapter 7. The final Chapter 8 will shortly discuss the consequences of statistical thermodynamics for macromolecular systems and introduce the concepts of lattice models, random walks, and entropic elasticity.

The time available for this course does not permit to treat all aspects of statistical thermodynamics and statistical mechanics that are important in physical chemistry, chemical physics, polymer physics, and biophysics, let alone in solid-state physics. The most important omissions are probably kinetic aspects of chemical reactions, which are treated in detail in a lecture course on Advanced Kinetics, and the topic of phase transitions, including the famous Ising chain model. We believe that the foundations laid in the present course will allow students to understand these topics from reading in the textbooks listed in the following Section.

1.2 Suggested Reading

General textbooks on physical chemistry, such as [AP13; ER06] provide some overview of the most important concepts of statistical thermodynamics as well as some of the key formulas, but they are not quite on the level of this advanced course. If you already own these books, it might still be useful to read what they write on the topic. If you have the choice, Engel/Reid [ER06] is better on this topic than Atkins [AP13]. The best introduction in a general physical chemistry textbook can be found in the German book by Wedler and Freund [WF12].

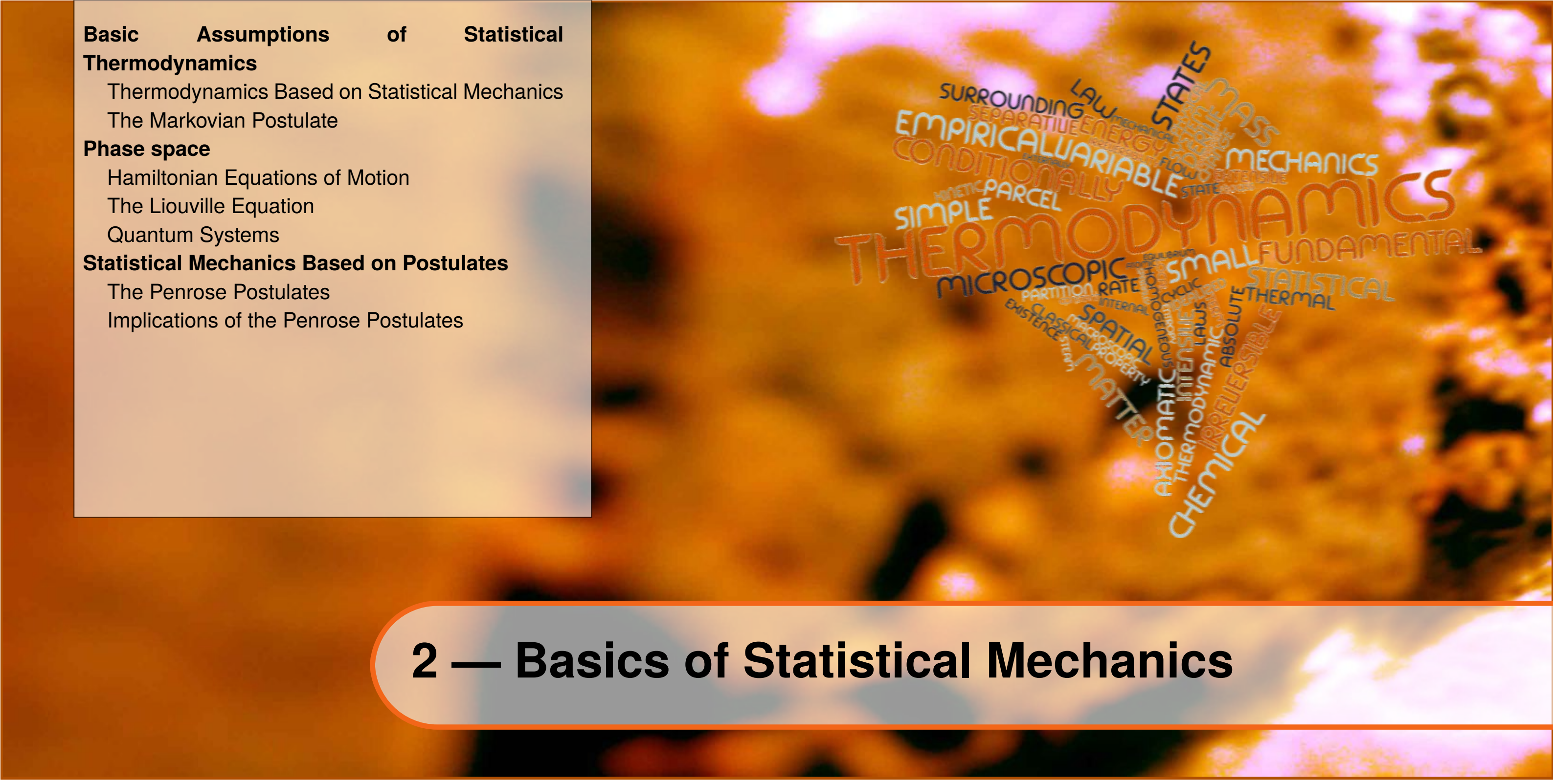
A rather good and modern introduction at an advanced level has been published in English language by Swendsen [Swe12]. Swendsen introduces statistical thermodynamics together with phenomenological thermodynamics and covers more examples than we can treat in this course. He does not introduce some concepts that are widely used in the field, because he dislikes them. In this course we do introduce these concepts and discuss the problems associated with them.

A modern German-language introduction is the one by Schwabl [Sch06], which caters more to the physicist than to the physical chemist. Schwabl is stronger on phase transitions and dynamic phenomena, but probably harder to read than Swendsen, even if German is your native language. For Chapter 7, but only for this Chapter, Maczek's book from the Oxford Chemistry Primers series [Mac98] can be quite useful. Several topics that are not or only superficially covered in my lecture notes are treated in the notes by Cohen from Ben Gurion University [Coh11], which are strongly focused on quantum applications. Finally, I want to mention Penrose's book [Pen70], which is certainly not an introductory textbook and may be most appealing to the strongly mathematically or philosophically inclined. If you look for guidance on *applying* statistical thermodynamics to real systems this book is certainly deficient, but from an epistemological point of view it is probably the best one.

For many of the central concepts I have looked up (English) Wikipedia articles and have found that these articles are, on average, of rather good quality. They do differ quite strongly from each other in style and notation. When using only Wikipedia or other internet resources it is difficult to fit the pieces of information together. If, on the other hand, you already do have a basic level of understanding, but some difficulties with a particular concept, such sources may provide just the missing piece of information. The NIST guide for computing thermodynamical state functions from the results of *ab initio* computations is a particularly good example for a useful internet resource [Iri98].

1.3 Acknowledgment

I am grateful to M. Schäfer, U. Hollenstein, and F. Merkt for making their lecture notes for this course available and to Takuya Segawa for thorough proofreading of the first manuscript of these notes. All remaining errors are my own sole responsibility.



Basic Assumptions of Statistical Thermodynamics
Thermodynamics Based on Statistical Mechanics
The Markovian Postulate
Phase space
Hamiltonian Equations of Motion
The Liouville Equation
Quantum Systems
Statistical Mechanics Based on Postulates
The Penrose Postulates
Implications of the Penrose Postulates

2 — Basics of Statistical Mechanics

2.1 Basic Assumptions of Statistical Thermodynamics

2.1.1 Thermodynamics Based on Statistical Mechanics

Phenomenological thermodynamics describes relations between observable quantities that characterize macroscopic material objects. We know that these objects consist of a large number of small particles, molecules or atoms, and, for all we know, these small particles adhere to the laws of quantum mechanics and often in good approximation to the laws of Newtonian mechanics. Statistical mechanics is the theory that explains macroscopic properties, not only thermodynamic state functions, by applying probability theory to the mechanic equations of motion for a large ensemble of systems of particles. In this lecture course we are concerned with the part of statistical mechanics that relates to phenomenological thermodynamics.

In spite of its name, phenomenological (equilibrium) thermodynamics is essentially a static theory that provides an *observational*, *macroscopic* description of matter. The underlying mechanical description is *dynamical* and *microscopic*, but it is observational only for systems consisting of a small number of particles. To see this, we consider a system of N identical classical point particles that adhere to Newton's equations of motion.

Concept 2.1.1 — Newtonian equations of motion. With particle mass m, Cartesian coordinates q_i $(i = 1, 2, \ldots, 3N)$ and velocity coordinates $\dot{q}_i$, a system of N identical classical point particles evolves by

$$m\frac{\mathrm{d}^2 q_i}{\mathrm{d}t^2} = -\frac{\partial}{\partial q_i} V\left(q_1, \ldots, q_{3N}\right) \ , \tag{2.1}$$

where $V(q_1, \ldots, q_{3N})$ is the potential energy function.

Notation 2.1. The *dynamical state* or *microstate* of the system at any instant is defined by the $6N$ Cartesian and velocity coordinates, which span the *dynamical space* of the system. The curve of the system in dynamical space is called a *trajectory*.

The concept extends easily to atoms with different masses m_i. If we could, at any instant, precisely measure all $6N$ dynamical coordinates, i.e., spatial coordinates and velocities, we could precisely predict the future trajectory. The system as described by the Newtonian equations of motions behaves deterministically.

For any system that humans can see and handle directly, i.e., without complicated technical devices, the number N of particles is too large (at least of the order of 10^{18}) for such complete measurements to be possible. Furthermore, for such large systems even tiny measurement errors would make the trajectory prediction useless after a rather short time. In fact, atoms are quantum objects and the measurements are subject to the Heisenberg uncertainty principle, and even the small uncertainty introduced by that would make a deterministic description futile.

We can only hope for a theory that describes what we can observe. The number of *observational states* or *macrostates* that can be distinguished by the observer is much smaller than the number of dynamical states. Two classical systems in the same dynamical state are necessarily also in the same observational state, but the converse is not generally true. Furthermore, the observational state also evolves with time, but we have no equations of motion for this state (but see Section 2.2.2). In fact we cannot have deterministic equations of motion for the observational state of an individual system, precisely because the same observational state may correspond to different dynamical states that will follow different trajectories.

Still we can make predictions, only these predictions are necessarily statistical in nature. If we consider a large *ensemble* of identical systems in the same observational state we can even make fairly precise predictions about the outcome. Penrose [Pen70] gives the example of a women at a time when ultrasound diagnosis can detect pregnancy, but not sex of the foetus. The observational state is pregnancy, the two possible dynamical states are on path to a boy or girl. We have no idea what will happen in the individual case, but if the same diagnosis is performed on a million of women, we know that about 51-52% will give birth to a boy.

How then can we derive stable predictions for an ensemble of systems of molecules? We need to consider probabilities of the outcome and these probabilities will become exact numbers in the limit where the number N of particles (or molecules) tends to infinity. The theory required for computing such probabilities will be treated in Chapter 3.

R Our current usage of the term *ensemble* is loose. We will devote the whole Chapter 4 to clarifying what types of ensembles we use in computations and why.

2.1.2 The Markovian Postulate

There are different ways for defining and interpreting probabilities. For abstract discussions and mathematical derivations the most convenient definition is the one of *physical* or *frequentist* probability.

Definition 2.1.1 — Physical probability. Given a reproducible trial $\mathcal{T}$ of which A is one of the possible outcomes, the physical probability P of the outcome A is defined as

$$P(A|\mathcal{T}) = \lim_{\mathcal{N}\to\infty} \frac{n(A,\mathcal{N},\mathcal{T})}{\mathcal{N}} \tag{2.2}$$

where $n(A,\mathcal{N},\mathcal{T})$ is the number of times the outcome A is observed in the first $\mathcal{N}$ trials.

A trial $\mathcal{T}$ conforming to this definition is statistically regular, i.e., the limit exists and is the same for all infinite series of the same trial. If the physical probability is assumed to be a stable property of the system under study, it can be measured with some experimental error. This experimental error has two contributions: (i) the actual error of the measurement of the quantity A and (ii) the deviation of the experimental frequency of observing A from the limit defined in Eq. (2.2). Contribution (ii) arises from the experimental number of trials $\mathcal{N}$ not being infinite.

We need some criterion that tells us whether $\mathcal{T}$ is statistically regular. For this we split the trial into a preparation period, an evolution period, and the observation itself. The evolution period is a waiting time during which the system is under controlled conditions. Together with the preparation period it needs to fulfill the *Markovian postulate*.

Concept 2.1.2 — Markovian postulate. A trial $\mathcal{T}$ that invariably ends up in the observational state $\mathcal{O}$ of the system after the preparation stage is called statistically regular. The start of the evolution period is assigned a time $t = 0$.

Note that the system can be in different observational states at the time of *observation*; otherwise the postulate would correspond to a trivial experiment. The Markovian postulate is related to the concept of a Markovian chain of events. In such a chain the outcome of the next event depends only on the current state of the system, but not on states that were encountered earlier in the chain. Processes that lead to a Markovian chain of events can thus be considered as memoryless.

2.2 Phase space

2.2.1 Hamiltonian Equations of Motion

The Newtonian equations of motion (2.1) are very convenient for atomistic molecular dynamics (MD) computations. Statistical analysis of trajectories encountered during such MD simulations can be analyzed in terms of thermodynamic quantities, such as free energy. However, for analyzing evolution of the system in terms of spectroscopic properties, the Newtonian description is very inconvenient. Since spectroscopic measurements can provide the most stringent tests of theory, we shall use the Hamiltonian formulation of mechanics in the following. This formulation is particularly convenient for molecules that also have rotational degrees of freedom. For that, we replace the velocity coordinates by momentum coordinates $p_j = m_j \dot{q}_j$, where index j runs over all atoms and for each atom over the three Cartesian coordinates. Furthermore, we assume M identical molecules, with each of them having f degrees of freedom, so that the total number of degrees of freedom is $F = fM$. Such as system can be described by $2F$ differential equations

Concept 2.2.1 — Hamiltonian equations of motion. With the single-molecule Hamiltonian $\mathcal{H}(\mathbf{p}_i, \mathbf{q}_i)$ the equations of motion for M non-interacting identical molecules with f degrees of freedom for each molecule read

$$\frac{\mathrm{d}\mathbf{q}_i}{\mathrm{d}t} = \frac{\partial \mathcal{H}\left(\mathbf{p}_i, \mathbf{q}_i\right)}{\partial \mathbf{p_i}} \tag{2.3}$$

$$\frac{\mathrm{d}\mathbf{p}_i}{\mathrm{d}t} = -\frac{\partial \mathcal{H}\left(\mathbf{p}_i, \mathbf{q}_i\right)}{\partial \mathbf{q_i}} , \tag{2.4}$$

where $i = 1 \ldots M$. Each of the dynamical variables $\mathbf{q}_i$ and $\mathbf{p}_i$ is a vector of length f. The $2fM$ dynamical variables span the *phase space*.

Definition 2.2.1 — Phase space and state space. *Phase space* is the space where microstates of a system reside. Sometimes the term is used only for problems that can be described in spatial and momentum coordinates, sometimes for all problems where some type of a Hamiltonian equation of motion applies. Sometimes the term *state space* is used for the space of microstates in problems that cannot be described by (only) spatial and momentum coordinates.

If the molecule is just a single atom, we have only $f = 3$ translational degrees of freedom and the Hamiltonian is given by

$$\mathcal{H}\left(\mathbf{p}_i, \mathbf{q}_i\right) = \frac{1}{2m}\left(p_{x,i}^2 + p_{y,i}^2 + p_{z,i}^2\right) \ , \tag{2.5}$$

describing translation. For molecules with n atoms, three of the $f = 3n$ degrees of freedom are translational, two or three are rotational for linear and non-linear molecules, respectively, and the remaining $3n-5$ or $3n-6$ degrees of freedom are vibrational.

2.2.2 The Liouville Equation

Our observations do not allow us to specify phase space trajectories, i.e. the trajectory of microstates for a single system. Instead, we consider an ensemble of identical systems that all represent the same (observational) macrostate $\mathcal{O}$ but may be in different microstates. At a given time we can characterize such an ensemble by a probability density $\rho(\mathbf{p}, \mathbf{q}, t)$ in phase space, where $\mathbf{p}$ and $\mathbf{q}$ are the vectors of *all* momentum and spatial coordinates in the system, respectively. We are interested in an equation of motion for this probability density ρ, which corresponds to the full knowledge that we have on the system. This equation can be derived from an integral representation of ρ and the Hamiltonian equations of motion [Sch06].

Concept 2.2.2 — Liouville Equation. The probability density $\rho(\mathbf{p}, \mathbf{q}, t)$ in phase space evolves in time according to

$$\frac{\partial \rho}{\partial t} = \sum_i \left(\frac{\partial \rho}{\partial p_i}\frac{\partial \mathcal{H}}{\partial q_i} - \frac{\partial \rho}{\partial q_i}\frac{\partial \mathcal{H}}{\partial p_i} \right) \ . \tag{2.6}$$

With the *Poisson brackets*

$$\{u, v\} = \sum_i \left[\frac{\partial u}{\partial p_i}\frac{\partial v}{\partial q_i} - \frac{\partial u}{\partial q_i}\frac{\partial v}{\partial p_i} \right] \ . \tag{2.7}$$

this *Liouville equation* can be expressed as

$$\frac{\partial \rho}{\partial t} = -\left\{\mathcal{H}, \rho\right\} \ . \tag{2.8}$$

For the probability density along a phase space trajectory, i.e., along a trajectory that is taken by microstates, we find

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho\left(q(t), p(t), t\right) = 0 \ . \tag{2.9}$$

If we consider a uniformly distributed number $\mathrm{d}N$ of ensemble members in a volume element $\mathrm{d}\Gamma_0$ in phase space at time $t = 0$ and ask about the volume element $\mathrm{d}\Gamma$ in which these ensemble members are distributed at a later time, we find

$$\mathrm{d}\Gamma = \mathrm{d}\Gamma_0 \ . \tag{2.10}$$

This is the *Liouville theorem* of mechanics.

2.2.3 Quantum Systems

Hamiltonian mechanics can be applied to quantum systems, with the Hamiltonian equations of motion being replaced by the time-dependent Schrödinger equation. The probability density in phase space is replaced by the density operator $\widehat{\rho}$ and the Liouville equation by the Liouville-von-Neumann equation

$$\frac{\partial \widehat{\rho}}{\partial t} = -\frac{i}{\hbar}\left[\widehat{\mathcal{H}}, \widehat{\rho}\right] . \tag{2.11}$$

In quantum mechanics, observables are represented by operators $\widehat{A}$. The expectation value of an observable can be computed from the density operator that represents the distribution of the ensemble in phase space,

$$\langle \widehat{A} \rangle = \mathrm{Trace}\left(\widehat{\rho}\widehat{A}\right) . \tag{2.12}$$

We note that the Heisenberg uncertainty relation does not introduce an additional complication in statistical mechanics. Determinism had been lost before and the statistical character of the measurement on an individual system is unproblematic, as we seek only statistical predictions for a large ensemble. In the limit of an infinite ensemble, $N \to \infty$, there is no uncertainty and the expectation values of incompatible observables are well defined and can be measured simultaneously. Such an infinitely large system is not perturbed by the act of observing it. The only difference between the description of classical and quantum systems arises from their statistical behavior on permutation of the coordinates of two particles, see Section 6.2.

2.3 Statistical Mechanics Based on Postulates

2.3.1 The Penrose Postulates

Penrose [Pen70] has made the attempt to strictly specify what results can be expected from statistical mechanics if the theory is based on a small number of plausible postulates.

1. Macroscopic physical systems are composed of molecules that obey classical or quantum mechanical equations of motion (*dynamical description of matter*).
2. An observation on such a macroscopic system can be idealized as an instantaneous, simultaneous measurement of a set of dynamical variables, each of which takes the values 1 or 0 only (*observational description of matter*).
3. A measurement on the system has no influence whatsoever on the outcome of a later measurement on the same system (*compatibility*).
4. The Markovian postulate. (Concept 2.1.2)
5. Apart from the Bose and Fermi symmetry conditions for quantum systems, the whole phase space can, in principle, be accessed by the system (*accessibility*).

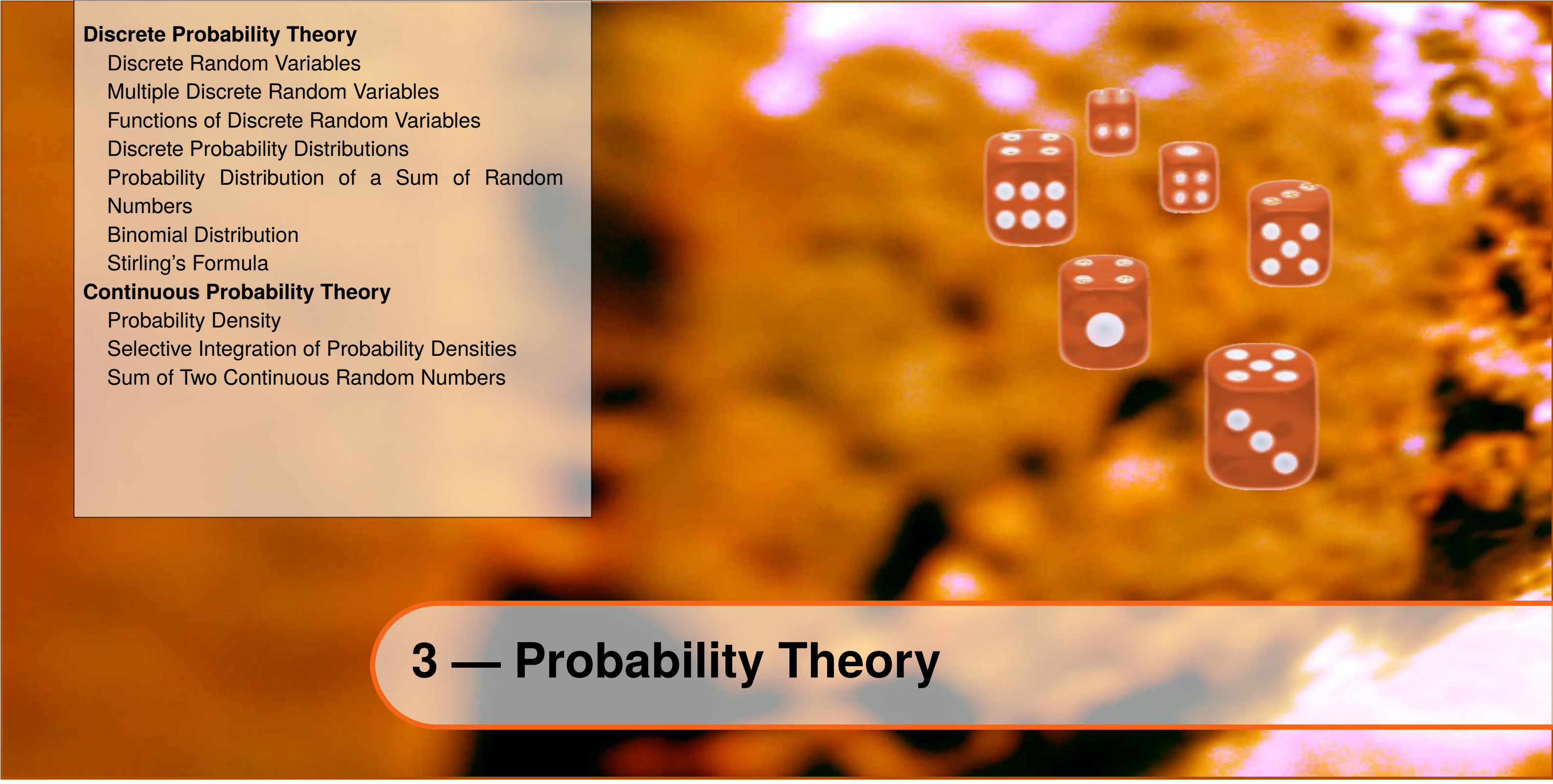
After the discussion above, only the second of these postulates may not immediately appear plausible. In the digital world of today it appears natural enough: Measurements have resolution limits and their results are finally represented in a computer by binary numbers, which can be taken to be the dynamical variables in this postulate.

2.3.2 Implications of the Penrose Postulates

Entropy is one of the central quantities of thermodynamics, as it tells in which direction a spontaneous process in an isolated system will proceed. For closed systems that can exchange heat and work with their environment, such predictions on spontaneous processes are based on free energy, of which the entropy contribution is usually an important part. To keep such considerations consistent, entropy must have two fundamental properties

1. If the system does not exchange energy with its environment, its entropy cannot decrease. (*non-decrease*).
2. The entropy of two systems considered together is the sum of their separate entropies. (*additivity*).

Based on the Penrose postulates it can be shown [Pen70] that the definition of Boltzmann entropy (Chapter 5) ensures both properties, but that statistical expressions for entropy ensure only the non-decrease property, not in general the additivity property. This appears to leave us in an inconvenient situation. However, it can also be shown that for large systems, in the sense that the number of macrostates is much smaller than the number of microstates, the term that quantifies non-additivity is negligibly small compared to the total entropy [Pen70]. The problem is thus rather a mathematical beauty spot than a serious difficulty in application of the theory.



Discrete Probability Theory
- Discrete Random Variables
- Multiple Discrete Random Variables
- Functions of Discrete Random Variables
- Discrete Probability Distributions
- Probability Distribution of a Sum of Random Numbers
- Binomial Distribution
- Stirling's Formula

Continuous Probability Theory
- Probability Density
- Selective Integration of Probability Densities
- Sum of Two Continuous Random Numbers

3 — Probability Theory

3.1 Discrete Probability Theory

3.1.1 Discrete Random Variables

Consider a trial $\mathcal{T}$ where the observation is a measurement of the z component $\hbar m_S$ of spin angular momentum of a spin $S = 5/2$. There are just six possible outcomes (events) that can be labeled with the magnetic spin quantum number m_S or indexed by integer numbers 1, 2, . . . 6. In general, the probabilities of the six possible events will differ from each other. They will depend on preparation and may depend on evolution time before the observation. To describe such situations, we define a set of elementary events

$$A = \{a_j\} \ , \tag{3.1}$$

where in our example index j runs from 1 to 6, whereas in general it runs from 1 to the number N_A of possible events. Each of the events is assigned a probability $0 \leq P(a_j) \leq 1$. Impossible events (for a given preparation) have probability zero and a certain event has probability 1. Since one and only one of the events must happen in each trial, the probabilities are normalized, $\sum_j^{N_A} P(a_j) = 1$. A simplified model of our example trial is the rolling of a die. If the die is fair, we have the special situation of a uniform probability distribution, i.e., $P(a_j) = 1/6$ for all j.

Concept 3.1.1 — Random variable. A set of random events with their associated probabilities is called a *random variable*. If the number of random events is countable, the random variable is called *discrete*. In a computer, numbers can be assigned to the events, which makes the random variable a *random number*. A series of trials can then be simulated by generating a series of $\mathcal{N}$ pseudo-random numbers that assign the events observed in the $\mathcal{N}$ trials. Such simulations are called *Monte Carlo* simulations. Pseudo-random numbers obtained from a computer function need to be adjusted so that they reproduce the given or assumed probabilities of the events.

Problem 3.1 Using the Matlab® function `rand`, which provides uniformly distributed random numbers in the open interval $(0, 1)$, write a program that simulates throwing a die with six faces. The outer function should have trial number $\mathcal{N}$ as an input and a vector of the numbers of encountered ones, twos, ... and sixes as an output. It should be based on an inner function that simulates a single throw of the die. Test the program by determining the difference from the expectation $P(a_j) = 1/6$ for ever larger numbers of trials.

3.1.2 Multiple Discrete Random Variables

For two sets of events A and B and their probabilities, we define a *joint probability* $P(a_j, b_k)$ that is the probability of observing both a_j and b_k in the same trial. An example is the throwing of two dice, one black and one red, and asking about the probability that the black die shows a 2 and the red die a 3. A slightly more complicated example is the measurement of the individual z components of spin angular momentum of two coupled spins $S_\mathrm{A} = 5/2$ and $S_\mathrm{B} = 5/2$. Like individual probabilities, joint probabilities fall in the closed interval $[0, 1]$. Joint probabilities are normalized,

$$\sum_a \sum_b P(a, b) = 1 \; . \tag{3.2}$$

Note that we have introduced a brief notation that suppresses indices j and k. This notation is often encountered because of its convenience in writing.

If we know the probabilities $P(a, b)$ for all $N_A \cdot N_B$ possible combinations of the two events, we can compute the probability of a single event, for instance a,

$$P_A(a) = \sum_b P(a, b) \; , \tag{3.3}$$

where $P_A(a)$ is the *marginal probability* of event a.

R The unfortunate term 'marginal' does not imply a small probability. Historically, these probabilities were calculated in the margins of probability tables [Swe12].

Another quantity of interest is the *conditional probability* $P(a|b)$ of an event a, provided that b has happened. For instance, if we call two cards from a full deck, the probability of the second card being a Queen is conditional on the first card having been a Queen. With the definition for the conditional probability we have

$$P(a, b) = P(a|b) P_B(b) \tag{3.4}$$
$$= P(b|a) P_A(a) \; . \tag{3.5}$$

Theorem 3.1.2 — Bayes' theorem. If the marginal probability of event b is not zero, the conditional probability of event a given b is

$$P(a|b) = \frac{P(b|a) P_A(a)}{P_B(b)} \; . \tag{3.6}$$

Bayes' theorem is the basis of Bayesian inference, where the probability of proposition a is sought given prior knowledge (short: the prior) b. Often Bayesian probability is interpreted subjectively, i.e., different persons, because they have different prior knowledge b, will come to different assessments for the probability of proposition a. This interpretation is incompatible with theoretical physics, where, quite successfully, an objective reality is assumed. Bayesian probability theory can also be applied with an objective interpretation in mind and is nowadays used, among else, in structural modeling of biomacromolecules to assess agreement of a model (the proposition) with experimental data (the prior).

In experimental physics, biophysics, and physical chemistry, Bayes' theorem can be used to assign experimentally informed probabilities to different models for reality. For example assume that a theoretical modeling approach, for instance an MD simulation, has provided a set of conformations $A = \{a_j\}$ of a protein molecule and associated probabilities $P_A(a_j)$. The probabilities are related, *via* the Boltzmann distribution, to the free energies of the conformations (this point is discussed later in the lecture course). We further assume that we have a measurement B with output b_k and we know the marginal probability $P_B(b)$ of encountering this output for a random set of conformations of the protein molecule. Then we need only a physical model that provides the conditional probabilities $P(b_k|a_j)$ of measuring b_k given the conformations a_j and can compute the probability $P(a_j|b_k)$ that the true conformation is a_j, given the result of our measurement, via Bayes' theorem. Eq. (3.6). This procedure can be generalized to multiple measurements. The required $P(b_k|a_j)$ depend on measurement errors. The approach allows for combining possibly conflicting modeling and experimental results to arrive at a 'best estimate' for the distribution of conformations.

The events associated with two random variables can occur completely independent of each other. This is the case for throwing two dice: the number shown on the black die does not depend on the number shown on the red die. Hence, the probability to observe a 2 on the black and a 3 on the red die is $(1/6) \cdot (1/6) = 1/36$. In general, joint probabilities of *independent* events factorize into the individual (or marginal) probabilities, which leads to huge simplifications in computations. In the example of two coupled spins $S_\mathrm{A} = 5/2$ and $S_\mathrm{B} = 5/2$ the two random variables $m_{S,\mathrm{A}}$ and $m_{S,\mathrm{B}}$ may or may not be independent. This is decided by the strength of the coupling, the preparation of trial $\mathcal{T}$, and the evolution time t before observation.

Concept 3.1.3 — Independent variables. If two random variables are *independent*, the joint probability of two associated events is the product of the two marginal probabilities,

$$P(a,b) = P_A(a)P_B(b) \; . \tag{3.7}$$

As a consequence, the conditional probability $P(a|b)$ equals the marginal probability of a (and *vice versa*),

$$P(a|b) = P_A(a) \; . \tag{3.8}$$

For a set of more than two random variables two degrees of independence can be established, a weak type of *pairwise independence* and a strong type of *mutual independence*. The set is mutually independent if the marginal probability distribution in any subset, i.e. the set of marginal probabilities for all event combinations in this subset, is given by the product of the corresponding marginal distributions for the individual events.[1] This corresponds to complete independence. Weaker pairwise independence implies that the marginal distributions for any pair of random variables are given by the product of the two corresponding distributions. Note that even weaker independence can exist within the set, but not throughout the set. Some, but not all pairs or subsets of random variables can exhibit independence.

Another important concept for multiple random variables is whether or not they are distinguishable. In the example above we used a black and a red die to specify our events. If both dice would be black, the event combinations (a_2, b_3) and (a_3, b_2) would be indistinguishable and the corresponding composite event of observing a 2 and a 3 would have a probability of $1/18$, i.e. the product of the probability $1/36$ of the basic composite event with its multiplicity 2. In general, if n random variables are indistinguishable, the multiplicity equals the number of permutations of the n variables, which is $n! = 1 \cdot 2 \cdots (n-1) \cdot n$.

---

[1] As the distributions are vectors and all combinations have to be considered, an outer product must be taken.

3.1.3 Functions of Discrete Random Variables

We consider an event g that depends on two other events a and b. For example, we ask for the probability that the sum of the numbers shown by the black and red die is g, where g can range from 2 to 12, given that we know the probabilities $P(a,b)$, which in our example all have the value 1/36. In general, the probability distribution of random variable G can be computed by

$$P_G(g) = \sum_a \sum_b \delta_{g,G(a,b)} P(a,b) \ , \tag{3.9}$$

where $G(a,b)$ is an arbitrary function of a and b and the Kronecker delta $\delta_{g,G(a,b)}$ assumes the value one if $g = G(a,b)$ and zero otherwise. In our example, $g = G(a,b) = a + b$ will assume the value of 5 for the event combinations $(1,4), (2,3), (3,2), (4,1)$ and no others. Hence, $P_G(5) = 4/36 = 1/9$. There is only a single combination for $g = 2$, hence $P_G(2) = 1/36$, and there are 6 combinations for $g = 7$, hence $P_G(7) = 1/6$. Although the probability distributions for the individual random numbers A and B are uniform, the one for G is not. It peaks at the value of $g = 7$ that has the most realizations. Such peaking of probability distributions that depend on multiple random variables occurs very frequently in statistical mechanics. The peaks tend to become the sharper the larger the number of random variables that contribute to the sum. If this number N tends to infinity, the distribution of the sum g is so sharp that the distribution width (to be specified below) is smaller than the error in the measurement of the mean value g/N (see Section 3.1.5). This effect is the very essence of statistical thermodynamics: Although quantities for a single molecule may be broadly distributed and unpredictable, the mean value for a large number of molecules, let's say 10^{18} of them, is very well defined and perfectly predictable.

In a numerical computer program, Eq. (3.9) for only two random variables can be implemented very easily by a loop over all possible values of g with inner loops over all possible values of a and b. Inside the innermost loop, $G(a,b)$ is computed and compared to loop index g to add or not add $P(a,b)$ to the bin corresponding to value g. Note however that such an approach does not carry to large numbers of random variables, as the number of nested loops increases with the number of random variables and computation time thus increases exponentially. Analytical computations are simplified by the fact that $\delta_{g,G(a,b)}$ usually deviates from zero only within certain ranges of the summation indexes j (for a) and k (for b). The trick is then to find the proper combinations of index ranges.

Problem 3.2 Compute the probability distribution for the sum g of the numbers shown by two dice in two ways. First, write a computer program using the approach sketched above. Second, compute the probability distribution analytically by making use of the uniform distribution for the individual events ($P(a,b) = 1/36$ for all a, b. For this, consider index ranges that lead to a given value of the sum g.[2]

3.1.4 Discrete Probability Distributions

In most cases random variables are compared by considering the *mean values* and widths of their probability distributions. As a measure of the width, the *standard deviation* σ of the values from the mean value is used, which is the square root of the *variance* σ^2. The concept can be generalized by considering functions $F(A)$ of the random variable. In the following expressions, $F(A) = A$ provides the mean value and standard deviation of the original random variable A.

Concept 3.1.4 — Mean value and standard deviation. For any function $F(A)$ of a random variable A, the mean value $\langle F\rangle$ is given by,

---

[2] The solution of the second task can be found in [Swe12].

$$\langle F\rangle = \sum_a F(a)P_A(a)\ . \tag{3.10}$$

The standard deviation, which characterizes the width of the distribution of the function values $F(a)$, is given by,

$$\sigma = \sqrt{\sum_a \left(F(a) - \langle F\rangle\right)^2 P_A(a)}\ . \tag{3.11}$$

The mean value is the first moment of the distribution, with the n^{th} moment being defined by

$$\langle F^n\rangle = \sum_a F^n(a)P_A(a)\ . \tag{3.12}$$

The n^{th} central moment is

$$\langle (F - \langle F\rangle)^n\rangle = \sum_a \left(F(a) - \langle F\rangle\right)^n P_A(a)\ . \tag{3.13}$$

For the variance, which is the second central moment, we have

$$\sigma^2 = \langle F^2\rangle - \langle F\rangle^2\ . \tag{3.14}$$

Assume that we know the mean values for functions $F(A)$ and $G(B)$ of two random variables as well as the mean value $\langle FG\rangle$ of their product, which we can compute if the joint probability function $P(a,b)$ is known. We can then compute a *correlation function*

$$R_{FG} = \langle FG\rangle - \langle F\rangle\langle G\rangle\ , \tag{3.15}$$

which takes the value of zero, if F and G are independent random numbers.

Problem 3.3 Compute the probability distribution for the normalized sum g/M of the numbers obtained on throwing M dice in a single trial. Start with $M = 1$ and proceed via $M = 10, 100, 1000$ to $M = 10000$. Find out how many Monte Carlo trials $\mathcal{N}$ you need to guess the converged distribution. What is the mean value $\langle g/M\rangle$? What is the standard deviation σ_g? How do they depend on $\mathcal{N}$?

3.1.5 Probability Distribution of a Sum of Random Numbers

If we associate the random numbers with N molecules, identical or otherwise, we will often need to compute the sum over all molecules. This generates a new random number

$$S = \sum_{j=1}^{N} F_j\ , \tag{3.16}$$

whose mean value is the sum of the individual mean values,

$$\langle S\rangle = \sum_{j=1}^{N}\langle F_j\rangle\ . \tag{3.17}$$

If motion of the individual molecules is uncorrelated, the individual random numbers F_j are independent. It can then be shown that the variances add [Swe12],

$$\sigma_S^2 = \sum_{j=1}^{N} \sigma_j^2 \tag{3.18}$$

For identical molecules, all random numbers have the same mean $\langle F \rangle$ and variance σ_F^2 and we find

$$\langle S \rangle = N \langle F \rangle \tag{3.19}$$

$$\sigma_S^2 = N \sigma_F^2 \tag{3.20}$$

$$\sigma_S = \sqrt{N} \sigma_F \ . \tag{3.21}$$

This result relates to the concept of *peaking* of probability distributions for a large number of molecules that was introduced above on the example of the probability distribution for sum of the numbers shown by two dice. The width of the distribution normalized to its mean value,

$$\frac{\sigma_S}{\langle S \rangle} = \frac{1}{\sqrt{N}} \frac{\sigma_F}{\langle F \rangle} \ , \tag{3.22}$$

scales with the inverse square root of N. For 10^{18} molecules, this *relative width* of the distribution is one billion times smaller than for a single molecule. Assume that for a certain physical quantity of a single molecule the standard deviation is as large as the mean value. No useful prediction can be made. For a macroscopic sample, the same quantity can be predicted with an accuracy better than the precision that can be expected in a measurement.

3.1.6 Binomial Distribution

We consider the measurement of the z component of spin angular momentum for an ensemble of N spins $S = 1/2$.[3] The random number associated with an individual spin can take only two values, $-\hbar/2$ or $+\hbar/2$. Additive and multiplicative constants can be taken care of separately and we can thus represent each spin by a random number A that assumes the value $a = 1$ (for $m_S = +1/2$) with probability P and, accordingly, the value $a = 0$ (for $m_S = -1/2$) with probability $1 - P$. This is a very general problem, which also relates to the second postulate of Penrose (see Section 2.3.1). A simplified version with $P = 1 - P = 0.5$ is given by N flips of a fair coin. A fair coin or a biased coin with $P \neq 0.5$ can be easily implemented in a computer, for instance by using `a = floor(rand+P)` in Matlab®. For the individual random numbers we find $\langle A \rangle = P$ and $\sigma_A^2 = P(1-P)$, so that the relative standard deviation for the ensemble with N members becomes $\sigma_S / \langle S \rangle = \sqrt{(1-P)/(N \cdot P)}$.[4]

To compute the explicit probability distribution of the sum of the random numbers for the whole ensemble, we realize that the probability of a subset of n ensemble members providing a 1 and $N - n$ ensemble members providing a 0 is $P^n (1-P)^{N-n}$. The value of the sum associated with this probability is n.

Now we still need to consider the phenomenon already encountered for the sum of the numbers on the black and red dice: Different numbers n have different multiplicities. We have $N!$ permutations of the ensemble members. Let us assign a 1 to the first n members of each permutation. For our problem, it does not matter in which sequence these n members are numbered and it does not matter in which sequence the remaining $N - n$ members are numbered. Hence, we need to divide the total number of permutations $N!$ by the numbers of permutations in each subset, $n!$ and $(N-n)!$ for the first and second subset, respectively. The multiplicity that we need is the number of combinations of N elements to the n^{th} class, which is thus given by the binomial coefficient,

---

[3] Here, each individual spin is considered as a system and the N spins as an ensemble of identical systems

[4] Thanks to Takuya Segawa for pointing out a mistake in this expression in [Swe12]

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} , \tag{3.23}$$

providing the probability distribution

$$P_S(n) = \binom{N}{n} P^n (1-P)^{N-n} . \tag{3.24}$$

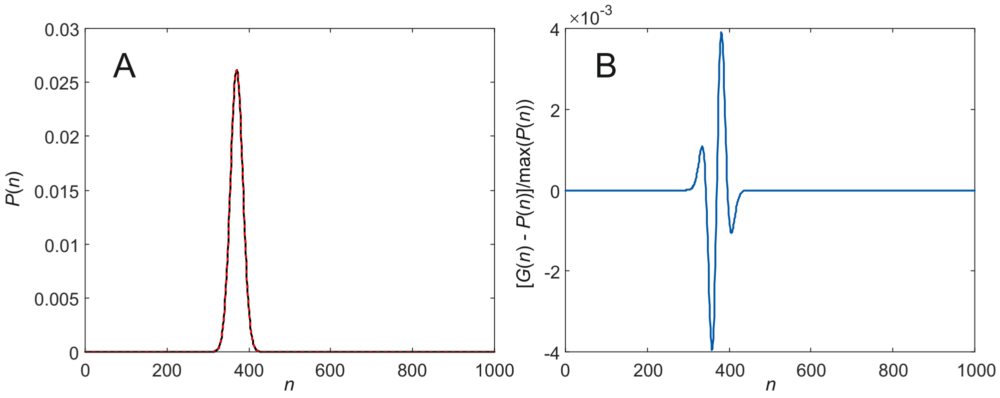


Figure 3.1: Gaussian approximation of the binomial distribution. (A) Gaussian approximation (red dashed line) and binomial distribution (black solid line) for $P = 0.37$ and $N = 1000$. (B) Error of the Gaussian approximation relative to the maximum value of the binomial distribution.

For large values of N the binomial distribution tends to a Gaussian distribution,

$$G(s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(s-\langle s\rangle)^2}{2\sigma^2}\right] . \tag{3.25}$$

As we already know the mean value $\langle s\rangle = \langle n\rangle = NP$ and variance $\sigma_S^2 = NP(1-P)$, we can immediately write down the approximation

$$P_S(n) \approx \frac{1}{\sqrt{2\pi P(1-P)N}} \exp\left[-\frac{(n-PN)^2}{2P(1-P)N}\right] = G(n) . \tag{3.26}$$

As shown in Figure 3.1 the Gaussian approximation of the binomial distribution is quite good already at $N = 1000$.

In fact, the Gaussian (or normal) distribution is a general distribution for the arithmetic mean of a large number of independent random variables:

Concept 3.1.5 — Central limit theorem. Suppose that a large number N of observations has been made with each observation corresponding to a random number that is independent from the random numbers of the other observations. According to the *central limit theorem*, the mean value $\langle S\rangle/N$ of the sum of all these random numbers is approximately normally distributed, regardless of the probability distribution of the individual random numbers, as long all the probability distributions of all individual random numbers are identical.[a] The central limit theorem applies, if each individual random variable has a well-defined mean value (expectation value) and a well-defined variance. These conditions are fulfilled for statistically regular trials $\mathcal{T}$.

---

[a] If the individual random numbers are not identically distributed, the theorem will still apply, if Lyapunov's condition or Lindeberg's condition is fulfilled. See the very useful and detailed Wikipedia article on the Central limit theorem for more information and proofs.

3.1.7 Stirling's Formula

The number $N!$ of permutations increases very fast with N, leading to numerical overflow in calculators and computers at values of N that correspond to nanoclusters rather than to macroscopic samples. Even binomial coefficients, which grow less strongly with increasing ensemble size, cannot be computed with reasonable precision for $N \gg 1000$. Furthermore, the factorial $N!$ is difficult to handle in calculus. The scaling problem can be solved by taking the logarithm of the factorial,

$$\ln N! = \ln \left(\prod_{n=1}^{N} n \right) = \sum_{n=1}^{N} \ln n \, . \tag{3.27}$$

Concept 3.1.6 — Stirling's formula. For large numbers N the natural logarithm of the factorial can be approximated by *Stirling's formula*

$$\ln N! \approx N \ln N - N + 1 \, , \tag{3.28}$$

which amounts to the approximation

$$N! \approx N^N \exp(1 - N) \tag{3.29}$$

for the factorial itself. For large numbers N it is further possible to neglect 1 in the sum and approximate $\ln N! \approx N \ln N - N$.

The absolute error of this approximation for $N!$ looks gross and increases fast with increasing N, but because $N!$ grows much faster, the relative error becomes insignificant already at moderate N. For $\ln N!$ it is closely approximated by $-0.55/N$. In fact, an even better approximation has been found by Gosper [Swe12],

$$\ln N! \approx N \ln N - N + \frac{1}{2} \ln \left[\left(2N + \frac{1}{3} \right) \pi \right] \, . \tag{3.30}$$

Gosper's approximation is useful for considering moderately sized systems, but note that several of our other assumptions and approximations become questionable for such systems and much care needs to be taken in interpreting results. For the macroscopic systems, in which we are mainly interested here, Stirling's formula is often sufficiently precise and Gosper's is not needed.

Slightly better than Stirling's original formula, but still a simple approximation is

$$N! \approx \sqrt{2\pi N} \left(\frac{N}{e} \right)^N \, . \tag{3.31}$$

3.2 Continuous Probability Theory

3.2.1 Probability Density

Although the outcomes of measurements can be discretized, and in fact, are invariably discretized when storing the data, in theory it is convenient to work with continuous variables where physical quantities are assumed to be continuous. For instance, spatial coordinates in phase space are assumed to be continuous, as are the momentum coordinates for translational motion in free space.

To work with continuous variables, we assume that an event can return a real number instead of an integer index. The real number with its associated *probability density* ρ is a *continuous random number*. Note the change from assigning a probability to an event to assigning a probability density. This is necessary as real numbers are not countable and thus the number of possible events is infinite. If we want to infer a probability in the usual sense, we need to specify an interval $[l, u]$ between a lower bound l and an upper bound u. The probability that trial $\mathcal{T}$ will turn up a real number in this closed interval is given by

$$P([l,u]) = \int_l^u \rho(x)\mathrm{d}x \; . \tag{3.32}$$

The probability density must be *normalized*,

$$\int_{-\infty}^{\infty} \rho(x)\mathrm{d}x = 1 \; . \tag{3.33}$$

A probability density distribution can be characterized by its moments.

Concept 3.2.1 — Moment analysis. The n^{th} moment of a probability density distribution is defined as,

$$\langle x^n \rangle = \int_{-\infty}^{\infty} x^n \rho(x)\mathrm{d}x \; . \tag{3.34}$$

The first moment is the mean of the distribution. With the mean $\langle x \rangle$, the central moments are defined

$$\langle (x - \langle x \rangle)^n \rangle = \int_{-\infty}^{\infty} (x - \langle x \rangle)^n \rho(x)\mathrm{d}x \; . \tag{3.35}$$

The second central moment is the variance σ_x^2 and its square root σ_x is the standard deviation.

Probability density is defined along some dimension x, corresponding to some physical quantity. The average of a function $F(x)$ of this quantity is given by

$$\langle F(x) \rangle = \int_{-\infty}^{\infty} F(x)\rho(x)\mathrm{d}x \; . \tag{3.36}$$

R In many books and articles, the same symbol P is used for probabilities and probability densities. This is pointed out by Swendsen [Swe12] who decided to do the same, pointing out that the reader must learn to deal with this. In the next section he goes on to confuse marginal and conditional probability densities with probabilities himself. In these lecture notes we use P for probabilities, which are always unitless, finite numbers in the interval $[0, 1]$ and ρ for probability densities, which are always infinitesimally small and may have a unit. Students are advised to keep the two concepts apart, which means using different symbols.

Computer representations of probability densities by a vector or array are discretized. Hence, the individual values are finite. We now consider the problem of generating a stream of random numbers that conforms to a given discretized probability density $\vec{\rho}$ [Pre+97]. Modern

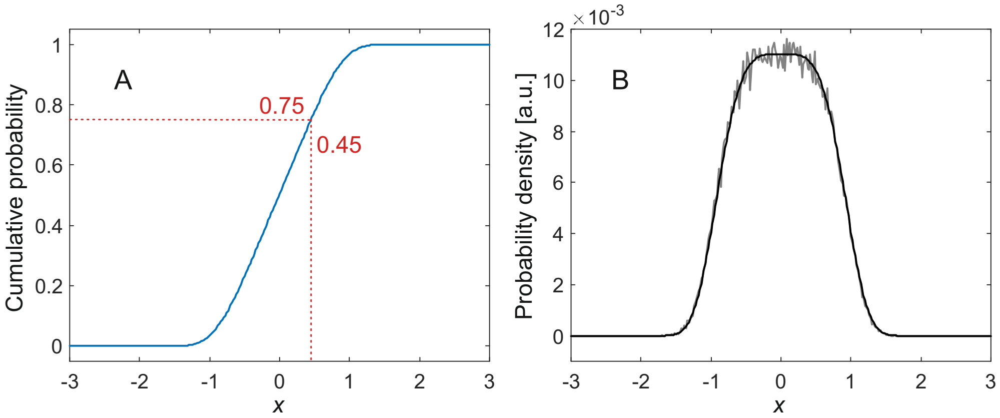


Figure 3.2: Generation of random numbers that conform to a given probability density distribution. (A) Cumulative probability distribution $P(x) = \int_{-\infty}^{x} \rho(\xi)\mathrm{d}\xi$ for $\rho(x) = \exp(-x^4)$ (blue). A pseudo-random number with uniform distribution in $(0, 1)$, here 0.75, selects the ordinate of $P(x)$ (red dashed horizontal line). The corresponding abscissa, here $x = 0.45$ (red dashed vertical line), is an instance of a random number with probability density distribution $\rho(x)$. (B) Distribution of 10^5 random numbers (grey line) and target probability density distribution $\rho(x) = \exp(-x^4)$ (black line).

programming languages or mathematical libraries include functions that provide uniformly distributed pseudo-random numbers in the interval $(0, 1)$ (Matlab®: `rand`) or pseudo-random numbers with a Gaussian (normal) distribution with mean 0 and standard deviation 1 (Matlab®: `randn`). A stream of uniformly distributed pseudo-random numbers in $(0, 1)$ can be transformed to a stream of numbers with probability density conforming to $\vec{\rho}$ by selecting for each input number the abscissa where the cumulative sum of $\vec{\rho}$ (Matlab®: `cumsum(rho)`) most closely matches the input number (Figure 3.2). Note that $\vec{\rho}$ must be normalized (Matlab®: `rho = rho/sum(rho)`). Since a random number generator is usually called very often in a Monte Carlo simulation, the cumulative sum `cumsum_rho` should be computed once for all before the loop over all trials. With this, generation of the abscissa index `poi` becomes a one-liner in Matlab®: `[~,poi] = min(abs(cumsum_rho - rand));`[5]

Coming back to physical theory, the concept of probability density can be extended to multiple dimensions, for instance to the $2F = 2fM$ dimensions of phase space. Probability then becomes a volume integral in this hyperspace. A simple example of a multidimensional continuous problem is the probability of finding a classical particle in a box. The probability to find it at a given point is infinitely small, as there are infinitely many of such points. The probability density is uniform, since all points are equally likely for a classical (unlike a quantum) particle. With the volume V of the box, this uniform probability density is $1/V$ if we have a single particle in the box. This follows from the normalization condition, which is $\int \rho \mathrm{d}V = 1$. Note that a probability density has a unit, in our example m^{-3}. In general, the unit is the inverse of the product of the units of all dimensions.

The marginal probability density for a subset of the events is obtained by 'integrating out' the other events. Let us assume a particle in a two-dimensional box with dimensions x and y and ask about the probability density along x. It is given by

$$\rho_x(x) = \int_{-\infty}^{\infty} \rho(x, y)\mathrm{d}y \; . \tag{3.37}$$

---

[5] This one-liner may cause efficiency problems if computational effort per trial besides random number generation is small.

Likewise, the conditional probability density $\rho(y|x)$ is defined at all points where $\rho_x(x) \neq 0$,

$$\rho(y|x) = \frac{\rho(x,y)}{\rho_x(x)} \; . \tag{3.38}$$

If two continuous random numbers are independent, their joint probability density is the product of the two individual probability densities,

$$\rho(x,y) = \rho_x(x)\rho_y(y) \; . \tag{3.39}$$

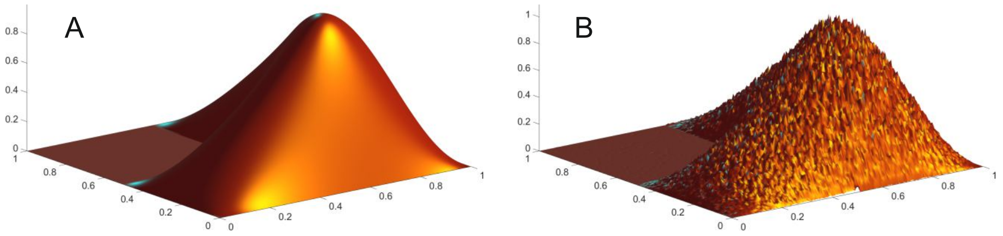


Figure 3.3: Monte Carlo simulation of a two-dimensional probability density distribution. (A) Two-dimensional probability density distribution corresponding to the first-order membrane function used in the Matlab® logo. (B) Distribution of 10^7 random numbers conforming to the probability density distribution shown in (A).

Problem 3.4 Write a Matlab® program that generates random numbers conforming to a two-dimensional probability density distribution ρ_{mem} that resembles the Matlab® logo (see Figure 3.3). The (not yet normalized) distribution ρ_{mem} is obtained with the function call `L = membrane(1,resolution,9,9);`. *Hint:* You can use the `reshape` function to generate a vector from a two-dimensional array as well as for reshaping a vector into a two-dimensional array. That way the two-dimensional problem (or, in general, a multi-dimensional problem) can be reduced to the problem of a one-dimensional probability density distribution.

3.2.2 Selective Integration of Probability Densities

We already know how to compute probability from probability density for a simply connected parameter range. Such a range can be an interval $[l,u]$ for a probability density depending on only one parameter x or a simply connected volume element for a probability density depending on multiple parameters. In a general problem, the points that contribute to the probability of interest may not be simply connected. If we can find a function $g(x)$ that is zero at the points that should contribute, we can solve this problem with the Dirac delta function, which is the continuous equivalent of the Kronecker delta that was introduced above.

Concept 3.2.2 — Dirac delta function. The Dirac delta function is a generalized function with the following properties

1. The function $\delta(x)$ is zero everywhere except at $x = 0$.
2. $\int_{-\infty}^{\infty} \delta(x)\mathrm{d}x = 1$.

The function can be used to select the value $f(x_0)$ of another continuous function $f(x)$,

$$f(x_0) = \int_{-\infty}^{\infty} f(x)\delta(x - x_0)\mathrm{d}x \; . \tag{3.40}$$

This concept can be used, for example, to compute the probability density of a new random variable s that is a function of two given random variables x and y with given joint probability density $\rho(x, y)$. The probability density $\rho(s)$ corresponding to $s = f(x, y)$ is given by

$$\rho(s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \rho(x, y)\delta\left(s - f(x, y)\right) \mathrm{d}x\mathrm{d}y\ . \tag{3.41}$$

Note that the probability density $\rho(s)$ computed that way is automatically normalized.

3.2.3 Sum of Two Continuous Random Numbers

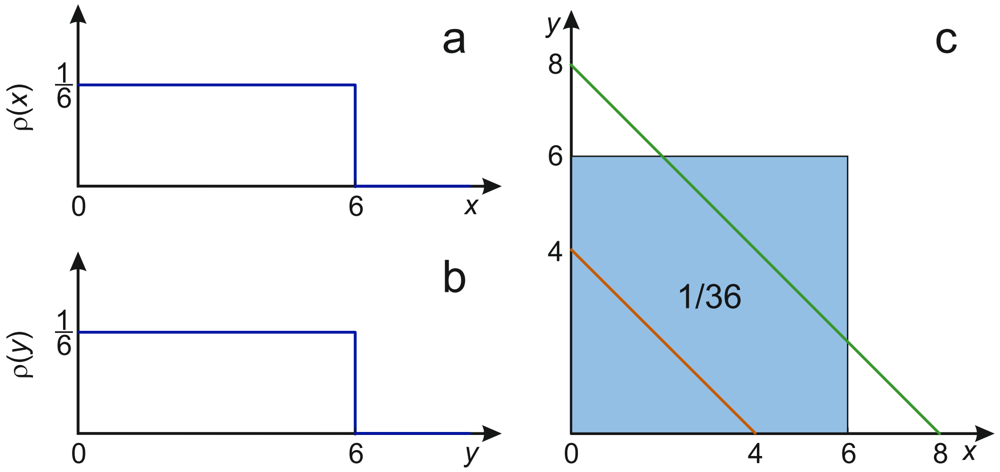


Figure 3.4: Probability density distributions for two continuous random numbers x and y that are uniformly distributed in the interval $[0, 6]$ and have zero probability density outside this interval. a) Marginal probability density $\rho_x(x)$. b) Marginal probability density $\rho_y(y)$. c) Joint probability density $\rho(x, y)$. In the light blue area, $\rho = 1/36$, outside $\rho = 0$. The orange line corresponds to $s = 4$ and the green line to $s = 8$.

We now use the concept of selective integration to compute the probability density $\rho(s)$ for the sum $s = x + y$ of the numbers shown by two continuous dice, with each of them having a uniform probability density in the interval $[0, 6]$ (see Fig. 3.4). We have

$$\rho(s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \rho(x, y)\delta\left(s - (x + y)\right) \mathrm{d}y\mathrm{d}x \tag{3.42}$$

$$= \frac{1}{36}\int_{0}^{6}\int_{0}^{6} \delta\left(s - (x + y)\right) \mathrm{d}y\mathrm{d}x\ . \tag{3.43}$$

The argument of the delta function in the inner integral over y can be zero only for $0 \leq s - x \leq 6$, since otherwise no value of y exists that leads to $s = x + y$. It follows that $x \leq s$ and $x \geq s - 6$. For $s = 4$ (orange line in Fig. 3.4c) the former condition sets the upper limit of the integration. Obviously, this is true for any s with $0 \leq s \leq 6$. For $s = 8$ (orange line in Fig. 3.4c) the condition $x \geq s - 6$ sets the lower limit of the integration, as is also true for any s with $6 \leq s \leq 12$. The lower limit is 0 for $0 \leq s \leq 6$ and the upper limit is 6 for $6 \leq s \leq 12$. Hence,

$$\rho(s) = \frac{1}{36}\int_{0}^{s} \mathrm{d}s = \frac{s}{36} \text{ for } s \leq 6\ , \tag{3.44}$$

and

$$\rho(s) = \frac{1}{36}\int_{s-6}^{6} \mathrm{d}s = \frac{12-s}{36} \text{ for } s \geq 6 \ . \tag{3.45}$$

From the graphical representation in Fig. 3.4c it is clear that $\rho(s)$ is zero at $s = 0$ and $s = 12$, assumes a maximum of $1/6$ at $s = 6$, increases linearly between $s = 0$ and $s = 6$ and decreases linearly between $s = 6$ and $s = 12$.

Statistical Ensembles
Concept of an Ensemble
Ergodicity
Microcanonical Ensemble
Canonical Ensemble
Boltzmann Distribution
Equipartition Theorem
Internal Energy and Heat Capacity of the Canonical Ensemble
Grand Canonical Ensemble

4 — Classical Ensembles

4.1 Statistical Ensembles

4.1.1 Concept of an Ensemble

Probability densities in phase space cannot be computed by considering only a single system at a single instant in time. Such a system will be in some random microstate, but what we need is the statistics of such microstates. This problem was solved by Gibbs, who considered ensembles that consist of a very large number of identical systems in possibly different microstates. The microstates for a system with M molecules with f degrees of freedom each are points in $2fM$-dimensional phase space. If we have information on the probability density assigned to such points, we can use probability theory to compute thermodynamical state functions.

4.1.2 Ergodicity

Instead of considering a large ensemble of systems at the same time (ensemble average), we could also consider a long trajectory of a single system in phase space. The single system will go through different microstates and if we observe it for a sufficiently long time, we might expect that it visits all accessible points in phase space with a frequency that corresponds to the associated probability density. This idea is the basis of analyzing MD trajectories in terms of thermodynamic state functions. The ensemble average $\langle A \rangle$ is replaced by the time average $\overline{A}$. We assume

$$\langle A \rangle = \overline{A} \, . \tag{4.1}$$

Systems where this assumption holds are called *ergodic* systems.

Often, experiments are performed on a large ensemble of identical systems. An example is a spectroscopic experiment on a dilute solution of chromophores: Each chromophore can be considered as an individual system and their number may be of the order of 10^{10} or higher. In some cases an equivalent experiment can be performed on a single chromophore, but such single-molecule experiments require many repetitions and measure a time-average. The results of ensemble and single-molecule experiments are equivalent if the system is ergodic and the measurement time in the single-molecule experiment is sufficiently long.

Whether or not a system is ergodic depends on kinetic accessibility of the whole thermodynamically accessible phase space. We shall see later that thermodynamic accessibility is related to temperature and to the energy assigned to points in phase space. Points are accessible if their energy is not too much higher than the energy minimum in phase space. Whether a single dynamic system visits all these points at the same given temperature- and what time it needs to sample phase space- depends on energy barriers. In MD simulations sampling problems are often encountered, where molecular conformations that are thermodynamically accessible are not accessed within reasonable simulation times. A multitude of techniques exists for alleviating such sampling problems, none of them perfect. In general, time-average methods, be they computational or experimental, should be interpreted only with care in terms of thermodynamics. In this lecture course we focus on ensemble-average methods, which suffer from a loss in dynamic information, but get the thermodynamic state functions right.

4.2 Microcanonical Ensemble

Assume that we have an isolated system with N particles in a fixed volume V. Because the system is isolated, the total energy E must also be fixed. If we know that the energy must be in an interval $[E, E+\Delta E]$ the probability density in phase space must be zero everywhere outside the region between the two hypersurfaces with constant energies E and $E+\Delta E$. We call this region the energy shell in which the system is confined. If the system is in equilibrium, i.e., the probability density ρ is stationary, ρ must be uniform in this energy shell, i.e., it must not depend on p and q within this shell. We can see this from the Liouville equation (2.8), whose left-hand side must be zero for a stationary probability density. The Poisson bracket on the right-hand side will vanish if ρ is uniform.[1]

Concept 4.2.1 — Microcanonical ensemble. An ensemble with a constant number N of particles in a constant volume V and with constant total energy E has a uniform probability density ρ_mc in the part of phase space, where it can reside, which is the energy hypersurface at energy E. Such an ensemble is called a *microcanonical ensemble*.

We are left with computing this constant probability density ρ_mc. As the energy is given by the Hamiltonian function $\mathcal{H}(\mathbf{p},\mathbf{q})$, we can formally write ρ_mc for an infinitely thin energy shell ($\Delta E \rightarrow 0$) as

$$\rho_\mathrm{mc} = \frac{1}{\Omega(E)} \delta\left(E - \mathcal{H}(\mathbf{p},\mathbf{q})\right) \, , \tag{4.2}$$

where the *statistical weight* Ω depends on energy, volume, and number of particles N, but at constant energy does not depend on momentum $\mathbf{p}$ or spatial coordinates $\mathbf{q}$. Since the probability density is normalized, we have

$$\Omega(E) = \int \int \delta\left(E - \mathcal{H}(\mathbf{p},\mathbf{q})\right) \mathrm{d}\mathbf{q}\mathrm{d}\mathbf{p} \, . \tag{4.3}$$

The probability density in phase space of the microcanonical ensemble is thus relatively easy to compute. However, the restriction to constant energy, i.e. to an isolated system, severely limits application of the microcanonical ensemble. To see this, we consider the simplest system, an electron spin $S = 1/2$ in an external magnetic field B_0. This system is neither classical nor describable in phase space, but it will nicely serve our purpose. The system has a state space consisting of only two states $|\alpha\rangle$ and $|\beta\rangle$ with energies $\epsilon_\alpha = \hbar g_e \mu_\mathrm{B} B_0/2$ and

---

[1] It is more tricky to argue that it will *only* vanish if ρ is uniform. However, as the individual particles follow random phase space trajectories, it is hard to imagine that the right-hand side could be stationary zero unless ρ is uniform.

$\epsilon_\beta = -\hbar g_e \mu_\mathrm{B} B_0/2$.[2] In magnetic resonance spectroscopy, one would talk of an ensemble of 'isolated' spins, if the individual spins do not interact with each other. We shall see shortly that this ensemble is not isolated in a thermodynamical sense, and hence not a microcanonical ensemble.

The essence of the microcanonical ensemble is that all systems in the ensemble have the *same* energy E, this restricts probability density to the hypersurface with constant E. If our ensemble of N spins would be a microcanonical ensemble, this energy would be either $E = \hbar g_e \mu_\mathrm{B} B_0/2$ or $E = -\hbar g_e \mu_\mathrm{B} B_0/2$ and all spins in the ensemble would have to be in the same state, i.e., the ensemble would be in a pure state. In almost any experiment on spins $S = 1/2$ the ensemble is in a mixed state and the populations of states $|\alpha\rangle$ and $|\beta\rangle$ are of interest. The system is not isolated, but, via spin relaxation processes, in thermal contact with its environment. To describe this situation, we need another type of ensemble.

4.3 Canonical Ensemble

Equilibrium thermodynamics describes systems that are in thermal equilibrium. In an ensemble picture, this can be considered by assuming that the system is in contact with a very large— for mathematical purposes infinitely large— heat bath. Because of this, the individual systems in the ensemble can differ in energy. However, the probability density distribution in phase space or state space must be consistent with constant temperature T, which is the temperature of the heat bath. In experiments, it is the temperature of the environment.

Concept 4.3.1 — Canonical ensemble. An ensemble with a constant number N of particles in a constant volume V and at thermal equilibrium with a heat bath at constant temperature T can be considered as an ensemble of microcanonical subensembles with different energies ϵ_i. The energy dependence of probability density conforms to the Boltzmann distribution. Such an ensemble is called a *canonical ensemble*.

R Because each system can exchange heat with the bath and thus change its energy, systems will transfer between subensembles during evolution. This does not invalidate the idea of microcanonical subensembles with constant particle numbers N_i. For a sufficiently large ensemble at thermal equilibrium the N_i are constants of motion.

There are different ways of deriving the Boltzmann distribution. Most of them are rather abstract and rely on a large mathematical apparatus. The derivation gets lengthy if one wants to create the illusion that we know why the constant β introduced below *always* equals $1/k_\mathrm{B}T$, where $k_\mathrm{B} = R/N_\mathrm{Av}$ is the Boltzmann constant, which in turn is the ratio of the universal gas constant R and the Avogadro constant N_Av. Here we follow a derivation [WF12] that is physically transparent and relies on a minimum of mathematical apparatus that we have already introduced.

4.3.1 Boltzmann Distribution

Here we digress from the ensemble picture and use a system of N particles that may exist in r different states with energies ϵ_i with $i = 0 \ldots r-1$. The number of particles with energy ϵ_i is N_i. The particles do not interact, they are completely independent from each other. We could therefore associate theses particles with microcanonical subensembles of a canonical ensemble, but the situation is easier to picture with particles. The probability $P_i = N_i/N$ to find a particle with energy ϵ_i can be associated with the probability density for the microcanonical subensemble at energy ϵ_i. The difference between this simple derivation and the more elaborate derivation for a canonical ensemble is thus essentially the difference between discrete and continuous probability theory. We further assume that the particles are classical particles and thus distinguishable.

---

[2] Where g_e is the g value of the free electron and μ_B the Bohr magneton.

To compute the probability distribution $P(i) = N_i/N$, we note that

$$\sum_0^{r-1} N_i = N \tag{4.4}$$

and

$$\sum_0^{r-1} N_i \epsilon_i = E \; , \tag{4.5}$$

where E is a constant total energy of the system. We need to be careful in interpreting the latter equation in the ensemble picture. The quantity E corresponds to the energy of the whole canonical ensemble, which is indeed a constant of motion, if we consider a sufficiently large number of systems in contact with a thermal bath. We can thus use our simple model of N particles for guessing the probability density distribution in the canonical ensemble.

What we are looking for is the most likely distribution of the N particles on the r energy levels. This is equivalent to putting N distinguishable balls into r boxes. We did already solve the problem of distributing N objects to 2 states when considering the binomial distribution in Section 3.1.6. The statistical weight of a configuration with n objects in the first state and $N - n$ objects in the second state was $\binom{N}{n}$. With this information we would already be able to solve the problem of a canonical ensemble of N spins $S = 1/2$ in thermal contact with the environment, disregarding for the moment differences between classical and quantum statistics (see Section 6.2).

Coming back to N particles and r energy levels, we still have $N!$ permutations. If we assign the first N_0 particles to the state with energy ϵ_0, the next N_1 particles to ϵ_1 and so on, we need to divide each time by the number of permutations $N_i!$ in the same energy state, because the sequence of particles with the same energy does not matter. We call the vector of the occupation numbers N_i a *configuration*. The configuration specifies one particular macrostate of the system and the relative probability of the macrostates is given by their statistical weights,

$$\Omega = \frac{N!}{N_0! N_1! \ldots N_{r-1}!} \; . \tag{4.6}$$

The most probable macrostate is the one with maximum statistical weight Ω. Because of the peaking of probability distributions for large N, we need to compute only this most probable macrostate; it is representative for the whole ensemble. Instead of maximizing Ω we can as well maximize $\ln \Omega$, as the natural logarithm is a strictly monotonous function. This allows us to apply Stirling's formula,

$$\begin{aligned} \ln \Omega &= \ln N! - \sum_0^{r-1} \ln N_i! & (4.7) \\ &= N \ln N - N + 1 - \sum_0^{r-1} N_i \ln N_i + \sum_0^{r-1} N_i - r \; . & (4.8) \end{aligned}$$

By inserting Eq. (4.4) we find

$$\ln \Omega = N \ln N - \sum_0^{r-1} N_i \ln N_i + 1 - r \; . \tag{4.9}$$

Note that the second term on the right-hand side of Eq. (4.9) has some similarity to the entropy of mixing, which suggests that $\ln \Omega$ is related to entropy.

At the maximum of $\ln \Omega$ the derivative of $\ln \Omega$ with respect to the N_i must vanish,

$$0 = \delta \sum_i N_i \ln N_i = \sum_i \left(N_i \delta \ln N_i + \delta N_i \ln N_i \right) = \sum_i \delta N_i + \sum_i \ln N_i \delta N_i \; . \tag{4.10}$$

In addition, we need to consider the boundary conditions of constant particle number, Eq. (4.4),

$$\delta N = \sum_i \delta N_i = 0 \tag{4.11}$$

and constant total energy, Eq. (4.5),

$$\delta E = \sum_i \epsilon_i \delta N_i = 0 \; . \tag{4.12}$$

It might appear that Eq. (4.11) could be used to cancel a term in Eq. (4.10), but this would be wrong as Eq. (4.11) is a constraint that must be fulfilled separately. For the constrained maximization we can use the method of Lagrange multipliers.

Concept 4.3.2 — Lagrange multipliers. The maximum or minimum of a function $f(x_1 \ldots, x_n)$ of n variables is a stationary point that is attained at

$$\delta f = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)_{x_k \neq x_i} \partial x_i = 0 \; . \tag{4.13}$$

We now consider the case where the possible sets of the n variables are constrained by c additional equations

$$g_j(x_1, x_2, \ldots, x_n) = 0 \; , \tag{4.14}$$

where index j runs over the c constraints ($j = 1 \ldots c$). Each constraint introduces another equation of the same form as the one of Eq. (4.13),

$$\delta g_j = \sum_{i=1}^{n} \left(\frac{\partial g_j}{\partial x_i} \right)_{x_k \neq x_i} \partial x_i = 0 \; . \tag{4.15}$$

The constraints can be introduced by multiplying each of the c equations by a multiplier λ_j and subtracting it from the equation for the stationary point without the constraints,

$$\mathcal{L} = \sum_{i=1}^{n} \left[\left(\frac{\partial f}{\partial x_i} \right)_{x_k \neq x_i} - \sum_{j=1}^{c} \lambda_j \left(\frac{\partial g_j}{\partial x_i} \right)_{x_k \neq x_i} \right] \partial x_i \; . \tag{4.16}$$

If a set of variables $\{x_{0,1} \ldots, x_{0,n}\}$ solves the constrained problem then there exists a set $\{\lambda_{0,1} \ldots \lambda_{0,r}\}$ for which $\{x_{0,1}, x_{0,2}, \ldots, x_{0,n}\}$ also corresponds to a stationary point of the Lagrangian function $\mathcal{L}(x_1, \ldots, x_n, \lambda_1, \ldots \lambda_r)$. Note that not all stationary points of the Lagrangian function are necessarily solutions of the constrained problem. This needs to be checked separately.

With this method, we can write

$$0 = \sum_i \delta N_i + \sum_i \ln N_i \delta N_i + \alpha \sum_i \delta N_i + \beta \sum_i \epsilon_i \delta N_i \tag{4.17}$$

$$= \sum_i \delta N_i \left(1 + \ln N_i + \alpha + \beta\epsilon_i\right) \ . \tag{4.18}$$

The two boundary conditions fix only two of the population numbers N_i. We can choose the multipliers α and β in a way that $(1 + \ln N_i + \alpha + \beta\epsilon_i) = 0$ for these two N_i, which ensures that the partial derivatives of $\ln \Omega$ with respect to these two N_i vanishes. The other $r - 2$ population numbers can, in principle, be chosen freely, but again we must have

$$1 + \ln N_i + \alpha + \beta\epsilon_i = 0 \tag{4.19}$$

for all i to make sure that we find a maximum with respect to variation of any of the r population numbers. This gives

$$N_i = \gamma e^{-\beta\epsilon_i} \tag{4.20}$$

with $\gamma = e^{-(1+\alpha)}$. We can eliminate γ by using Eq. (4.4),

$$\sum_i N_i = \gamma \sum_i e^{-\beta\epsilon_i} = N \ , \tag{4.21}$$

giving

$$\gamma = \frac{N}{\sum_i e^{-\beta\epsilon_i}} \ , \tag{4.22}$$

and finally leading to

$$P(i) = \frac{N_i}{N} = \frac{e^{-\beta\epsilon_i}}{\sum_i e^{-\beta\epsilon_i}} \ . \tag{4.23}$$

For many problems in statistical thermodynamics, the Lagrange multiplier α is related to the chemical potential by $\alpha = \mu/(k_\mathrm{B}T)$. The Lagrange multiplier β must have the reciprocal dimension of an energy, as the exponent must be dimensionless. As indicated above, we cannot at this stage prove that β is the same energy for all problems of the type that we have posed here, let alone for all of the analogous problems of canonical ensembles. The whole formalism can be connected to phenomenological thermodynamics *via* Maxwell's kinetic gas theory (see also Section 4.3.2). For this problem one finds

$$\beta = \frac{1}{k_\mathrm{B}T} \ . \tag{4.24}$$

Concept 4.3.3 — Boltzmann distribution. For a classical canonical ensemble with energy levels ϵ_i the probability distribution for the level populations is given by the *Boltzmann distribution*

$$P_i = \frac{N_i}{N} = \frac{e^{-\epsilon_i/k_\mathrm{B}T}}{\sum_i e^{-\epsilon_i/k_\mathrm{B}T}} \ . \tag{4.25}$$

The sum over states

$$Z(N, V, T) = \sum_i e^{-\epsilon_i/k_\mathrm{B}T} \tag{4.26}$$

required for normalization is called *canonical partition function*.[a] The partition function is a thermodynamical state function.

[a] The dependence on N and V arises, because these parameters influence the energy levels

For the partition function, we use the symbol Z relating to the German term *Zustandssumme*, which is a more lucid description of this quantity.

4.3.2 Equipartition Theorem

Comparison of Maxwell's kinetic theory of gases with the state equation of the ideal gas from phenomenological thermodynamics provides a mean kinetic energy of a point particle of $\langle \epsilon_{\mathrm{kin}} \rangle = 3k_{\mathrm{B}}T/2$. This energy corresponds to

$$\epsilon_{\mathrm{trans}} = \frac{1}{2}mv^2 = \frac{1}{2m}p^2 \; , \tag{4.27}$$

i.e., it is quadratic in the velocity coordinates of dynamic space or the momentum coordinates of phase space. Translational energy is distributed via three degrees of freedom, as the velocities or momenta have components along three pairwise orthogonal directions in space. Each quadratic degree of freedom thus contributes a mean energy of $k_{\mathrm{B}}T/2$.

If we accept that the Lagrange multiplier β assumes a value $1/k_{\mathrm{B}}T$, we find a mean energy $k_{\mathrm{B}}T$ of an harmonic oscillator in the high-temperature limit [WF12]. Such an oscillator has two degrees of freedom that contribute quadratically in the degrees of freedom to energy,

$$\epsilon_{\mathrm{vib}} = \frac{1}{2}\mu v^2 + \frac{1}{2}fx^2 \; , \tag{4.28}$$

where μ is the reduced mass and f the force constant. The first term contributes to kinetic energy, the second to potential energy. In the time average, each term contributes the same energy and assuming ergodicity this means that each of the two degrees of freedom contributes with $k_{\mathrm{B}}T/2$ to the average energy of a system at thermal equilibrium.

The same exercise can be performed for rotational degrees of freedom with energy

$$\epsilon_{\mathrm{rot}} = \frac{1}{2}I\omega^2 \; , \tag{4.29}$$

where I is angular momentum and ω angular frequency. Each rotational degree of freedom, being quadratic in ω again contributes a mean energy of $k_{\mathrm{B}}T/2$.

Based on Eq. (4.23) it can be shown [WF12] that for an energy

$$\epsilon_i = \eta_0 + \eta_1 + \eta_2 + \ldots = \sum_{k=1}^{f} \eta_k \; , \tag{4.30}$$

where index k runs over the individual degrees of freedom, the number of molecules that contribute energy η_k does not depend on the terms η_j with $j \neq k$. It can be further shown that

$$\langle \eta_k \rangle = \frac{1}{2\beta} \tag{4.31}$$

for all terms that contribute quadratically to energy.[3]

This result has two consequences. First, we can generalize $\beta = 1/k_{\mathrm{B}}T$, which we strictly knew only for translational degrees of freedom, to any canonical ensemble for which all individual energy contributions are quadratic along one dimension in phase space. Second, we can formulate the

---

[3] The condition of a quadratic contribution arises from an assumption that is made when integrating over the corresponding coordinate.

Theorem 4.3.4 — Equipartition theorem. Each degree of freedom, whose energy scales quadratically with one of the coordinates of state space, contributes a mean energy of $k_\mathrm{B}T/2$.

R The equipartition theorem applies to all degrees of freedom that are *activated*. Translational degrees of freedom are always activated and rotational degrees of freedom are activated at ambient temperature, which corresponds to the high-temperature limit of rotational dynamics. To vibrational degrees of freedom the equipartition theorem applies only in the high-temperature limit. In general, the equipartition theorem fails for quantized degrees of freedom if the quantum energy spacing is comparable to $k_\mathrm{B}T/2$ or exceeds this value. We shall come back to this point when discussing the vibrational partition function.

4.3.3 Internal Energy and Heat Capacity of the Canonical Ensemble

The internal energy u of a system consisting of N particles that are distributed to r energy levels can be identified as the total energy E of the system considered in Section (4.3.1). Using Eqs. (4.5) and (4.25) we find

$$u = N\frac{\sum_i \epsilon_i e^{-\epsilon_i/k_\mathrm{B}T}}{\sum_i e^{-\epsilon_i/k_\mathrm{B}T}} = N\frac{\sum_i \epsilon_i e^{-\epsilon_i/k_\mathrm{B}T}}{Z} . \tag{4.32}$$

The sum in the numerator can be expressed by the partition function, since

$$\frac{\mathrm{d}Z}{\mathrm{d}T} = \frac{1}{k_\mathrm{B}T^2}\sum_i \epsilon_i e^{-\epsilon_i/k_\mathrm{B}T} . \tag{4.33}$$

Thus we obtain

$$u = Nk_\mathrm{B}T^2 \cdot \frac{1}{Z} \cdot \frac{\mathrm{d}Z}{\mathrm{d}T} = Nk_\mathrm{B}T^2\frac{\mathrm{d}\ln Z}{\mathrm{d}T} . \tag{4.34}$$

Again the analogy of our simple system to the canonical ensemble holds. At this point we have computed one of the state functions of phenomenological thermodynamics from the set of energy levels. The derivation of the Boltzmann distribution has also indicated that $\ln\Omega$, and thus the partition function Z are probably related to entropy. We shall see in Section 5.2 that this is indeed the case and that we can compute *all* thermodynamic state functions from Z.

Here we can still derive the heat capacity c_V at constant volume, which is the partial derivative of internal energy with respect to temperature. To that end we note that the partition function for the canonical ensemble relates to constant volume and constant number of particles.

$$c_V = \left(\frac{\partial u}{\partial T}\right)_V = N\frac{\partial}{\partial T}\left(k_\mathrm{B}T^2\frac{\partial \ln Z}{\partial T}\right)_V = -N\frac{\partial}{\partial T}\left(k_\mathrm{B}\frac{\partial \ln Z}{\partial 1/T}\right)_V \tag{4.35}$$

$$= -Nk_\mathrm{B}\left(\frac{\partial\left[\partial \ln Z/\partial 1/T\right]}{\partial T}\right)_V = \frac{Nk_\mathrm{B}}{T^2}\left(\frac{\partial\left[\partial \ln Z/\partial 1/T\right]}{\partial 1/T}\right)_V \tag{4.36}$$

$$= \frac{k_\mathrm{B}}{T^2}\left(\frac{\partial^2 \ln z}{\partial\left(1/T\right)^2}\right)_V . \tag{4.37}$$

In the last line of Eq (4.37) we have substituted the molecular partition function Z by the partition function for the whole system, $\ln z = N\ln Z$. Note that this implies a generalization. Before, we were considering a system of N identical particles. Now we implicitly assume that Eq. (4.37), as well as $u = k_\mathrm{B}T^2\frac{\mathrm{d}\ln z}{\mathrm{d}T}$ will hold for any system, as long as we correctly derive the system partition function z.

We note here that the canonical ensemble describes a closed system that can exchange heat with its environment, but by definition it cannot exchange work, because its volume V is constant.

This does not present a problem, since the state functions can be computed at different V. In particular, pressure p can be computed from the partition function as well (see Section 5.2). However, because the canonical ensemble is closed, it cannot easily be applied to all problems that involve chemical reactions. For this we need to remove the restriction of a constant number of particles in the systems that make up the ensemble.

4.4 Grand Canonical Ensemble

For the description of an open system in the thermodynamical sense, i.e., a system that can exchange not only heat, but also matter with its environment, we need to replace particle number N with another constant of motion. If we would fail to introduce a new constant of motion, we would end up with a system that is not at equilibrium and thus cannot be fully described by time-independent state functions. If we assume that the system is in chemical as well as thermal equilibrium with its environment, the new constant of motion is the chemical potential μ, or more precisely, a vector $\boldsymbol{\mu}$ of the chemical potentials μ_k of all components.

Concept 4.4.1 — Grand canonical ensemble. An ensemble with constant chemical potential μ_k of all components, and constant volume V that is at thermal equilibrium with a heat bath at constant temperature T and in chemical equilibrium with its environment is called a *grand canonical ensemble*. It can be considered as consisting of canonical subensembles with different particle numbers N. The grand canonical state energies and partition function contain an additional chemical potential term. With this additional term the results obtained for the canonical ensemble apply to the grand canonical ensemble, too.

The partition function for the grand canonical ensemble is given by

$$Z_{\mathrm{gc}}(\boldsymbol{\mu}, V, T) = \sum_i e^{\left(\sum_k N_{i,k}\mu_k - \epsilon_i\right)/k_{\mathrm{B}}T} \,, \tag{4.38}$$

whereas the probability distribution over the levels and particle numbers is

$$P_i = \frac{e^{\left(\sum_k N_{i,k}\mu_k - \epsilon_i\right)/k_{\mathrm{B}}T}}{Z_{\mathrm{gc}}} \,. \tag{4.39}$$

Note that the index range i is much larger than for a canonical ensemble, because each microstate is now characterized by a set of particle numbers $N_{i,k}$, where k runs over the components.

R At this point we are in conflict with the notation that we used in the lecture course on phenomenological thermodynamics (PC I `http://www.epr.ethz.ch/education/Thermodynamics_PCI/`). In that course we defined the chemical potential μ as a *molar* quantity, here it is a molecular quantity. The relation is $\mu_{\mathrm{PCI}} = N_{\mathrm{Av}}\mu_{\mathrm{PCVI}}$. Using the PC I notation in the current lecture notes would be confusing in other ways, as μ is generally used in statistical thermodynamics for the molecular chemical potential. A similar remark applies to capital letters for state functions. Capital letters denote either a molecular quantity or a molar quantity. The difference will be clear from the context. We note that in general small letters for state functions (except for pressure p) denote extensive quantities and capital letters (except for volume V) denote intensive quantities.

Swendsen’s Postulates of Thermodynamics
  Cautionary Remarks on Entropy
  Swendsen’s Postulates
  Entropy in Phenomenological Thermodynamics
  Boltzmann’s Entropy Definition
The Relation of State Functions to the Partition Function
  Entropy and the Partition Function
  Helmholtz Free Energy
  Gibbs Free Energy, Enthalpy, and Pressure
Irreversibility
  Historical Discussion
  Irreversibility as an Approximation
Entropy and Information
  Gibbs Entropy
  Von Neumann Entropy
  Shannon Entropy

UNCERTAINTY
DISPERSAL
CHANGE
CYCLE
ENTROPY
CONSTANT
STATISTICAL
FUNDAMENTAL
MECHANICS
LAW
NUMBER
POSITED
CONSERVED
CLASSICAL
MOLECULAR
ROOM
PRESSURE
QUANTUM
EQUILIBRIUM
SURROUNDING
RESERVOIR
THERMODYNAMIC
WATER
INTERPRETATIVE
VOLUME
THERMAL
REACTANTS
INTERNAL
MOLAR
ENTROPY
ARROW
PROBABILITY
VARIABLES
DYNAMICAL
ISOLATED
MACROSCOPIC
SYSTEM
KELVIN
STATE
GAS
AMOUNT
UNIVERSE
CHEMICAL
TOTAL
ENERGY
LOGARITHM
DISSIPATION
MICROSCOPIC
INFORMATION

5 — Entropy

5.1 Swendsen’s Postulates of Thermodynamics

5.1.1 Cautionary Remarks on Entropy

You will search in vain for a mathematical derivation or clear condensed explanation of entropy in textbooks and textbook chapters on statistical thermodynamics (please tell me if you have found one). There is a simple reason for it: no such derivation or explanation exists. With entropy being a central concept, probably *the* central concept of the theory, this may appear very strange. However, the situation is not as bad as it may appear. The theory and the expressions that can be derived work quite well and have predictive power. There are definitions of entropy in statistical thermodynamics (unfortunately, more than one) and they make some sense. Hence, while it may be unnerving that we cannot derive the central state function from scratch, we can still do many useful things and gain *some* understanding.

Textbooks tend to sweep the problem under the rug. We won’t do that here. We try to make an honest attempt to clarify what we do know and what we don’t know about entropy before accepting one working definition and base the rest of theory on this definition. It is probably best to start with a set of postulates that explains what we expect from the quantity that we want to define.

5.1.2 Swendsen’s Postulates

The following postulates are introduced and shortly discussed in Section 9.6 of Swendsen’s book [Swe12]. We copy the long form of these postulates verbatim with very small alterations that improve consistency or simplify the expression.

1. There exist equilibrium states of a macroscopic system that are characterized uniquely by a small number of extensive variables.
2. The values assumed at equilibrium by the extensive variables of an isolated system in the absence of internal constraints are those that maximize the entropy over the set of all constrained macroscopic states.
3. The entropy of a composite system is additive over the constituent subsystems.
4. For equilibrium states the entropy is a monotonically increasing function of the energy.
5. The entropy is a continuous and differentiable function of the extensive variables.

We have omitted Swendsen’s last postulate (The entropy is an extensive function of the extensive variables), because, strictly speaking, it is superfluous. If the more general third postulate of additivity is fulfilled, entropy is necessarily an extensive property.

Swendsen's first postulate (*Equilibrium States*) establishes the formalism of thermodynamics, while all the remaining postulates constitute a wish list for the quantity *entropy* that we need to predict the equilibrium states. They are a wish list in the sense that we cannot prove that a quantity with all these properties must exist. We can, however, test any proposed definition of entropy against these postulates.

Some points need explanation. First, the set of postulates defines entropy as a state function, although this may be hidden. The first postulate implies that in equilibrium thermodynamics some extensive variables are state functions and that a small set of such state functions completely specifies all the knowledge that we can have about a macroscopic system. Because entropy in turn specifies the other state functions for an isolated system at equilibrium, according to the second postulate (*Entropy Maximization*), it must be a state function itself. It must be an extensive state function because of the third postulate (*Additivity*), but the third postulate requires more, namely that entropies can be added not only for subsystems of the same type in the same state, but also for entirely different systems. This is required if we want to compute a new equilibrium state (or entropy change) after unifying different systems. Otherwise, the simple calorimetry experiment of equilibrating a hot piece of copper with a colder water bath would already be outside our theory. The fourth postulate (*Monotonicity*) is new compared to what we discussed in phenomenological thermodynamics. For a classical ideal gas this postulate can be shown to hold. This postulate is needed because it ensures that temperature is positive. The fifth postulate is a matter of mathematical convenience, although it may come as a surprise in a theory based on integer numbers of particles. We assume, as at many other points, that the system is sufficiently large for neglecting any errors that arise from treating particle number as a real rather than an integer number. In other words, these errors must be smaller than the best precision that we can achieve in experiments. As we already know from phenomenological thermodynamics, the fifth postulate does not apply to first-order phase transitions, where entropy has a discontinuity. We further note that the second postulate is an alternative way of writing the Second Law of Thermodynamics. The term 'in the absence of internal constraints' in the second postulate ensures that the whole state space (or, for systems fully described by Hamiltonian equations of motion, the whole phase space) is accessible.

5.1.3 Entropy in Phenomenological Thermodynamics

Textbook authors are generally much more comfortable in discussing entropy as an abstract state function in phenomenological thermodynamics than in discussing its statistical thermodynamics aspects. We recall that the concept of entropy is not unproblematic in phenomenological thermodynamics either. We had accepted the *definition* of Clausius entropy,

$$\mathrm{d}s = \frac{\mathrm{d}q_{\mathrm{rev}}}{T} \,, \tag{5.1}$$

where $\mathrm{d}q_{\mathrm{rev}}$ is the differentially exchanged heat for a reversible process that leads to the same differential change in other state variables as an irreversible process under consideration and T is the temperature. We could then show that entropy is a state function (Carnot process and its generalization) and relate entropy via its total differential to other state functions. With this definition we could further show that for closed systems, which can exchange heat, but not volume work with their environment ($\mathrm{d}V = 0$), minimization of Helmholtz free energy $f = u - Ts$ provides the equilibrium state and that for closed systems at constant pressure ($\mathrm{d}p = 0$), minimization of Gibbs free energy $g = h - Ts$ provides the equilibrium state. Partial molar Gibbs free energy is the chemical potential $\mu_{k,\mathrm{molar}}$ and via $\mu_{k,\mathrm{molecular}} = \mu_{k,\mathrm{molar}}/N_{\mathrm{Av}}$ it is related to terms in the partition function of the grand canonical ensemble, where we have abbreviated $\mu_{k,\mathrm{molecular}}$ as μ_k (Section 4.4).

We were unable in phenomenological thermodynamics to prove that the definition given in Eq. (5.1) ensures fulfillment of the Second Law. We were able to give plausibility arguments why such a quantity should increase in some spontaneous processes, but not more.

5.1.4 Boltzmann's Entropy Definition

Boltzmann provided the first statistical definition of entropy, by noting that it is the logarithm of probability, up to a multiplicative and an additive constant. The formula $s = k \ln W$ by Planck, which expresses Boltzmann's definition, omits the additive constant. We shall soon see why.

We now go on to test Boltzmann's definition against Swendsen's postulates. From probability theory and considerations on ensembles we know that for a macroscopic system, probability density distributions for an equilibrium state are sharply peaked at their maximum. In other words, the macrostate with largest probability is such a good representative for the equilibrium state that it serves to predict state variables with better accuracy that the precision of experimental measurements. It follows strictly that any definition of entropy that fulfills Swendsen's postulates must make s a monotonously increasing function of probability density[1] *for an isolated system*.

Why the logarithm? Let us express probability (for the moment discrete again) by the measure of the statistical weights of macrostates. We consider the isothermal combination of two independent systems A with entropies s_A and s_B to a total system with entropy $s = s_\mathrm{A} + s_\mathrm{B}$ [WF12]. The equation for total entropy is a direct consequence of Swendsen's third postulate. On combination, the statistical weights Ω_A and Ω_B multiply, since the subsystems are independent. Hence, with the monotonously increasing function $f(\Omega)$ we must have

$$s = f(\Omega) = f(\Omega_\mathrm{A} \cdot \Omega_\mathrm{B}) = f(\Omega_\mathrm{A}) + f(\Omega_\mathrm{B}) \ . \tag{5.2}$$

The only solutions of this functional equation are logarithm functions. What logarithm we choose will only influence the multiplicative constant. Hence, we can write

$$s = k \ln \Omega \ , \tag{5.3}$$

where, for the moment, constant k is unknown. Boltzmann's possible additive constant must vanish at this point, because with such a constant, the functional equation (5.2), which specifies additivity of entropy, would not have a solution.

It is tempting to equate Ω in Eq. (5.3) in the context of phase space problems with the volume of phase space occupied by the system. Indeed, this concept is known as *Gibbs entropy* (see Section 5.4.1). It is plausible, since the phase space volume specifies a statistical weight for a continuous problem. No problem arises if Gibbs entropy is used for equilibrium states as it then coincides with Boltzmann entropy. There exists a conceptual problem, however, if we consider *approach* to equilibrium. The Liouville theorem (see Section 2.2.2) states that the volume in phase space taken up by a system is a constant of motion.[2]. Hence, Gibbs entropy is a constant of motion for an isolated system and the equilibrium state would be impossible to reach from any non-equilibrium state, which would necessarily occupy a smaller phase space volume. This leads to the following cautionary remark:

> R Statistical thermodynamics, as we introduce it in this lecture course, does not describe dynamics that leads from non-equilibrium to equilibrium states. Different equilibrium states can be compared and the equilibrium state can be determined, but we have made a number of assumptions that do not allow us to apply our expressions and concepts to non-equilibrium states without further thought. Non-equilibrium statistical thermodynamics is explicitly outside the scope of the theory that we present here.

---

[1] Boltzmann was thinking in terms of discrete probability theory. As we want to use continuous probability theory here, we have made the transition from probability to probability density.

[2] The theorem relies on uniform distribution in this volume at some point in time, but it applies here, as we have seen before that such uniform distribution in an energy shell is a feature of the equilibrium state of an isolated system.

A conceptual complication with Boltzmann's definition is that one might expect s to be maximal at equilibrium for a closed system, too, not only for an isolated system. In classical thermodynamics we have seen, however, that the equilibrium condition for a closed system is related to free energy. Broadly, we could say that for a closed system probability must be maximized for the system and its environment together. Unfortunately, this cannot be done mathematically as the environment is very large (in fact, for mathematical purposes infinite). The solution to this problem lies in the treatment of the canonical ensemble (Section 4.3). In that treatment we have seen that energy enters into the maximization problem via the boundary condition of constant total energy of the system that specifies what exactly is meant by thermal contact between the system and its environment. We can, therefore, conclude that Boltzmann's entropy definition, as further specified in Eq. (5.3), fulfills those of Swendsen's postulates that we have already tested and that the core idea behind it, maximization of probability (density) at equilibrium is consistent with our derivation of the partition function for a canonical ensemble at thermal equilibrium. We can thus fix k in Eq. (5.3) by deriving s from the partition function.

5.2 The Relation of State Functions to the Partition Function

5.2.1 Entropy and the Partition Function

We recall that we already computed internal energy u and heat capacity c_V at constant volume from the system partition function z (Section 4.3.3). For a canonical system ($V = \mathrm{const.}$), which is by definition at thermal equilibrium (reversible), we can identify q_rev in Eq. (5.1) with[3]

$$\mathrm{d}q_\mathrm{rev} = c_V \mathrm{d}T \; . \tag{5.4}$$

Definite integration with substitution of c_V by Eq. (4.35) gives [WF12],

$$s - s_0 = \int_0^T \frac{c_V}{T}\mathrm{d}T' = \int_0^T \frac{1}{T}\frac{\partial}{\partial T}\left(k_\mathrm{B}T^2\frac{\partial \ln z}{\partial T}\right)_V \mathrm{d}T' \tag{5.5}$$

$$= \int_0^T \frac{1}{T}\left[k_\mathrm{B}T^2\left(\frac{\partial^2 \ln z}{\partial T^2}\right)_V + 2k_\mathrm{B}T\left(\frac{\partial \ln z}{\partial T}\right)_V\right]\mathrm{d}T' \tag{5.6}$$

$$= k_\mathrm{B}\int_0^T T\left(\frac{\partial^2 \ln z}{\partial T^2}\right)_V \mathrm{d}T + 2k_\mathrm{B}\int_0^T \left(\frac{\partial \ln z}{\partial T}\right)_V \mathrm{d}T' \; . \tag{5.7}$$

Partial integration provides

$$s - s_0 = k_\mathrm{B}T\left(\frac{\partial \ln z}{\partial T}\right)_V - k_\mathrm{B}\int_0^T \left(\frac{\partial \ln z}{\partial T}\right)_V \mathrm{d}T + 2k_\mathrm{B}\int_0^T \left(\frac{\partial \ln z}{\partial T}\right)_V \mathrm{d}T \tag{5.8}$$

$$= k_\mathrm{B}T\left(\frac{\partial \ln z}{\partial T}\right)_V + k_\mathrm{B}\ln z\Big|_0^T \tag{5.9}$$

$$= \frac{u}{T} + k_\mathrm{B}\ln z - k_\mathrm{B}\left(\ln z\right)_{T=0} \; , \tag{5.10}$$

where we have used Eq. (4.34) to substitute the first term on the right hand side of Eq. (5.9). If we assume that $\lim\limits_{T\rightarrow 0} u/T = 0$, the entropy at an absolute temperature of zero can be identified as $s_0 = k_\mathrm{B}\left(\ln z\right)_{T=0}$. If there are no degenerate ground states, $s_0 = 0$ in agreement with

---

[3] Purists of statistical thermodynamics will shudder, as we now rely on the entropy definition of phenomenological thermodynamics. We hide the fact that we are incapable of a strict general derivation and just relate the new concepts of statistical thermodynamics to the concepts of phenomenological thermodynamics. In effect, we show how state functions of phenomenological thermodynamics must be computed if both Boltzmann's and Clausius' entropy definitions apply.

Nernst's theorem (Third Law of Thermodynamics), as will be discussed in Section 7.2.2. Thus, by associating $u = 0$ with $T = 0$ we obtain

$$s = \frac{u}{T} + k_\mathrm{B} \ln z = k_\mathrm{B} \left[\left(\frac{\partial \ln z}{\partial \ln T} \right)_V + \ln z \right] . \tag{5.11}$$

We see that under the assumptions that we have made the entropy can be computed from the partition function. In fact, there should be a unique mapping between the two quantities, as both the partition function and the entropy are state functions and thus must be uniquely defined by the state of the system.

We now proceed with computing constant k in the mathematical definition of Boltzmann entropy, Eq. (5.3). By inserting Eq. (4.9) into Eq. (5.3) we have

$$s = k \left(N \ln N - \sum_{i=0}^{r-1} N_i \ln N_i \right) . \tag{5.12}$$

We have neglected the term r on the right-hand side of Eq. (4.9), as is permissible if the number N of particles is much larger than the number r of energy levels. Furthermore, according to Eq. (4.25) and the definition of the partition function, we have $N_i = N e^{-\epsilon_i/k_\mathrm{B}T}/Z$. Hence,

$$s = k \left[N \ln N - N \sum_{i=0}^{r-1} \frac{e^{-\epsilon_i/k_\mathrm{B}T}}{Z} \ln N \left(\frac{e^{-\epsilon_i/k_\mathrm{B}T}}{Z} \right) \right] \tag{5.13}$$

$$= k \left[N \ln N - N \sum_{i=0}^{r-1} \frac{e^{-\epsilon_i/k_\mathrm{B}T}}{Z} \ln N + N \sum_{i=0}^{r-1} \frac{e^{-\epsilon_i/k_\mathrm{B}T}}{Z} \ln Z + N \sum_{i=0}^{r-1} \frac{e^{-\epsilon_i/k_\mathrm{B}T}}{Z} \cdot \frac{\epsilon_i}{k_\mathrm{B}T} \right] \tag{5.14}$$

$$= k \left[N \ln N - N \ln N + N \ln Z + \frac{N}{k_\mathrm{B}T} \frac{\sum_{i=0}^{r-1} \epsilon_i e^{-\epsilon_i/k_\mathrm{B}T}}{Z} \right] , \tag{5.15}$$

where we have used the definition of the partition function of going from Eq. (5.14) to (5.15). Using Eq. (4.32) for substitution in the last term on the right-hand side of Eq. (5.15), we find

$$s = k \left[N \ln Z + \frac{u}{k_\mathrm{B}T} \right] . \tag{5.16}$$

Comparison of Eq. (5.16) with Eq. (5.11) gives two remarkable results. First, the multiplicative constant k in Boltzmann's entropy definition can be identified as $k = k_\mathrm{B} = R/N_\mathrm{Av}$. Second, for the system of N identical, distinguishable classical particles, we must have

$$z_\mathrm{dist} = Z^N . \tag{5.17}$$

In other words, the partition function of a system of N identical, distinguishable, non-interacting particles is the N^th power of the molecular partition function.

It turns out that Eq. (5.17) leads to a contradiction if we apply it to an ideal gas. Assume that we partition the system into two subsystems with particle numbers $N_\mathrm{sub} = N/2$. The internal-energy dependent term in Eq. (5.16) obviously will not change during this partitioning. For the partition-function dependent term we have $N \ln Z$ for the total system and $2(N/2) \ln Z'$ for the sum of the two subsystems. The molecular partition function in the subsystems differs, because volume available to an individual particle is only half as large as in the total system. For the inverse process of unifying the two subsystems we would thus obtain a mixing entropy, although the gases in the subsystems are the same. This appearance of a mixing entropy for two identical ideal gases is called the Gibbs paradox. The Gibbs paradox can be healed by treating the particles as indistinguishable. This reduces the statistical weight Ω by $N!$ for the total system and by $(N/2)!$ for each subsystem, which just offsets the volume effect. Hence, for an ideal gas we have

$$z_{\mathrm{indist}} = \frac{1}{N!} Z^N \, . \tag{5.18}$$

It may appear artificial to treat classical particles as indistinguishable, because the trajectory of each particle could, in principle, be followed if they adhere to classical mechanics equations of motion, which we had assumed. Note, however, that we discuss a macrostate and that we have explicitly assumed that we cannot have information on the microstates, i.e., on the trajectories. In the macrostate picture, particles in an ideal gas are, indeed, indistinguishable. For an ideal crystal, on the other hand, each particle could be individually addressed, for instance, by high resolution microscopy. In this case, we need to use Eq. (5.17).

5.2.2 Helmholtz Free Energy

Helmholtz free energy (*German:* Freie Energie) f is defined as

$$f = u - Ts \, . \tag{5.19}$$

This equation has a simple interpretation. From phenomenological thermodynamics we know that the equilibrium state of a closed systems corresponds to a minimum in free energy. Among all macrostates with the same energy u at a given temperature T, the equilibrium state is the one with maximum entropy s. Furthermore, using Eq. (5.11) we have

$$f = u - T\left(u/T + k_{\mathrm{B}} \ln z\right) \tag{5.20}$$

$$= -k_{\mathrm{B}} T \ln z \, . \tag{5.21}$$

We note that this value of f, which can be computed from only the canonical partition function and temperature, corresponds to the global minimum over all macrostates. This is not surprising. After all, the partition function was found in a maximization of the probability of the macrostate.

5.2.3 Gibbs Free Energy, Enthalpy, and Pressure

All ensembles that we have defined correspond to equilibrium states at constant volume. To make predictions for processes at constant pressure or to compute enthalpies $h = u + pV$ and Gibbs free energies $g = f + pV$ we need to compute pressure from the partition function. The simplest way is to note that $p = -\left(\partial f/\partial V\right)_{T,n}$. With Eq. (5.21) it then follows that

$$p = k_{\mathrm{B}} T \left(\frac{\partial \ln z}{\partial V}\right)_T \, , \tag{5.22}$$

where we have skipped the lower index n indicating constant molar amount. This is permissible for the canonical ensemble, where the number of particles is constant by definition. From Eq. (5.22) it follows that

$$pV = k_{\mathrm{B}} T \left(\frac{\partial \ln z}{\partial \ln V}\right)_T \tag{5.23}$$

and

$$h = u + pV = k_{\mathrm{B}} T \left[\left(\frac{\partial \ln z}{\partial \ln T}\right)_V + \left(\frac{\partial \ln z}{\partial \ln V}\right)_T\right] \, . \tag{5.24}$$

Connoisseurs will notice the beautiful symmetry of this equation.

With Eq. (5.22) we can also compute Gibbs free energy (*German:* freie Enthalpie),

$$g = f + pV = -k_\mathrm{B}T\left[\ln z - \left(\frac{\partial \ln z}{\partial \ln V}\right)_T\right] . \tag{5.25}$$

5.3 Irreversibility

5.3.1 Historical Discussion

Daily experience tells us that some processes are irreversible. Phenomenological thermodynamics had provided recipes for recognizing such processes by an increase in entropy for an isolated system or decrease of free energy for a closed system. When Boltzmann suggested a link between classical mechanics of molecules on a microscopic level and irreversibility of processes on the macroscopic level, many physicists were irritated nevertheless. In retrospect it is probably fair to say that a controversial discussion of Boltzmann's result could only ensue because atomistic or molecular theory of matter was not yet universally accepted at the time. It is harder to understand why this discussion is still going on in textbooks. Probably this is related to the fact that physicists in the second half of the 19^th and first half of the 20^th believed that pure physics has implications in philosophy, beyond the obvious ones in epistemology applied to experiments in the sciences. If statistical mechanics is used to predict the future of the universe into infinite times, problems ensue. If statistical mechanics is properly applied to well-defined experiments there are no such problems.

Classical mechanics of particles does not involve irreversibility. The equations of motion have time reversal symmetry and the same applies to quantum-mechanical equations of motion. If the sign of the Hamiltonian can be inverted, the system will evolve backwards along the same trajectory in phase space (or state space) that it followed to the point of inversion. This argument is called *Umkehreinwand* or Loschmidt paradox and was brought up (in its classical form) by Loschmidt. The argument can be refined and is then known as the *central paradox*: Each microstate can be assigned a time-reversed state that evolves, under the same Hamiltonian, backwards along the same trajectory. The two states should have the same probability. The central paradox confuses equilibrium and non-equilibrium dynamics. At equilibrium a state and the corresponding time-reversed state indeed have the same probability, which explains that the macrostate of the system does not change and why processes that can be approximated by a series of equilibrium states are reversible. If, on the other hand, we are not at equilibrium, there is no reason for assuming that the probabilities of any two microstates are related. The system is at some initial condition with a given set of probabilities and we are not allowed to pose symmetry requirements to this initial condition.

The original *Umkehreinwand*, which is based on sign inversion of the Hamiltonian rather then the momenta of microstates, is more serious than the central paradox. Time-reversal experiments of this type can be performed, for instance, echo experiments in magnetic resonance spectroscopy and optical spectroscopy. In some of these echo experiments, indeed the Hamiltonian is sign-inverted, in most of these experiments application of a perturbation Hamiltonian for a short time (pulse experiment) causes sign inversion of the density matrix. Indeed, the first paper on observation of such a spin echo by Erwin Hahn was initially rejected with the argument that he could not have observed what he claimed, as this would have violated the Second Law of Thermodynamics. A macroscopic 'time-reversal' experiment that creates a 'colorant echo' in corn syrup can be based on laminar flow [UNM12]. We note here that all these time-reversal experiments are based on preparing a system in a non-equilibrium state. To analyze them, changes in entropy or Helmholtz free energy must be considered during the evolution that can be reversed. These experiments do not touch the question whether or not the same system will irreversibly approach an equilibrium state if left to itself for a sufficiently long time. We can see this easily for the experiment with colorants and corn syrup. If, after setup of the initial state and evolution to the point of time reversal, a long time would pass, the colorant echo would no longer be observed, because diffusion of the colorants in corn syrup would destroy spatial correlation. The echo relies on the fact that diffusion of the colorants in corn syrup can be neglected on the time scale of the experiment, i.e., that equilibrium cannot be reached. The same is true for the spin echo experiment, which fails if the evolution time is much longer than the transverse relaxation time of the spins.

Another argument against irreversibility was raised by Zermelo, based on a theorem by Poincaré. The theorem states that any isolated classical system will return repeatedly to a point in phase space that is arbitrarily close to the starting point. This argument is known as *Wiederkehreinwand* or Zermelo paradox. We note that such quasi-periodicity is compatible with the probability density formalism of statistical mechanics. The probability density distribution is very sharply peaked at the equilibrium state, but it is not zero at the starting point in phase space. The system fluctuates around the equilibrium state and, because the distribution is sharply peaked, these fluctuations are very small *most of the time*. Once in a while the fluctuation is sufficiently large to revisit even a very improbable starting point in phase space, but for a macroscopic system this while is much longer than the lifetime of our galaxy. For *practical purposes* such large fluctuations can be safely neglected, because they occur so rarely. That a system will never evolve far from the equilibrium state once it had attained equilibrium is an approximation, but the approximation is better than many other approximations that we use in physics. The statistical error that we make is certainly much smaller than our measurement errors.

5.3.2 Irreversibility as an Approximation

If the whole of phase space is accessible the system will always *tend* to evolve from a less probable macrostate to a more probable macrostate, until it has reached the most probable macrostate, which is the equilibrium state. Equilibrium is dynamic. The microstate of each individual system evolves in time. However, for most microstates the values of all state variables are the same as for equilibrium within experimental uncertainty. In fact, the fraction of such microstates does not significantly differ from unity. Hence, a system that has attained equilibrium once will be found at equilibrium henceforth, as long as none of the external parameters is changed on which the probability density distribution in phase space depends. In that sense, processes that run from a non-equilibrium state to an equilibrium state are irreversible.

We should note at this point that all our considerations in this lecture course assume systems under thermodynamic control. If microstate dynamics in phase space is slow compared to the time scale of the experiment or simulation, the equilibrium state may not be reached. This may also happen if dynamics is fast in the part of phase space where the initial state resides but exchange dynamics is too slow between this part of phase space and the part of phase space where maximum probability density is located.

5.4 Entropy and Information

5.4.1 Gibbs Entropy

For a system with a countable number of microstates an ensemble entropy can be defined by a weighted sum over entropies of all microstates that are in turn expressed as $-k_\mathrm{B} \ln P_i$, which is analogous to Boltzmann's entropy definition for a macrostate.

$$S = -k_\mathrm{B} \sum_i P_i \ln P_i \; . \tag{5.26}$$

This is the definition of Gibbs entropy, while Boltzmann entropy is assigned to an individual microstate. Note that we have used a capital S because Gibbs entropy is a molecular entropy. Using Eq. (4.25), we obtain for the system entropy $s = NS$,

$$s = -k_{\mathrm{B}} N \sum_i P_i \left(-\frac{\epsilon_i}{k_{\mathrm{B}} T} - \ln Z \right) \tag{5.27}$$

$$= \frac{u}{T} + k_B \ln z \; , \tag{5.28}$$

where we have assumed distinguishable particles, so that $\ln z = N \ln Z$. We have recovered Eq. (5.11) that we had derived for the system entropy starting from Boltzmann entropy and assuming a canonical ensemble. For a canonical ensemble of distinguishable particles, either concept can be used. As noted above, Gibbs entropy leads to the paradox of a positive mixing entropy for combination of two subsystems made up by the same ideal gas. More generally, Gibbs entropy is not extensive if the particles are indistinguishable. The problem can be solved by redefining the system partition function as in Eq. (5.18).

This problem suggests that entropy is related to the information we have on the system. Consider mixing of $^{13}\mathrm{CO}_2$ with $^{12}\mathrm{CO}_2$.[4] At a time when nuclear isotopes were unknown, the two gases could not be distinguished and mixing entropy was zero. With a sufficiently sensitive spectrometer we could nowadays observe the mixing process by ^{13}C NMR. We will observe spontaneous mixing. Quite obviously, the mixing entropy is not zero anymore.

This paradox cautions against philosophical interpretation of entropy. Entropy is a quantity that can be used for predicting the outcome of physical experiments. It *presumes an observer* and depends on the information that the observer has or can obtain.[5] Statistical mechanics provides general recipes for defining entropy, but the details of a proper definition depend on experimental context.

Unlike the system entropy derived from Boltzmann entropy via the canonical ensemble, Gibbs entropy is, in principle, defined for non-equilibrium states. Because it is based on the same probability concept, Gibbs entropy in an isolated system is smaller for non-equilibrium states than for equilibrium states.

5.4.2 Von Neumann Entropy

The concept of Gibbs entropy for a countable set of discrete states and their probabilities is easily extended to continuous phase space and probability densities. This leads to the *von Neumann entropy*,

$$S = -k_{\mathrm{B}} \mathrm{Trace} \left\{ \rho \ln \rho \right\} \; , \tag{5.29}$$

where ρ is the density matrix. Some physics textbooks don't distinguish von Neumann entropy from Gibbs entropy [Sch06]. Von Neumann entropy is a constant of motion if an ensemble of classical systems evolves according to the Liouville equation or a quantum mechanical system evolves according to the Liouville-von-Neumann equation. It cannot describe the approach of an isolated system to equilibrium. Coupling of the quantum mechanical system to an environment can be described by the *stochastic Liouville equation*[Kub63; Tan06; VF74]

$$\frac{\partial \widehat{\rho}}{\partial t} = -\frac{i}{\hbar} \left[\widehat{\mathcal{H}}, \widehat{\rho} \right] + \widehat{\widehat{\Gamma}} \left(\widehat{\rho} - \widehat{\rho}_{\mathrm{eq}} \right) \; , \tag{5.30}$$

---

[4] This thought experiment was suggested to me by Roland Riek.

[5] One can speculate on philosophical interpretations. Irreversibility could be a consequence of partitioning the universe into an observer and all the rest, a notion that resonates with intuitions of some mystical thinkers across different religious traditions. Although the idea is appealing, it cannot be rationally proved. Rational thought already implies that an observer exists.

where $\widehat{\widehat{\Gamma}}$ is a Markovian operator and ρ_{eq} the density matrix at equilibrium. This equation of motion can describe quantum dissipative systems, i.e., the approach to equilibrium, without relying explicitly on the concept of entropy, except for the computation of ρ_{eq}, which relies on generalization of the Boltzmann distribution (see Section 6.1.2). However, to derive the Markovian operator $\widehat{\widehat{\Gamma}}$, explicit assumptions on the coupling between the quantum mechanical system and its environment must be made, which is beyond the scope of this lecture course.

5.4.3 Shannon Entropy

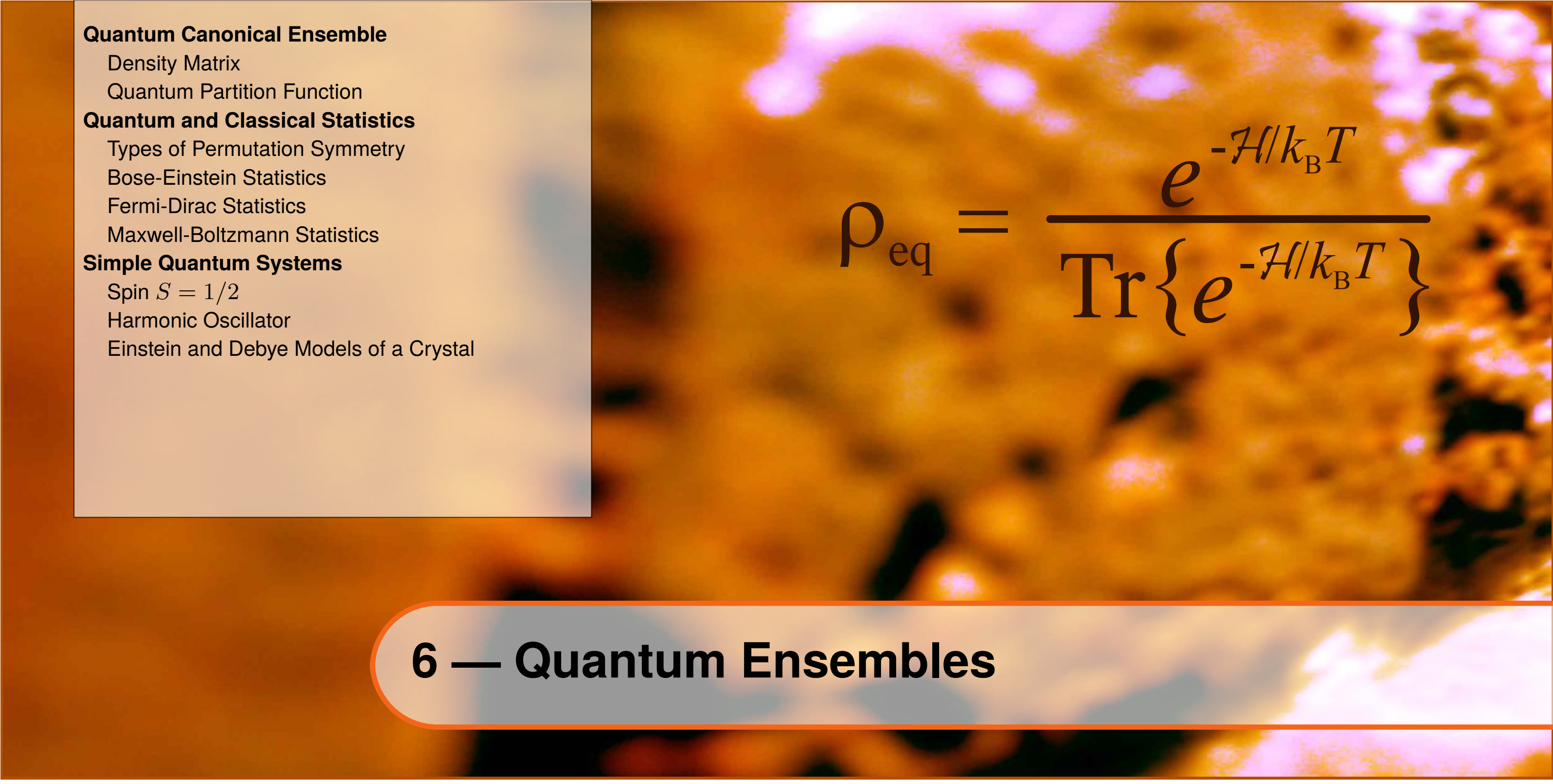
The concept of entropy has also been introduced into information theory. For any discrete random number that can take values a_j with probabilities $P(a_j)$, the *Shannon entropy* is defined as

$$H_{\mathrm{Shannon}}\left(a\right) = -\sum_j P(a_j) \log_2 P(a_j) \ . \tag{5.31}$$

A logarithm to the basis of 2 is used here as the information is assumed to be coded by binary numbers. Unlike for discrete states in statistical mechanics, an event may be in the set but still have a probability $P(a_j) = 0$. In such cases, $P(a_j) \log_2 P(a_j)$ is set to zero. Shannon entropy is the larger the 'more random' the distribution is, or, more precisely, the closer the distribution is to a uniform distribution. Information is considered as deviation from a random stream of numbers or characters. The higher the information content is, the lower the entropy.

Shannon entropy can be related to reduced Gibbs entropy $\sigma = S/k_{\mathrm{B}}$. It is the amount of Shannon information that is required to specify the microstate of the system if the macrostate is known. When expressed with the binary logarithm, this amount of Shannon information specifies the number of yes/no questions that would have to be answered to specify the microstate. We note that this is exactly the type of experiment presumed in the second Penrose postulate (Section 2.3.1). The more microstates are consistent with the observed macrostate, the larger is this number of questions and the larger are Shannon and Gibbs entropy. The concept applies to non-equilibrium states as well as to equilibrium states. It follows, what was stated before Shannon by G. N. Lewis: "Gain in entropy always means loss of information, and nothing more". The equilibrium state is the macrostate that lacks most information on the underlying microstate.

We can further associate order with information, as any ordered arrangement of objects contains information on how they are ordered. In that sense, loss of order is loss of information and increase of disorder is an increase in entropy. The link arises via probability, as the total number of arrangements is much larger than the number of arrangements that conform to a certain order principle. Nevertheless, the association of entropy with disorder is only colloquial, because in most cases we do not have quantitative descriptions of order.



Quantum Canonical Ensemble
  Density Matrix
  Quantum Partition Function
Quantum and Classical Statistics
  Types of Permutation Symmetry
  Bose-Einstein Statistics
  Fermi-Dirac Statistics
  Maxwell-Boltzmann Statistics
Simple Quantum Systems
  Spin $S = 1/2$
  Harmonic Oscillator
  Einstein and Debye Models of a Crystal

$$\rho_{\mathrm{eq}} = \frac{e^{-\mathcal{H}/k_{\mathrm{B}}T}}{\mathrm{Tr}\{e^{-\mathcal{H}/k_{\mathrm{B}}T}\}}$$

6 — Quantum Ensembles

6.1 Quantum Canonical Ensemble

6.1.1 Density Matrix

We have occasionally referred to the quantum-mechanical density matrix ρ in previous sections. Before we discuss quantum ensembles, we need to fully specify this concept.

Concept 6.1.1 — Density Matrix. The microstates that can be assumed by a system in a quantum ensemble are specified by a possible set of wavefunctions ψ_i $(i = 1 \ldots r)$. The probability or *population* of the i^{th} microstate is denoted as P_i, and for the continuous case the probability density for a given wavefunction is denoted as $p(\psi)$. The *density operator* is then given by

$$\widehat{\rho} = \sum_{i=0}^{r-1} P(i) \left|\psi_i\right\rangle \left\langle\psi_i\right| \;\; \text{(discrete)} \tag{6.1}$$

$$\widehat{\rho} = \int_{\psi} p(\psi) \left|\psi_i\right\rangle \left\langle\psi_i\right| \;\; \text{(continuous)} \; . \tag{6.2}$$

Note that the discrete case is closely related to the problem with r energy levels that we discussed in deriving the Boltzmann distribution for a classical canonical ensemble. The density operator can be expressed as a *density matrix* ρ with respect to a set of basis functions $|k\rangle$. For exact computations the basis functions must form a countable complete set that allows for expressing the system wavefunctions ψ_i as linear combinations of basis functions. For approximate computations, it suffices that this linear combination is a good approximation. The matrix elements of the density matrix are then given by

$$\rho_{nm} = \sum_{i=0}^{r-1} P(i) \left\langle m \left|\psi_i\right\rangle \left\langle\psi_i\right| n\right\rangle \;\; \text{(discrete)} \tag{6.3}$$

$$\rho_{nm} = \int_{\psi} p(\psi) \left\langle m \left|\psi_i\right\rangle \left\langle\psi_i\right| n\right\rangle \;\; \text{(continuous)} \; . \tag{6.4}$$

With the complex coefficients c_k in the linear combination representation $|\psi\rangle = \sum_k c_k |k\rangle$, the matrix elements are

$$\rho_{nm} = \overline{c_n c_m^*} \,, \tag{6.5}$$

where the asterisk denotes the complex conjugate and the bar for once denotes the ensemble average. It follows that diagonal elements ($m = n$) are necessarily real, $\rho_{nn} = |c_n|^2$ and that ρ_{mn} is the complex conjugate of ρ_{nm}. Therefore, the density matrix is Hermitian and the density operator is self-adjoint. The matrix dimension is the number of basis functions. It is often convenient to use the eigenfunctions of the system Hamiltonian $\widehat{\mathcal{H}}$ as the basis functions, but the concept of the density matrix is not limited to this choice. The meaning of elemnts of the density matrix is visualized in Figure 6.1.

R That the density matrix can be expressed in the basis of eigenstates does not imply that the ensemble can be represented as consisting of only eigenstates, as erroneously stated by Swendsen [Swe12]. Off-diagonal elements of the density matrix denote coherent superpositions of eigenstates, or short *coherences*. This is not apparent in Swendsen's simple example where coherence is averaged to zero by construction. The ensemble can be represented as consisting of only eigenstates if coherence is absent. In that case the density matrix is diagonal in the eigenbasis. Diagonal elements of the density matrix denote *populations* of basis states.

$$\begin{array}{c|c|c|c|c|}
 & |0\rangle & |1\rangle & \dots & |r\text{-}1\rangle \\ \hline
\langle 0| & p_0 & \overline{c_n c_m^*} & \dots & \dots \\ \hline
\langle 1| & \overline{c_m c_n^*} & p_1 & \dots & \dots \\ \hline
\dots & \dots & \dots & \dots & \dots \\ \hline
\langle r\text{-}1| & \dots & \dots & \dots & p_{r\text{-}1} \\ \hline
\end{array}$$

Figure 6.1: Density matrix for a system with r states numbered $|0\rangle$, $|1\rangle$, $\ldots |r-1\rangle$. The columns correspond to kets, the rows to bras of the wavefunctions corresponding to these states. Diagonal elements $p_k = \rho_{kk}$ denote state populations and are real numbers. Off-diagonal elements $\overline{c_n c_m^*} = \rho_{nm}$ are complex numbers and denote coherent superpositions of states $|n\rangle$ and $|m\rangle$ or, short, coherences.

In quantum mechanics, it is well defined what information we can have about the macrostate of a system, because quantum measurements are probabilistic even for a microstate. We can observe only quantities that are quantum-mechanical observables and these observables are represented by operators $\widehat{A}$. It can be shown that the expectation value $\left\langle \widehat{A} \right\rangle$ of any observable can be computed from the density matrix by [SJ01; Swe12]

$$\left\langle \widehat{A} \right\rangle = \mathrm{Trace}\left\{ \widehat{\rho} \widehat{A} \right\} \,, \tag{6.6}$$

where we have used operator notation for $\widehat{\rho}$ to point out that $\widehat{\rho}$ and $\widehat{A}$ must be expressed in the same basis.

Since the expectation values of all observables are the full information that we can have on a quantum system, the density matrix specifies the full information that we can have on the ensemble. However, the density matrix does not fully specify the ensemble itself, i.e., we cannot infer the probabilities $P(i)$ or probability density function $\rho(\psi)$ from the density matrix (Swendsen gives a simple example [Swe12]). This is another example for the information loss on microstates that comes about when we can only observe macrostates and that is conceptually equivalent to entropy. The von-Neumann entropy can be computed from the density matrix by Eq. (5.29).

We note that there is one important distinction between classical and quantum-mechanical observations *for an individual system*. In the quantum case we can specify only an expectation value, and the second and third Penrose postulates (Section 2.3.1) do not apply: neither can we simultaneously measure all observables (they may be incompatible), nor is the outcome of a later measurement independent of the current measurement. However, quantum uncertainty is much smaller than measurement errors for the large ensembles that we treat by statistical thermodynamics. Hence, the Penrose postulates apply to the quantum-mechanical *ensembles* that represent macrostates, although they do not apply to the microstates.

If all systems in a quantum ensemble populate the same microstate, i.e., they correspond to the same wavefunction, the ensemble is said to be in a *pure state*. A pure state corresponds to minimum rather than maximum entropy. Otherwise the system is said to be in a *mixed state*.

6.1.2 Quantum Partition Function

Energy quantization leads to a difficulty in using the microcanonical ensemble. The difficulty arises because the microcanonical ensemble requires constant energy, which restricts our abilities to assign probabilities in a set of discrete energy levels. However, as we derived the Boltzmann distribution, partition function, entropy and all other state functions for classical systems from the canonical ensemble anyway, we can simply ignore this problem. The canonical ensemble is considered to be at thermal equilibrium with a heat bath (environment) of infinite size. It does not matter whether this heat bath is of classical or quantum mechanical nature. For an infinitely sized quantum system, the energy spectrum is continuous, which allows us to exchange energy between the bath and any constituent system of the quantum canonical ensemble at will.

We can derive Boltzmann distribution and partition function for the density matrix by analogy to the classical case. For that we consider the density matrix in the eigenbasis. The energies of the eigenstates are the eigenvalues ϵ_i of the Hamiltonian $\mathcal{H}$. All arguments and mathematical steps from Section 4.3.1 still apply, with a single exception: Quantum mechanics allows for microstates that are coherent superpositions of eigenstates. The classical derivation carries over if and only if we can be sure that the equilibrium density matrix can be expressed without contributions from such microstates, which would lead to off-diagonal elements in the representation in the eigenbasis of $\widehat{\mathcal{H}}$. This argument can indeed be made. Any superposition of two eigenstates $|n\rangle$ and $|m\rangle$ with amplitudes $|c_n|$ and $|c_m|$ can be realized with arbitrary phase difference $\Delta\phi$ between the two eigenfunctions. The microstates with the same $|c_n|$ and $|c_m|$ but different $\Delta\phi$ all have the same energy. The entropy of a subensemble that populates these microstates is maximal if the distribution of phase differences $\Delta\phi$ is uniform in the interval $[0, 2\pi)$. In that case $\overline{c_m^* c_n}$ vanishes, i.e., such subensembles will not contribute off-diagonal elements to the equilibrium density matrix.

We can now arrange the $e^{-\epsilon_i/k_\mathrm{B}T}$ in matrix form,

$$\xi = e^{-\widehat{\mathcal{H}}/k_\mathrm{B}T} , \tag{6.7}$$

with the matrix elements $\xi_{ii} = e^{-\epsilon_i/k_\mathrm{B}T}$ and $\xi_{ij} = 0$ for $i \neq j$. The partition function is the sum of all the diagonal elements of this matrix, i.e. the trace of ξ. Hence,

$$\widehat{\rho}_{\mathrm{eq}} = \frac{e^{-\widehat{\mathcal{H}}/k_{\mathrm{B}}T}}{\mathrm{Trace}\left\{e^{-\widehat{\mathcal{H}}/k_{\mathrm{B}}T}\right\}} , \tag{6.8}$$

where we have used operator notation. This implies that Eq. (6.8) can be evaluated in any basis, not only the eigenbasis of $\widehat{\mathcal{H}}$. In a different basis, $e^{-\epsilon_i/k_{\mathrm{B}}T}$ needs to be computed as a matrix exponential and, in general, the density matrix ρ_{eq} will have non-zero off-diagonal elements in such a different basis.

The quantum-mechanical partition function,

$$Z = \mathrm{Trace}\left\{e^{-\widehat{\mathcal{H}}/k_{\mathrm{B}}T}\right\} , \tag{6.9}$$

is independent of the choice of basis, as the trace of a matrix is invariant under unitary transformations. Note that we have used a capital Z for a molecular partition function. This is appropriate, as the trace of $\widehat{\rho}_{\mathrm{eq}}$ in Eq. (6.8) is unity. In the eigenbasis, the diagonal elements of ρ_{eq} are the populations of the eigenstates at thermal equilibrium. There is no coherence for a sufficiently large quantum ensemble at thermal equilibrium.

We note that the density matrix at thermal equilibrium can be derived in a more strict manner by explicitly considering a system that includes both the canonical ensemble and the heat bath and by either tracing out the degrees of freedom of the heat bath [Sch06] or relying on a series expansion that reduces to only two terms in the limit of an infinite heat bath [Sch06; Swe12].

When approaching zero absolute temperature, the matrix element of ρ in the eigenbasis that corresponds to the lowest energy ϵ_i becomes much larger than all the others. At $T = 0$, the corresponding *ground state* is exclusively populated and the ensemble is in a pure state if there is just one state with this energy. For $T \rightarrow \infty$ on the other hand, differences between the diagonal matrix elements vanish and all states are equally populated. The ensemble is in a maximally mixed state.

6.2 Quantum and Classical Statistics

6.2.1 Types of Permutation Symmetry

Classical particles are either distinguishable or non-distinguishable, a difference that influences the relation between the system partition function and the molecular partition function (Section 5.2.1). Quantum particles are special. They are always indistinguishable, but there exist two types that behave differently when two particles are permuted. For *bosons*, the wavefunction is unchanged on such permutation, whereas for *fermions* the wavefunction changes sign. This sign change does not make the particles distinguishable, as absolute phase of the wavefunction does not correspond to an observable. However, it has important consequences for the population of microstates. Two (or more) bosons can occupy the same energy level. In the limit $T \rightarrow 0$ they will all occupy the ground state and form a *Bose-Einstein condensate*. Bosons are particles with integer spin, with the composite boson ^{4}He (two protons, two neutrons, two electrons) probably being the most famous example. In contrast, two fermions (particles with half-integer spin) cannot occupy the same state, a fact that is known as *Pauli exclusion principle*. Protons, neutrons, and electrons are fermions (spin 1/2), whereas photons are bosons (spin 1).

This difference in permutation symmetry influences the distribution of particles over energy levels. The simplest example is the distribution of two particles to two energy levels ϵ_{l} (for 'left') and ϵ_{r} (for 'right') [FR03]. For distinguishable classical particles four possible configurations exist:

1. ϵ_{l} is doubly occupied

2. ϵ_l is occupied by particle A and ϵ_r is occupied by particle B
3. ϵ_l is occupied by particle B and ϵ_r is occupied by particle A
4. ϵ_r is doubly occupied.

For bosons and for indistinguishable classical particles as well, the second and third configuration above cannot be distinguished. Only three configurations exist:

1. ϵ_l is doubly occupied
2. ϵ_l is occupied by one particle and ϵ_r is occupied by one particle
3. ϵ_r is doubly occupied.

For fermions, the first and third configuration of the boson case are excluded by the Pauli principle. Only one configuration is left:

1. ϵ_l is occupied by one particle and ϵ_r is occupied by one particle.

Since the number of configurations enters into all probability considerations, we shall find different probability distributions for systems composed of bosons, fermions, or distinguishable classical particles. The situation is most transparent for an ideal gas, i.e. N non-interacting point particles that have only translational degrees of freedom [WF12]. For such a system the spectrum of energy levels is continuous.

6.2.2 Bose-Einstein Statistics

We want to derive the probability distribution for the occupation of energy levels by bosons. To that end, we first pose the question how many configurations exist for distributing N_i particles to A_i energy levels in the interval between ϵ_i and $\epsilon_i + \mathrm{d}\epsilon$. Each level can be occupied by an arbitrary number of particles. We picture the problem as a common set of particles P_k $(k = 1 \dots N_i)$ and levels L_k $(k = 1 \dots A_i)$ that has $N_i + A_i$ elements. Now we consider all permutations in this set and use the convention that particles that stand left from a level are assigned to this level. For instance, the permutation $\{P_1, P_2, L_1, P_3, L_2, L_3\}$ for three particles and three levels denotes a state where level L_1 s occupied by particles P_1 and P_2, level L_2 is occupied by particle P_3 and level A_3 is empty. With this convention the last energy level is necessarily the last element of the set (any particle standing right from it would not have an associated level), hence only $(N_i + A_i - 1)!$ such permutations exist. Each permutation also encodes a sequence of particles, but the particles are indistinguishable. Thus we have to divide by $N_i!$ in order to not double count configurations that we cannot distinguish. It also does not matter in which sequence we order the levels with their associated subsets of particles. Without losing generality, we can thus consider only the sequence with increasing level energy, so that the level standing right (not included in the number of permutations $(N_i + A_i - 1)!$) is the level with the highest energy. For the remaining $A_i - 1$ lower levels we have counted $(A_i - 1)!$ permutations, but should have counted only the properly ordered one. Hence, we also have to divide by $(A_i - 1)!$. Therefore, the number of configurations and thus the number of microstates in the interval between ϵ_i and $\epsilon_i + \mathrm{d}\epsilon$ is

$$C_i = \frac{(N_i + A_i - 1)!}{N_i!\,(A_i - 1)!} \ . \tag{6.10}$$

The configurations in energy intervals with different indices i are independent of each other. Hence, the statistical weight of a macrostate is

$$\Omega = \prod_i \frac{(N_i + A_i - 1)!}{N_i!\,(A_i - 1)!} \tag{6.11}$$

As the number of energy levels is, in practice, infinite, we can choose the A_i sufficiently large for neglecting the 1 in $A_i - 1$. In an exceedingly good approximation we can thus write

$$\Omega = \prod_i \frac{(N_i + A_i)!}{N_i! A_i!} \; . \tag{6.12}$$

The next part of the derivation is the same as for the Boltzmann distribution in Section 4.3.1, i.e., it relies on maximization of $\ln \Omega$ using the Stirling formula and considering the constraints of conserved total particle number $N = \sum_i N_i$ and conserved total energy of the system [WF12]. The initial result is of the form

$$\frac{N_i}{A_i} = \frac{1}{Be^{-\beta\epsilon_i} - 1} \; , \tag{6.13}$$

where B is related to the Lagrange multiplier α by $B = e^{-\alpha}$ and thus to the chemical potential by $B = e^{-\mu/(k_\mathrm{B}T)}$. After a rather tedious derivation using the definitions of Boltzmann entropy and $(\partial u/\partial s)_V = T$ we can identify β with $-1/k_\mathrm{B}T$ [WF12]. We refrain from reproducing this derivation here, as the argument is circular: It uses the identification of k with k_B in the definition of Boltzmann entropy that we had made earlier on somewhat shaky grounds. We accept the identification of $|\beta|$ with $1/k_\mathrm{B}T$ as general for this type of derivations, so that we finally have

$$\frac{N_i}{A_i} = \frac{1}{Be^{\epsilon_i/k_\mathrm{B}T} - 1} \; . \tag{6.14}$$

Up to this point we have supposed nothing else than a continuous, or at least sufficiently dense, energy spectrum and identical bosons. To identify B we must have information on this energy spectrum and thus specify a concrete physical problem. When using the *density of states* for an ideal gas consisting of quantum particles with mass m in a box with volume V (see Section 7.2 for derivation),

$$D(\epsilon) = 4\sqrt{2}\pi \frac{V}{h^3} m^{3/2} \epsilon^{1/2} \; , \tag{6.15}$$

we find, for the special case $Be^{\epsilon_i/k_\mathrm{B}T} \gg 1$,

$$B = \frac{(2\pi m k_\mathrm{B} T)^{3/2}}{h^3} \cdot \frac{V}{N} \; . \tag{6.16}$$

6.2.3 Fermi-Dirac Statistics

The number N_i of fermions in an energy interval with A_i levels cannot exceed A_i. The number of allowed configurations is now given by the number of possibilities to select N_i out of A_i levels that are populated, whereas the remaining levels remain empty. As each level can exist in only one of two conditions, populated or empty, this is a binomial distribution problem as we have solved in Section 3.1.6. In Eq. (3.23) we need to substitute N by A_i and n by N_i. Hence, the number of allowed configurations in the energy interval between ϵ_i and $\epsilon_i + \Delta\epsilon_i$ is given by

$$C_i = \frac{A_i!}{N_i! \, (A_i - N_i)!} \tag{6.17}$$

and, considering mutual independence of the configurations in the individual energy intervals, the statistical weight of a macrostate for fermions is

$$\Omega = \prod_i \frac{A_i!}{N_i! \, (A_i - N_i)!} \; . \tag{6.18}$$

Again, the next step of the derivation is analogous to derivation of the Boltzmann distribution in Section 4.3.1 [WF12]. We find

$$\frac{N_i}{A_i} = \frac{1}{Be^{\epsilon_i/k_\mathrm{B}T}+1} \,. \tag{6.19}$$

For the special case $Be^{\epsilon_i/k_\mathrm{B}T} \gg 1$, B is again given by Eq. (6.16). Comparison of Eq. (6.19) with Eq. (6.14) reveals as the only difference the sign of the additional number 1 in the denominator on the right-hand side of the equations. In the regime $Be^{\epsilon_i/k_\mathrm{B}T} \gg 1$, for which we have specified B, this difference is negligible.

It is therefore of interest when this regime applies. As $\epsilon_i \geq 0$ in the ideal gas problem, we have $e^{\epsilon_i/k_\mathrm{B}T} \geq 1$, so that $B \gg 1$ is sufficient for the regime to apply. Wedler and Freund [WF12] have computed values of B according to Eq. (6.16) for the lightest ideal gas, H_2, and have found $B \gg 1$ for $p = 1$ bar down to $T = 20$ K and at ambient temperature for pressures up to $p = 100$ bar. For heavier molecules, B is larger under otherwise identical conditions. Whether a gas atom or molecule is a composite boson or fermion thus does not matter, except at very low temperatures and very high pressures. However, if conduction electrons in a metal, for instance in sodium, are considered as a gas, their much lower mass and higher number density N/V leads to $B \ll 1$ at ambient temperature and even at temperatures as high as 1000 K. Therefore, a gas model for conduction electrons (spin 1/2) must be set up with Fermi-Dirac statistics.

6.2.4 Maxwell-Boltzmann Statistics

In principle, atoms and molecules are quantum objects and not classical particles. This would suggest that the kinetic theory of gases developed by Maxwell before the advent of quantum mechanics is deficient. However, we have already seen that for particles as heavy as atoms and molecules and number densities as low as in gases at atmospheric pressure or a bit higher, the difference between Bose-Einstein and Fermi-Dirac statistics vanishes, unless temperature is very low. This suggests that, perhaps, classical Maxwell-Boltzmann statistics is indeed adequate for describing gases under common experimental conditions.

We assume distinguishable particles. Each of the N_i particles can be freely assigned to one of the A_i energy levels. All these configurations can be distinguished from each other, as we can picture each of the particles to have an individual tag. Therefore,

$$C_i = (A_i)^{N_i} \tag{6.20}$$

configurations can be distinguished in the energy interval between ϵ_i and $\epsilon_i + \Delta\epsilon_i$. Because the particles are distinguishable ('tagged'), the configurations in the individual intervals are generally not independent from each other, i.e. the total number of microstates does not factorize into the individual numbers of microstates in the intervals. We obtain *more* configurations than that because we have the additional choice of distributing the N 'tagged' particles to r intervals. We have already solved this problem in Section 4.3.1, the solution is Eq. (4.6). By considering the additional number of choices, which enters multiplicatively, we find for the statistical weight of a macrostate

$$\Omega = \frac{N!}{N_0!N_1!\ldots N_{r-1}!} \cdot A_0^{N_0} \cdot A_1^{N_1} \cdot \ldots A_{r-1}^{N_{r-1}} \tag{6.21}$$

$$= N! \prod_i \frac{A_i^{N_i}}{N_i!} \,. \tag{6.22}$$

It appears that we have assumed a countable number r of intervals, but as in the derivations for the Bose-Einstein and Fermi-Dirac statistics, nothing prevents us from making the intervals arbitrarily narrow and their number arbitrarily large.

Again, the next step in the derivation is analogous to derivation of the Boltzmann distribution in Section 4.3.1 [WF12]. All the different statistics differ only in the expressions for Ω, constrained maximization of $\ln \Omega$ uses the same Lagrange ansatz. We end up with

$$\frac{N_i}{A_i} = \frac{1}{Be^{\epsilon_i/k_\mathrm{B}T}} \,. \tag{6.23}$$

Comparison of Eq. (6.23) with Eqs. (6.14) and (6.19) reveals that, again, only the 1 in the denominator on the right-hand side makes the difference, now it is missing. In the regime, where Bose-Einstein and Fermi-Dirac statistics coincide to a good approximation, both of them also coincide with Maxwell-Boltzmann statistics.

There exist two caveats. First, we already know that the assumption of distinguishable particles leads to an artificial mixing entropy for two subsystems consisting of the same ideal gas or, in other words, to entropy not being extensive. This problem does not, however, influence the probability distribution, it only influences scaling of entropy with system size. We can solve it by an *ad hoc* correction when computing the system partition function from the molecular partition function. Second, to be consistent we should not use the previous expression for B, because it was derived under explicit consideration of quantization of momentum.[1] However, for Maxwell-Boltzmann statistics B can be eliminated easily. With $\sum_i N_i = N$ we have from Eq. (6.23)

$$N = \frac{1}{B} \sum_i A_i e^{-\epsilon_i/k_\mathrm{B}T} \,, \tag{6.24}$$

which gives

$$\frac{1}{B} = \frac{N}{\sum_i A_i e^{-\epsilon_i/k_\mathrm{B}T}} \,. \tag{6.25}$$

With this, we can express the distribution function as

$$P_i = \frac{N_i}{N} = \frac{A_i e^{-\epsilon_i/k_\mathrm{B}T}}{\sum_i A_i e^{-\epsilon_i/k_\mathrm{B}T}} \,. \tag{6.26}$$

Comparison of Eq. (6.26) with the Boltzmann distribution given by Eq. (4.25) reveals the factors A_i as the only difference. Thus, the probability distribution for Maxwell-Boltzmann statistics deviates from the most common form by the *degree of degeneracy* A_i of the individual levels. This degeneracy entered the derivation because we assumed that within the intervals between ϵ_i and $\epsilon_i + \Delta\epsilon_i$ several levels exist. If $\Delta\epsilon_i$ is finite, we speak of *near degeneracy*. For quantum systems, degeneracy of energy levels is a quite common phenomenon even in small systems where the energy spectrum is discrete. In order to describe such systems, the influence of degeneracy on the probability distribution must be taken into account.

Concept 6.2.1 — Degeneracy. In quantum systems with discrete energy levels there may exist g_i quantum states with the same energy ϵ_i that do not coincide in all their quantum numbers. This phenomenon is called *degeneracy* and g_i the *degree of degeneracy*. A set of g_i degenerate levels can be populated by up to g_i fermions. In the regime, where Boltzmann statistics is applicable to the quantum system, the probability distribution considering such degeneracy is given by

---

[1] This is more a matter of taste than of substance. As long as $Be^{\epsilon_i/k_\mathrm{N}T} \gg 1$, we can approximate any type of quantum statistics by Maxwell-Boltzmann statistics before solving for B. We are thus permitted to freely mix Maxwell-Boltzmann statistics with quantum-mechanical equations of motion.

$$P_i = \frac{N_i}{N} = \frac{g_i e^{-\epsilon_i/k_\mathrm{B}T}}{\sum_i g_i e^{-\epsilon_i/k_\mathrm{B}T}} \tag{6.27}$$

and the molecular partition function by

$$Z = \sum_i g_i e^{-\epsilon_i/k_\mathrm{B}T} \; . \tag{6.28}$$

The condition that degenerate levels do not coincide in all quantum numbers makes sure that the Pauli exclusion principle does not prevent their simultaneous population with fermions.

At this point we can summarize the expected number of particles with chemical potential μ at level i with energy ϵ_i and arbitrary degeneracy g_i for Bose-Einstein, Fermi-Dirac, and Boltzmann statistics:

$$N_i = \frac{g_i}{e^{(\epsilon_i-\mu)/(k_\mathrm{B}T)} - 1} \qquad \text{Bose}-\text{Einstein statistics} \tag{6.29}$$

$$N_i = \frac{g_i}{e^{(\epsilon_i-\mu)/(k_\mathrm{B}T)} + 1} \qquad \text{Fermi}-\text{Dirac statistics} \tag{6.30}$$

$$N_i = \frac{g_i}{e^{(\epsilon_i-\mu)/(k_\mathrm{B}T)}} \qquad \text{Boltzmann statistics} \; . \tag{6.31}$$

Note that the chemical potential μ in these equations is determined by the condition $N = \sum_i N_i$. The constant B in the derivations above is given by $B = e^{-\mu/(k_\mathrm{B}T)}$. If N is not constant, we have $\mu = 0$ and thus $B = 1$.

6.3 Simple Quantum Systems

6.3.1 Spin $S = 1/2$

The simplest quantum system is a two-level system and probably the best approximation to isolated two-level systems is found in magnetic resonance spectroscopy of dilute $S = 1/2$ spin systems. The Hamiltonian for an electron spin $S = 1/2$ in an external magnetic field along z is given by

$$\widehat{\mathcal{H}} = \gamma\hbar B_0 \widehat{S}_z \; , \tag{6.32}$$

where $\gamma = g\mu_\mathrm{B}/\hbar$ is the gyromagnetic ratio and B_0 is the magnetic field expressed in units of 1 Tesla. The two states are designated by magnetic spin quantum number $m_S = \mp 1/2$ and have energies $\epsilon_\mp = \mp\gamma\hbar B_0/2$. The partition function is

$$Z = e^{\gamma\hbar B_0/2k_\mathrm{B}T} + e^{-\gamma\hbar B_0/2k_\mathrm{B}T} \; , \tag{6.33}$$

and the expectation value of $\widehat{S}_z$, which is proportional to longitudinal magnetization, by

$$\left\langle \widehat{S}_z \right\rangle = \sum m_S P(m_S) \tag{6.34}$$

$$= \frac{(-1/2)\cdot e^{\gamma\hbar B_0/2k_\mathrm{B}T} + (1/2)e^{-\gamma\hbar B_0/2k_\mathrm{B}T}}{Z} \tag{6.35}$$

$$= -\frac{1}{2}\tanh\left(\gamma\hbar B_0/2k_\mathrm{B}T\right) \; . \tag{6.36}$$

Usually one has $\gamma\hbar B_0 \ll 2k_\mathrm{B}T$, which is called the high-temperature approximation. The series expansion of the hyperbolic tangent,

$$\tanh(x) = x - \frac{1}{3}x^3 + \frac{2}{15}x^5 + \dots \ , \tag{6.37}$$

can then be restricted to the leading term, which gives

$$\left\langle \widehat{S}_z \right\rangle = -\frac{\gamma\hbar B_0}{4k_\mathrm{B}T} \ . \tag{6.38}$$

6.3.2 Harmonic Oscillator

A diatomic molecule has one vibrational mode along the bond direction x. If we assign masses m_A and m_B to the two atoms and a force constant f to the bond, we can write the Hamiltonian as

$$\widehat{\mathcal{H}} = \frac{1}{2}f\widehat{x}^2 + \frac{\widehat{p}^2}{2\mu} \ , \tag{6.39}$$

where the reduced mass μ is

$$\mu = \frac{m_\mathrm{A}m_\mathrm{B}}{m_\mathrm{A}+m_\mathrm{B}} \tag{6.40}$$

and where the first term on the right-hand side of Eq. (6.39) corresponds to potential energy and the second term to kinetic energy.

Eq. (6.39) can be cast in the form

$$\widehat{\mathcal{H}} = \frac{1}{2}\mu\omega^2\left(R - R_\mathrm{E}\right)^2 - \frac{\hbar^2}{2\mu}\frac{\mathrm{d}^2}{\mathrm{d}R^2} \ , \tag{6.41}$$

where we have substituted $\widehat{x}$ by the deviation of the atom-atom distance R from the bond length R_E and introduced the angular oscillation frequency ω of a classical oscillator with

$$\omega = \sqrt{\frac{f}{\mu}} \ . \tag{6.42}$$

Eq. (6.41) produces an infinite number of eigenstates with energies

$$\epsilon_v = \hbar\omega\left(v + \frac{1}{2}\right) \ , \tag{6.43}$$

where $v = 0, 1, \dots, \infty$ is the *vibrational quantum number*. All energies are positive, even the one of the ground state with $v = 0$. This residual zero-point vibration can be considered as a consequence of Heisenberg's uncertainty principle, since for a non-oscillating diatomic molecule atom coordinates as well as momentum would be sharply defined, which would violate that principle. In the context of statistical thermodynamics the unfortunate consequence is that for an ensemble of N diatomic molecules for $T \rightarrow 0$ the vibrational contribution to the internal energy u approaches $u_0 = N\hbar\omega/2$ and thus the term u/T in the entropy expression (5.11) approaches infinity. We ignore this problem for the moment.

The partition function of the harmonic oscillator is an infinite series,

$$Z = \sum_{v=0}^{\infty} e^{-\hbar\omega(v+1/2)/k_\mathrm{B}T} \tag{6.44}$$

$$= e^{-\hbar\omega/2k_\mathrm{B}T} \sum_{v=0}^{\infty} e^{-\hbar\omega v/k_\mathrm{B}T} \tag{6.45}$$

$$= e^{-\hbar\omega/2k_\mathrm{B}T} \sum_{v=0}^{\infty} \left(e^{-\hbar\omega/k_\mathrm{B}T}\right)^v \tag{6.46}$$

$$= e^{-\hbar\omega/2k_\mathrm{B}T} \sum_{n=0}^{\infty} x^n \,. \tag{6.47}$$

where we have substituted $x = e^{-\hbar\omega/k_\mathrm{B}T}$ and $n = v$ to obtain the last line. Since for finite temperatures $0 < e^{-\hbar\omega/k_\mathrm{B}T} < 1$, the infinite series $\sum_{n=0}^{\infty} x^n$ converges to $1/(1-x)$. Hence,

$$Z = \frac{e^{-\hbar\omega/2k_\mathrm{B}T}}{1 - e^{-\hbar\omega/k_\mathrm{B}T}} \,. \tag{6.48}$$

We can again discuss the behavior for $T \rightarrow 0$. In the denominator, the argument of the exponential function approaches $-\infty$, so that the denominator approaches unity. In the numerator the argument of the exponential function also approaches $-\infty$, so that the partition function approaches zero and Helmholtz free energy $f = -k_\mathrm{B}T \ln Z$ can only be computed as a limiting value. The term $k_\mathrm{B} \ln Z$ in the entropy equation (5.11) approaches $-\infty$.

This problem can be healed by shifting the energy scale by $\Delta E = -\hbar\omega/2$. We then have[2]

$$Z = \frac{1}{1 - e^{-\hbar\omega/k_\mathrm{B}T}} \,. \tag{6.49}$$

With this shift, the partition function and the population of the ground state $v = 0$ both approach 1 when the temperature approaches zero. For the term u/T in the entropy expression we still need to consider a limiting value, but it can be shown that $\lim\limits_{T \rightarrow 0} u/T = 0$. Since $k_\mathrm{B} \ln Z = 0$ for $Z = 1$, entropy of an ensemble of harmonic oscillators vanishes at the zero point in agreement with Nernst's theorem. Helmholtz free energy $f = -k_\mathrm{B}T \ln Z$ approaches zero.

R For computing a Boltzmann distribution we can shift all energy levels by the same offset ΔE without influencing the $P(i)$, as such a shift leads to a multiplication by the same factor of the numerator and of all terms contributing to the partition function. Such a shift can remove an infinity of the partition function.

This partition function can also be expressed with a *characteristic vibrational temperature*

$$\Theta_\mathrm{vib} = \frac{\hbar\omega}{k_\mathrm{B}} \,. \tag{6.50}$$

This temperature is usually higher than room temperature. We have

$$Z_\mathrm{vib} = \frac{1}{1 - e^{-\Theta_\mathrm{vib}/T}} \,. \tag{6.51}$$

Thus, $Z \approx 1$ at room temperature, which implies that only the vibrational ground state is significantly populated. Vibration does not significantly contribute to entropy at room temperature.

---

[2] The shift does not influence the denominator, as it merely removes the first factor on the right-hand side of Eq. (6.45).

6.3.3 Einstein and Debye Models of a Crystal

The considerations on the harmonic oscillator can be extended to a simple model for vibrations in a crystal. If we assume that all atoms except one are fixed at their average locations, the potential at the unique atom is parabolic. This assumption made by Einstein may at first sight violate his own dictum that "Everything should be made as simple as possible, but not simpler.". We shall come back to this point below. For the moment we consider Einstein's approach as a very simple *mean field approach*. Instead of the one-dimensional harmonic oscillator treated in Section 6.3.2, we now have a three-dimensional harmonic oscillator. For sufficiently high point symmetry at the unique atom, we can assume an isotropic force constant f. Each atom is then described by three independent harmonic oscillators along three orthogonal directions. The harmonic oscillators of different atoms are also independent by construction. Because we want to compute an absolute internal energy we revert to the partition function of the harmonic oscillator without energy shift given in Eq. (6.48). The partition function for a crystal with N atoms, considering that the atoms in a crystal lattice are distinguishable and that thus Eq. (5.17) applies, is then given by

$$Z = \left(\frac{e^{-\hbar\omega/2k_\mathrm{B}T}}{1 - e^{-\hbar\omega/k_\mathrm{B}T}} \right)^{3N} . \tag{6.52}$$

Internal energy can be computed by Eq. (4.34) [WF12],

$$\begin{aligned} u_\mathrm{vib} &= k_\mathrm{B}T^2 \left[\frac{\partial}{\partial T} \ln \left\{ \left(\frac{e^{-\hbar\omega/2k_\mathrm{B}T}}{1 - e^{-\hbar\omega/k_\mathrm{B}T}} \right)^{3N} \right\} \right]_V & (6.53) \\ &= 3k_\mathrm{B}NT^2 \left[\frac{\partial}{\partial T} \left\{ -\frac{\hbar\omega}{2k_\mathrm{B}T} - \ln\left(1 - e^{-\hbar\omega/k_\mathrm{B}T}\right) \right\} \right]_V & (6.54) \\ &= 3k_\mathrm{B}NT^2 \left[\frac{\hbar\omega}{2k_\mathrm{B}T^2} + \frac{\hbar\omega/k_\mathrm{B}T^2 \cdot e^{-\hbar\omega/k_\mathrm{B}T}}{1 - e^{-\hbar\omega/k_\mathrm{B}T}} \right] & (6.55) \\ &= \frac{3}{2}N\hbar\omega + \frac{3N\hbar\omega}{e^{\hbar\omega/k_\mathrm{B}T} - 1} . & (6.56) \end{aligned}$$

With the characteristic vibrational temperature Θ_vib introduced in Eq. (6.50) and by setting $N = N_\mathrm{Av}$ to obtain a molar quantity, we find

$$U_\mathrm{vib} = \frac{3}{2}R\Theta_\mathrm{vib} + \frac{3R\Theta_\mathrm{vib}}{e^{\Theta_\mathrm{vib}/T} - 1} . \tag{6.57}$$

The molar heat capacity of an Einstein solid is the derivative of U_vib with respect to T. We note that we do not need to specify constant volume or constant pressure, since this simple model depends on neither of these quantities. We find

$$C_\mathrm{vib} = 3R\frac{\left(\Theta_\mathrm{vib}/T\right)^2 e^{\Theta_\mathrm{vib}/T}}{\left(e^{\Theta_\mathrm{vib}/T} - 1\right)^2} . \tag{6.58}$$

According to the rule of Dulong and Petit we should obtain the value $3R$ for $T \to \infty$. Since the expression becomes indeterminate $(0/0)$, we need to compute a limiting value, which is possible with the approach of de l'Hospital where we separately differentiate the numerator and denominator[WF12]. The derivation is lengthy but it indeed yields the limiting value $3R$:

$$\begin{aligned}
\lim_{T\to\infty} C_\mathrm{vib} &= \lim_{T\to\infty} 3R \frac{\left(\Theta_\mathrm{vib}/T\right)^2 e^{\Theta_\mathrm{vib}/T}}{\left(e^{\Theta_\mathrm{vib}/T}-1\right)^2} && (6.59)\\
&= 3R \lim_{T\to\infty} \frac{2\left(\Theta_\mathrm{vib}/T\right)\left(-\Theta_\mathrm{vib}/T^2\right)}{2\left(1-e^{-\Theta_\mathrm{vib}/T}\right)\left(e^{\Theta_\mathrm{vib}/T}\right)\left(-\Theta_\mathrm{vib}/T^2\right)} && (6.60)\\
&= 3R \lim_{T\to\infty} \frac{\left(\Theta_\mathrm{vib}/T\right)}{\left(1-e^{-\Theta_\mathrm{vib}/T}\right)} && (6.61)\\
&= 3R \lim_{T\to\infty} \frac{\left(-\Theta_\mathrm{vib}/T^2\right)}{-e^{-\Theta_\mathrm{vib}/T}\left(-\Theta_\mathrm{vib}/T^2\right)} && (6.62)\\
&= 3R\,. && (6.63)
\end{aligned}$$

In Eq. (6.59) in the numerator and in going from Eq. (6.60) to (6.61) we have set $e^{\Theta_\mathrm{vib}/T}$ to 1, as we may for $T \to \infty$. As the expression was still indeterminate, we have computed the derivatives of numerator and denominator once again in going from Eq. (6.61) to (6.62) and finally we have once more set $e^{-\Theta_\mathrm{vib}/T}$ to 1 in going from Eq. (6.62) to (6.63). We see that Einstein's very simple model agrees with the rule of Dulong and Petit.

The model of the Einstein solid differs from a model of N_Av one-dimensional harmonic oscillators according to Section 6.3.2 only by a power of 3 in the partition function, which, after computing the logarithm, becomes a factor of 3 in the temperature-dependent term of U_vib and thus in C_vib. Hence, in the high-temperature limit the vibrational contribution to the molar heat capacity of a gas consisting of diatomic molecules is equal to R. It follows that, in this limit, each molecule contributes an energy $k_\mathrm{B}T$ to the internal energy, i.e. each of the two degrees of freedom (potential and kinetic energy of the vibration) that are quadratic in the coordinates contributes a term $k_\mathrm{B}T/2$. This agrees with the equipartition theorem. Likewise, the Einstein solid agrees with this theorem.

From experiments it is known that molar heat capacity approaches zero when temperature approaches zero. Again the limiting value can be computed by the approach of de 'l Hospital [WF12], where this time we can neglect the 1 in $e^{\Theta_\mathrm{vib}/T}-1$, as $e^{\Theta_\mathrm{vib}/T}$ tends to infinity for $T \to 0$. In the last step we obtain

$$\lim_{T\to 0} C_\mathrm{vib} = 6R \lim_{T\to 0} \frac{1}{e^{\Theta_\mathrm{vib}/T}} = 0\,. \tag{6.64}$$

Thus, the Einstein solid also agrees with the limiting behavior of heat capacity at very low temperatures.

Nevertheless the model is 'too simple', and Einstein was well aware of that. Vibrations of the individual atoms are not independent, but rather collective. The lattice vibrations, called phonons, have a spectrum whose computation is outside the scope of the Einstein model. A model that can describe this spectrum has been developed by Debye based on the density of states of frequencies ν. This density of states in turn has been derived by Rayleigh and Jeans based on the idea that the phonons are a system of standing waves in the solid. It is given by [WF12]

$$D(\nu) = \frac{4\pi\nu^2}{c^3} V\,. \tag{6.65}$$

Debye replaced c by a mean velocity of wave propagation in the solid, considered one longitudinal and two transverse waves and only the $3N$ states with the lowest frequencies, as the solid has only $3N$ vibrational degrees of freedom. These considerations lead to a maximum phonon frequency ν_max and, after resubstitution of the mean velocity, to a frequency spectrum that is still proportional to ν^2 and scales with ν_max^{-3}. Instead of the characteristic vibration temperature, it is now convenient to define the Debye temperature

$$\Theta_\mathrm{D} = \frac{h\nu_\mathrm{max}}{k_\mathrm{B}} . \tag{6.66}$$

In this model the molar heat capacity of the solid becomes

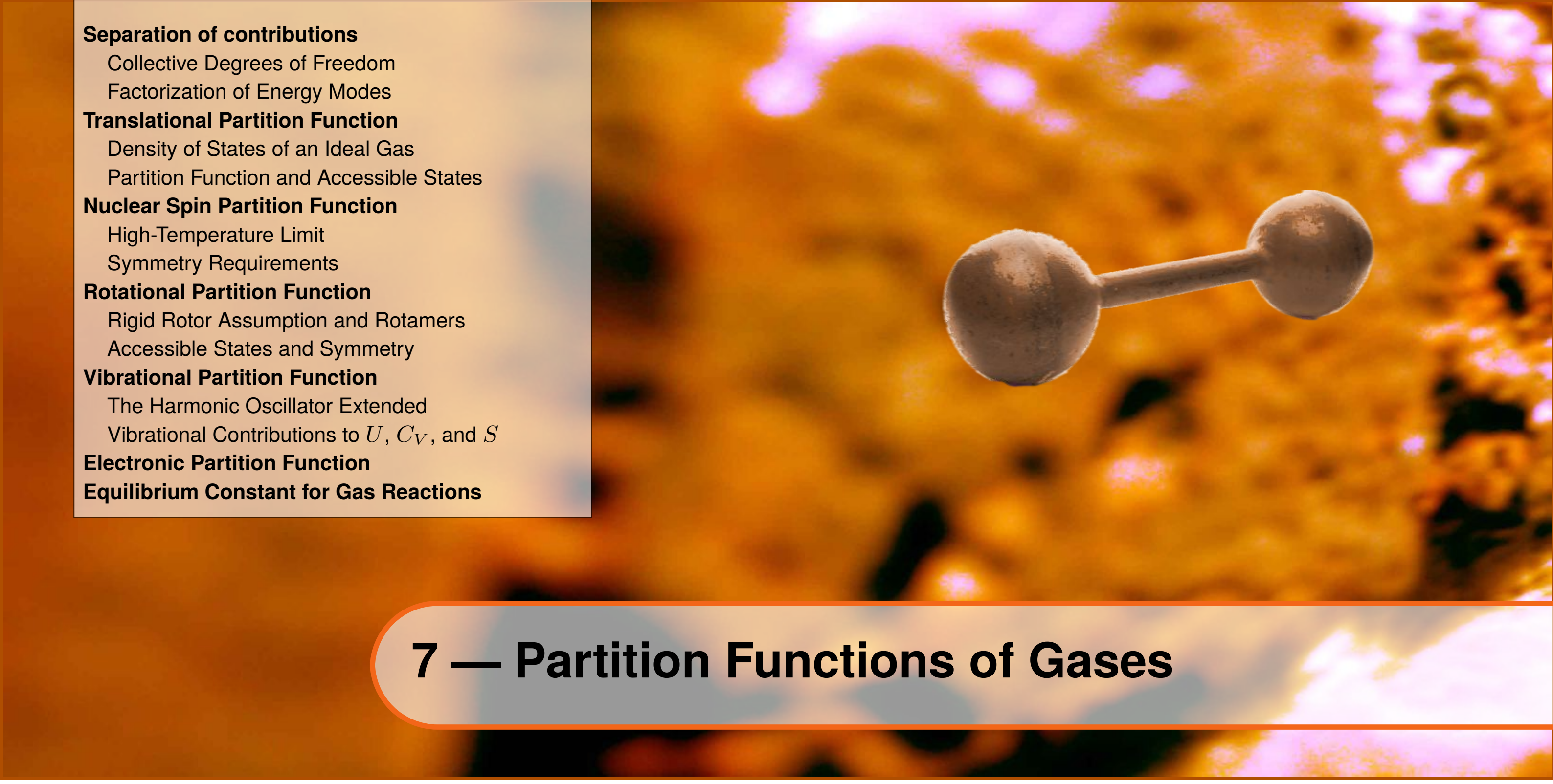
$$C_\mathrm{vib} = 9R\left(\frac{T}{\Theta_\mathrm{D}}\right)^3 \int_0^{\Theta_\mathrm{D}/T} \frac{x^4 e^x}{\left(e^x - 1\right)^2}\mathrm{d}x \tag{6.67}$$

The integral can be evaluated numerically after series expansion and finally Debye's T^3 law,

$$\lim_{T\to 0} C_\mathrm{vib} = 233.8R\frac{T^3}{\Theta_\mathrm{D}^3} , \tag{6.68}$$

results. This law does not only correctly describe that the heat capacity vanishes at absolute zero, it also correctly reproduces the scaling law, i.e., the T^3 dependence that is found experimentally. The high-temperature limit can also be obtained by series expansion and is again Dulong-Petit's value of $3R$.

The Debye model is still an approximation. Phonon spectra of crystalline solids are not featureless. They are approximated, but not fully reproduced, by a ν^2 dependence. The deviations from the Debye model depend on the specific crystal structure.



Separation of contributions
    Collective Degrees of Freedom
    Factorization of Energy Modes
Translational Partition Function
    Density of States of an Ideal Gas
    Partition Function and Accessible States
Nuclear Spin Partition Function
    High-Temperature Limit
    Symmetry Requirements
Rotational Partition Function
    Rigid Rotor Assumption and Rotamers
    Accessible States and Symmetry
Vibrational Partition Function
    The Harmonic Oscillator Extended
    Vibrational Contributions to U, C_V, and S
Electronic Partition Function
Equilibrium Constant for Gas Reactions

7 — Partition Functions of Gases

7.1 Separation of contributions

7.1.1 Collective Degrees of Freedom

In Section 6.3.3 we have seen that the treatment of condensed phases can be complicated by collective motion of particles. Such effects are absent in an ideal gas that consists of point particles, a model that is reasonable for noble gases far from condensation. For gases consisting of molecules, it does not suffice to consider only translational motion as in Maxwell's kinetic theory of gases. We see this already when considering H_2 gas, where each molecule can be approximated by a harmonic oscillator (Section 6.3.2). Neglect of the vibrational degrees of freedom will lead to wrong results for internal energy, heat capacity, the partition function, and entropy, at least at high temperatures. In fact, an H_2 molecule is not only an oscillator, it is also a rotor. As a linear molecule it has two rotational degrees of freedom, which also contribute to internal energy and to the partition function.

In principle, we could try to ignore all this and treat each atom as one particle. If the Hamiltonian includes the potentials that characterize interaction between the particles, our equations of motion would be correct. In practice, such a treatment is inconvenient and it is better to group the spatial degrees of freedom according to the type of motion. The H_2 molecule has 3 translational degrees of freedom, 2 rotational degrees of freedom, and 1 vibrational degree of freedom in the collective motion picture. The sum is 6, as expected for two atoms with each of them having 3 translational degrees of freedom in an 'independent' motion picture. In general, a molecule with n atoms has $f_s = 3n$ spatial degrees of freedom, 3 of which are translational, 3 are rotational, except for linear molecules, which have only 2 rotation degrees of freedom, and the rest are vibrational. We note that the number of degrees of freedom in phase space is $f = 2f_s$ because each spatial degree of freedom is also assigned a momentum degree of freedom.

These considerations take care of particle motion. Further contributions to internal energy and to the partition function can arise from spin. In both closed-shell and open-shell molecules, nuclear spin can play a role. This is indeed the case for H_2, which can exist in *ortho* and *para* states that differ in correlation of the nuclear spins of the two hydrogen atoms. For open-shell molecules electron spin degrees of freedom must be considered. This is the case, for instance, for O_2, which has a triplet ground state. In this case, rotational and spin degrees of freedom correspond to similar energies and couple. Finally, at sufficiently high temperatures electronic excitation becomes possible and then also makes a contribution to the partition function.

In many cases, the individual contributions are separable, i.e. the modes corresponding to different types of motions can be treated independently. Roughly speaking, this results from a separation of energy ranges (frequency bands) of the modes and a corresponding separation of time scales. Nuclear spin degrees of freedom have much lower energy than rotational degrees of freedom which usually have much lower energy than vibrational degrees of freedom which have much lower energies than electronic excitation. The independence of nuclear and electron motion is basis of the *Born-Oppenheimer approximation* and the independence of rotational and vibrational motion is invoked when treating a molecule as a rigid rotor. Separability of energy modes leads to a sum rule for the energy contributions for a single closed-shell molecule [Mac98],

$$\epsilon_j = \epsilon_{j,\mathrm{trs}} + \epsilon_{j,\mathrm{ns}} + \epsilon_{j,\mathrm{rot}} + \epsilon_{j,\mathrm{vib}} + \epsilon_{j,\mathrm{el}} \ , \tag{7.1}$$

where $\epsilon_{j,\mathrm{trs}}$, $\epsilon_{j,\mathrm{ns}}$, $\epsilon_{j,\mathrm{rot}}$, $\epsilon_{j,\mathrm{vib}}$, and $\epsilon_{j,\mathrm{el}}$ are the translational, nuclear spin, rotational, vibrational, and electronic contributions, respectively. For a monoatomic molecule (atom) $\epsilon_{j,\mathrm{rot}}$ and $\epsilon_{j,\mathrm{vib}}$ vanish. If both the number of neutrons and of protons in the nucleus is even, the nucleus has spin $I = 0$. In that case the nuclear spin contribution vanishes for an atom, even in the presence of an external magnetic field. If all nuclei have spin zero, the nuclear spin contribution also vanishes for a diatomic or multi-atomic molecule.

7.1.2 Factorization of Energy Modes

If we assume the equipartition theorem to hold, or even more generally, the whole system to attain thermal equilibrium, there must be some coupling between the different modes. If we say that the energy modes are separable, we assume *weak coupling*, which means that for statistical purposes we can assume the modes to be independent of each other. The consequence for the computation of the partition function can be seen by considering a system of N particles with an α mode associated with quantum number k and an ω mode associated with quantum number r [Mac98]. The total energy of a single molecule of this type is $\epsilon_j = \epsilon_{j,\alpha k} + \epsilon_{j,\omega r}$ The molecular partition function is given by

$$Z = \sum_k \sum_r e^{-\beta(\epsilon_{\alpha k} + \epsilon_{\omega r})} \ . \tag{7.2}$$

This sum can be rewritten as

$$\begin{aligned} Z &= \sum_k \sum_r e^{-\beta\epsilon_{\alpha k}} \cdot e^{-\beta\epsilon_{\omega r}} && (7.3)\\ &= \sum_k e^{-\beta\epsilon_{\alpha k}} \sum_r e^{-\beta\epsilon_{\omega r}} && (7.4)\\ &= \sum_k e^{-\beta\epsilon_{\alpha k}} Z_\omega = Z_\omega \sum_k e^{-\beta\epsilon_{\alpha k}} && (7.5)\\ &= Z_\alpha Z_\omega \ . && (7.6) \end{aligned}$$

We see that the total partition function is the product of the partition functions corresponding to the individual modes. This consideration can be extended to multiple modes. With Eq. (7.1) it follows that

$$Z = Z_{\mathrm{trs}} \cdot Z_{\mathrm{ns}} \cdot Z_{\mathrm{rot}} \cdot Z_{\mathrm{vib}} \cdot Z_{\mathrm{el}} \ . \tag{7.7}$$

By considering Eq. (5.17) or Eq. (5.18) we see that we can also compute the partition function for a given mode for all N particles before multiplying the modes. We have already seen that we must set $z_{\mathrm{trs}} = Z_{\mathrm{trs}}^N/N!$ to heal the Gibbs paradox. What about the other, internal degrees of freedom? If two particles with different internal states are exchanged, they must be considered distinguishable, exactly because their internal state 'tags' them. Hence, for all the other modes we have $z_\alpha = Z_\alpha^N$. Thus,

$$
\begin{aligned}
z &= \frac{1}{N!}\left(Z_{\mathrm{trs}} \cdot Z_{\mathrm{ns}} \cdot Z_{\mathrm{rot}} \cdot Z_{\mathrm{vib}} \cdot Z_{\mathrm{el}}\right)^N && (7.8)\\
&= \frac{Z_{\mathrm{trs}}^N}{N!} \cdot Z_{\mathrm{ns}}^N \cdot Z_{\mathrm{rot}}^N \cdot Z_{\mathrm{vib}}^N \cdot Z_{\mathrm{el}}^N \; . && (7.9)
\end{aligned}
$$

Accordingly, we can consider each of the partition functions in turn. We also note that separability of the energies implies factorization of the molecular wavefunction,

$$\psi = \psi_{\mathrm{trs}} \cdot \psi_{\mathrm{ns}} \cdot \psi_{\mathrm{rot}} \cdot \psi_{\mathrm{vib}} \cdot \psi_{\mathrm{el}} \tag{7.10}$$

7.2 Translational Partition Function

7.2.1 Density of States of an Ideal Gas

First, we derive the density of states that we had already used in computing the distribution functions for quantum gases. We consider a quantum particle in a three-dimensional cubic box with edge length a [WF12]. The energy is quantized with integer quantum numbers n_x, n_y, and n_z corresponding to the three pairwise orthogonal directions that span the cube,

$$
\begin{aligned}
\epsilon_{\mathrm{trs}} &= \frac{h^2}{8ma^2}\left(n_x^2 + n_y^2 + n_z^2\right) && (7.11)\\
&= \frac{1}{2m}\left(p_x^2 + p_y^2 + p_z^2\right) \; . && (7.12)
\end{aligned}
$$

It follows that momentum is also quantized with $|p_i| = (h/2a)n_i$ $(i = x, y, z)$. It is convenient to consider momentum in a Cartesian frame where $h/2a$ is the unit along the x, y, and z axes. Each state characterized by a unique set of *translational quantum numbers* (n_x, n_y, n_z) 'owns' a small cube with volume $h^3/8a^3$ in the octant with $x \geq 0$, $y \geq 0$, and $z \geq 0$. Since momentum can also be negative, we need to consider all eight octants, so that each state owns a cell in momentum space with volume h^3/a^3. In order to go to phase space, we need to add the spatial coordinates. The particle can move throughout the whole cube with volume a^3. Hence, each state owns a phase space volume of h^3.

By rearranging Eq. (7.11) we can obtain an equation that must be fulfilled by the quantum numbers,

$$\frac{n_x^2}{\left(\frac{a}{h}\sqrt{8m\epsilon}\right)^2} + \frac{n_y^2}{\left(\frac{a}{h}\sqrt{8m\epsilon}\right)^2} + \frac{n_z^2}{\left(\frac{a}{h}\sqrt{8m\epsilon}\right)^2} = 1 \tag{7.13}$$

and by using Eq. (7.12) we can convert it to an equation that must be fulfilled by the components of the momentum vector,

$$\frac{p_x^2}{\left(\frac{1}{2}\sqrt{8m\epsilon}\right)^2} + \frac{p_y^2}{\left(\frac{1}{2}\sqrt{8m\epsilon}\right)^2} + \frac{p_z^2}{\left(\frac{1}{2}\sqrt{8m\epsilon}\right)^2} = 1 \; . \tag{7.14}$$

All states with quantum numbers that make the expression on the left-hand side of Eq. (7.13) or Eq. (7.14) smaller than 1 correspond to energies that are smaller than ϵ. The momentum associated with these states lies in the sphere defined by Eq. (7.14) with radius $\frac{1}{2}\sqrt{8m\epsilon}$ and volume $\frac{\pi}{6}(8m\epsilon)^{3/2}$. With cell size h^3/a^3 in momentum space the number of cells with energies smaller than ϵ is

$$\mathcal{N}(\epsilon) = \frac{8\sqrt{2}}{3}\pi\frac{V}{h^3}(m\epsilon)^{3/2} \,, \tag{7.15}$$

where we have substituted a^3 by volume V of the box. The number of states in an energy interval between ϵ and $\epsilon + \mathrm{d}\epsilon$ is the first derivative of $\mathcal{N}(\epsilon)$ with respect to ϵ and is the sought density of states,

$$D(\epsilon) = 4\sqrt{2}\pi\frac{V}{h^3}m^{3/2}\epsilon^{1/2} \,. \tag{7.16}$$

7.2.2 Partition Function and Accessible States

This density of states is very high, so that we can replace the sum over the quantum numbers n_i in the partition function of the canonical ensemble by an integral [Mac98],

$$Z_{\mathrm{trs},i} = \int_o^\infty e^{-\beta n_i^2 h^2/8ma^2}\mathrm{d}n_i \;\; (i = x, y, z) \tag{7.17}$$

$$= \sqrt{\frac{2\pi m}{\beta}}\frac{a}{h} \,. \tag{7.18}$$

The contributions along orthogonal spatial coordinates are also independent of each other and factorize. Hence,

$$Z_{\mathrm{trs}} = Z_{\mathrm{trs},x} \cdot Z_{\mathrm{trs},y} \cdot Z_{\mathrm{trs},z} = \left(\frac{2\pi m k_{\mathrm{B}} T}{h^2}\right)^{3/2} V \,, \tag{7.19}$$

where we have again substituted a^3 by V and, as by now usual, also β by $1/k_{\mathrm{B}}T$. The corresponding molar partition function is

$$z_{\mathrm{trs}} = \frac{1}{N_{\mathrm{Av}}!}\left[\left(\frac{2\pi m k_{\mathrm{B}} T}{h^2}\right)^{3/2} V\right]^{N_{\mathrm{Av}}} \,. \tag{7.20}$$

At this point it is useful to introduce another concept:

Concept 7.2.1 — Number of accessible states. The molecular canonical partition function Z is a measure for the number of states that are accessible to the molecule at a given temperature.

This can be easily seen when considering

$$P_i = \frac{N_i}{N} = \frac{g_i e^{-\epsilon_i/k_{\mathrm{B}}T}}{Z} \tag{7.21}$$

and $\sum_i P_i = 1$. If we consider a mole of ^{4}He (bosons) at 4.2 K, where it liquifies, we find that $z_{\mathrm{trs}}/N_{\mathrm{Av}} \approx 7.5$, which is not a large number [Mac98]. This indicates that we are close to breakdown of the regime where Bose-Einstein statistics can be approximated by Boltzmann statistics.

R For $T \to 0$ only the g_0 lowest energy states are populated. In the absence of ground-state degeneracy, $g_0 = 1$, we find $Z = 1$ and with an energy scale where $U(T = 0) = 0$ we have $S(0) = 0$ in agreement with Nernst's theorem.

An expression for the translational contribution to the entropy of an ideal gas can be derived from Eq. (7.19), Eq. (5.18), and Eq. (5.11). We know that $u = 3Nk_\mathrm{B}T/2$, so that we only need to compute $\ln z_\mathrm{trs}$,

$$\begin{aligned}
\ln z_\mathrm{trs} &= \ln \frac{1}{N!} Z_\mathrm{trs}^N && (7.22)\\
&= -\ln N! + N \ln Z_\mathrm{trs} && (7.23)\\
&= -N \ln N + N + N \ln Z_\mathrm{trs} && (7.24)\\
&= N\left(1 + \ln \frac{Z_\mathrm{trs}}{N}\right) , && (7.25)
\end{aligned}$$

where we have used Stirling's formula to resolve the factorial. Thus we find

$$\begin{aligned}
s &= \frac{u}{T} + k_\mathrm{B} \ln z && (7.26)\\
&= \frac{3}{2} N k_\mathrm{B} + k_\mathrm{B} N \left(1 + \ln \frac{Z_\mathrm{trs}}{N}\right) && (7.27)\\
&= N k_\mathrm{B} \left(\frac{5}{2} + \ln \frac{Z_\mathrm{trs}}{N}\right) && (7.28)
\end{aligned}$$

By using Eq. (7.19) we finally obtain the *Sackur-Tetrode equation*

$$s = N k_\mathrm{B} \left\{ \frac{5}{2} + \ln \left[\left(\frac{2\pi m k_\mathrm{B} T}{h^2} \right)^{\frac{3}{2}} \frac{V}{N} \right] \right\} . \qquad (7.29)$$

To obtain the molar entropy S_m, N has to be replaced by N_Av. Volume can be substituted by pressure and temperature, by noting that the molar volume is given by $V_\mathrm{m} = RT/p = N_\mathrm{Av}V/N$. With $N_\mathrm{Av}k_\mathrm{B} = R$ and the molar mass $M = N_\mathrm{Av}m$ we obtain

$$S_\mathrm{m} = R \left\{ \frac{5}{2} + \ln \left[\left(\frac{2\pi M k_\mathrm{B} T}{N_\mathrm{Av} h^2} \right)^{\frac{3}{2}} \frac{RT}{N_\mathrm{Av} p} \right] \right\} \qquad (7.30)$$

7.3 Nuclear Spin Partition Function

7.3.1 High-Temperature Limit

In the absence of a magnetic field, all nuclear spin states are degenerate,[1] except for the very tiny splittings that arise from J couplings between the nuclear spins themselves. Even if we consider the largest magnetic fields available, it is safe to assume that all nuclear spin states are equally populated down to temperatures of at least 1.5 K and that the contribution of nuclear spins to the total energy is negligibly small. Of course, NMR spectroscopy relies on the fact that these states are *not exactly* equally populated, but in the context of statistical thermodynamics, the contribution to internal energy and the population differences are negligible.

Hence, in this high-temperature limit all nuclear spin states are fully accessible and the number of accessible states equals the total number of nuclear spin states. This gives

$$Z_\mathrm{ns} = \prod_i (2I_i + 1) , \qquad (7.31)$$

where the I_i are the nuclear spin quantum numbers for nuclei in the molecule. Magnetic equivalence leads to degeneracy of nuclear spin levels, but does not influence the total number of nuclear spin states. Since the term u/T in Eq. (5.11) is negligible and $z_\mathrm{ns} = Z_\mathrm{ns}^N$, we have

$$s = N k_\mathrm{B} \sum_i \ln (2I_i + 1) . \qquad (7.32)$$

---

[1] We neglect nuclear quadrupole coupling, which averages in gases.

This contribution to entropy is not generally negligible. Still it is generally ignored in textbooks, which usually does not cause problems, as the contribution is constant under most conditions where experiments are conducted and does not change during chemical reactions.

7.3.2 Symmetry Requirements

Nuclear spin states have another, more subtle effect that may prevent separation of state spaces. We consider this effect for H_2. In this molecule, the electron wavefunction arises from two electrons, which are fermions, and must thus be antisymmetric with respect to exchange of the two electrons. In quantum-chemical computations this is ensured by using a Slater determinant. Likewise, the nuclear wavefunction must be antisymmetric with respect to exchange of the two protons, which are also fermions. The spin part is antisymmetric for the singlet state with total nuclear spin quantum number $F = 0$,

$$\psi_{\mathrm{ns,S}} = \frac{1}{\sqrt{2}}\left(|\alpha\beta\rangle - |\beta\alpha\rangle\right) \ , \tag{7.33}$$

and symmetric for the triplet state with $F = 1$, as can be seen by the wavefunctions of each of the three triplet substates:

$$\psi_{\mathrm{ns,T_+}} = |\alpha\alpha\rangle \tag{7.34}$$

$$\psi_{\mathrm{ns,T_0}} = \frac{1}{\sqrt{2}}\left(|\alpha\beta\rangle + |\beta\alpha\rangle\right) \tag{7.35}$$

$$\psi_{\mathrm{ns,T_-}} = |\beta\beta\rangle \ . \tag{7.36}$$

The translational, vibrational, and electron wavefunction are generally symmetric with respect to the exchange of the two nuclei. The rotational wavefunction is symmetric for even rotational quantum numbers J and antisymmetric for odd quantum numbers. Hence, to ensure that the generalized Pauli principle holds and the total wavefunction is antisymmetric with respect to exchange of indistinguishable nuclei, even J can only be combined with the antisymmetric nuclear spin singlet state and odd J only with the symmetric triplet state. The partition functions for these two cases must be considered separately. For H_2 we have

$$Z_{\mathrm{para}} = \sum_{J\mathrm{even}} (2J+1)\, e^{-J(J+1)\hbar^2/2Ik_\mathrm{B}T} \ , \tag{7.37}$$

where $g_J = 2J + 1$ is the degeneracy of the rotational states and I is the moment of inertia, and

$$Z_{\mathrm{ortho}} = 3\sum_{J\mathrm{odd}} (2J+1)\, e^{-J(J+1)\hbar^2/2Ik_\mathrm{B}T} \ , \tag{7.38}$$

where $g_I = 3$ is the degeneracy of the nuclear spin states.

For H_2 the $(J = 0, F = 0)$ state is called para-hydrogen and the $(J = 1, F = 1)$ state ortho-hydrogen. At ambient temperature, both the $(J = 0, F = 0)$ state and the $(J = 1, F = 1)$ state are, approximately, fully populated and thus, the four nuclear spin substates described by Eqs. (7.33-7.36) are equally populated. Statistics then dictates a para-hydrogen to ortho-hydrogen ratio of 1:3 and no macroscopic spin polarization in a magnetic field. The splitting between the two states is

$$\frac{\epsilon_{J=1,F=1} - \epsilon_{J=0,F=0}}{k_\mathrm{B}} = 2\Theta_{\mathrm{rot}} = \frac{\hbar^2}{k_\mathrm{B}I} \approx 178.98\ \mathrm{K} \ , \tag{7.39}$$

where we have introduced a characteristic rotational temperature analogous to the characteristic vibrational temperature for the harmonic oscillator in Eq. (6.50). At temperatures well below this energy splitting, para-hydrogen is strongly enriched with respect to ortho-hydrogen. Equilibration in a reasonable time requires a catalyst. Still, no macroscopic spin polarization in a magnetic field is observed, as the two nuclear spins are magnetically equivalent and aligned antiparallel. If, however, para-hydrogen is reacted with a molecule featuring a multiple bond, magnetic equivalence of the two hydrogen atoms can be removed and in that case enhanced nuclear spin polarization is observable (para-hydrogen induced polarization, PHIP). We note that for $^2H_2$ the combination of nuclear spin states and rotational states to an allowed state reverses, as deuterons are bosons.

7.4 Rotational Partition Function

7.4.1 Rigid Rotor Assumption and Rotamers

Separation of the rotational partition function from the partition functions of other degrees of freedom does not only require consideration of nuclear spin states, but also the assumption that the *moment of inertia* is the same for all rotational states. This is generally true if the molecule behaves as a rigid rotor. For small molecules consisting of only a few atoms, this is often a good approximation. Larger molecules feature internal rotations, where a group of atoms rotates with respect to the rest of the molecule. In general, internal rotations are torsions about rotatable bonds, which are often not freely rotatable. The torsion potential has several minima and these minima are separated by energy barriers with heights that are larger, but not much larger than $k_\mathrm{B}T$. If we denote the number of potential minima for the i^th rotatable bond with $n_{\mathrm{min},i}$, the total number of rotameric states, short *rotamers* is

$$n_\mathrm{rot} = \prod_i n_{\mathrm{min},i} \, . \tag{7.40}$$

Each rotamer has it's own moment of inertia and, hence, its own set of states with respect to total rotation of the molecule. Because the energy scales of internal and total rotations are not well separated and because in larger molecules some vibrational modes may also have energies in this range, the partition function is not usually separable for large molecules. In such cases, insight into statistical thermodynamics can be best obtained by MD simulations. In the following, we consider small molecules that can be assumed to behave as a rigid rotor. We first consider diatomic molecules, where it certainly applies on the level of precision that we seek here.

The energy levels of a rigid rotor of a diatomic molecule are quantized by the *rotational quantum number* J and given by

$$\epsilon_J = J(J+1)\frac{h^2}{8\pi^2 I} = hcBJ(J+1) \, , \tag{7.41}$$

where

$$I = \mu r^2 \tag{7.42}$$

is the moment of inertia with the reduced mass μ and

$$B = \frac{h}{8\pi^2 Ic} \tag{7.43}$$

is the rotational constant. After introducing the characteristic rotational temperature,

$$\Theta_\mathrm{rot} = \frac{h^2}{8\pi^2 I k_\mathrm{B}} = \frac{hcB}{k_\mathrm{B}} \tag{7.44}$$

we have

$$\epsilon_J = J(J+1)k_\mathrm{B}\Theta_\mathrm{rot} \; . \tag{7.45}$$

As already mentioned, each rotational level has a degeneracy $g_J = 2J+1$. If all nuclei in the molecules are distinguishable (magnetically not equivalent), there is no correlation with nuclear spin states. In that case we have

$$Z_\mathrm{rot} = \sum (2J+1)\, e^{-J(J+1)\Theta_\mathrm{rot}/T} \; . \tag{7.46}$$

For sufficiently high temperatures and a sufficiently large moment of inertia, the density of states is sufficiently large for replacing the sum by an integral,

$$Z_\mathrm{rot} \approx \int_0^\infty (2J+1)\, e^{-J(J+1)\Theta_\mathrm{rot}/T} \mathrm{d}J = \frac{T}{\Theta_\mathrm{rot}} \; . \tag{7.47}$$

Deviations between the partition functions computed by Eq. (7.46) and Eq. (7.47) are visualized in Figure 7.1. As state functions depend on $\ln Z$, the continuum approximation is good for $T/\Theta_\mathrm{rot} \gg 1$, which applies to all gases, except at low temperatures for those that contain hydrogen. At ambient temperature it can be used in general, except for a further correction that we need to make because of symmetry considerations.

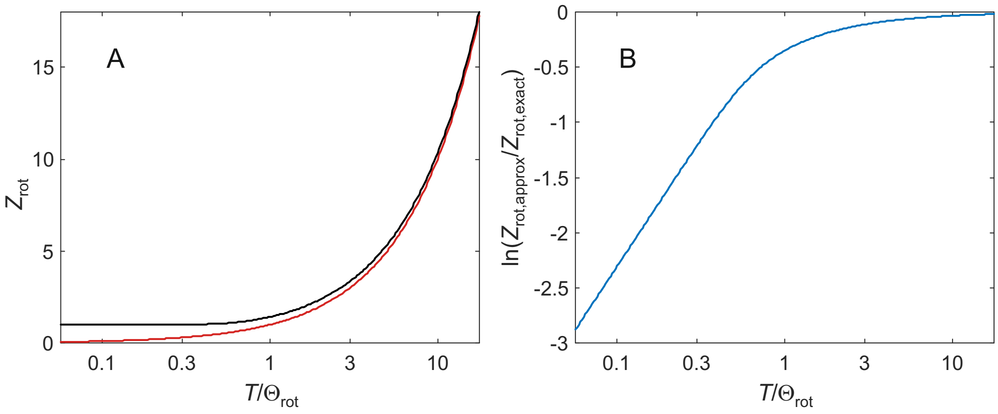


Figure 7.1: Continuum approximation for the rotational partition function. (A) Rotational partition function obtained by the sum expression (7.46) (black line) and by the integral expression (7.47) corresponding to the continuum approximation (red line). (B) Logarithm of the ratio between the rotational partition functions obtained by the continuum approximation (7.47) and the exact sum formula (7.46).

7.4.2 Accessible States and Symmetry

Even if all nuclei are magnetically inequivalent and, hence, distinguishable, rotational states may be not. For heteronuclear diatomic molecules they are, but for homonuclear diatomic molecules, they are not. To see this, we consider a symmetrical linear molecule that rotates by 180° about an axis perpendicular to the bond axis and centered in the middle of the bond. This rotation produces a configuration that is indistinguishable from the original configuration. In other words, the nuclear wavefunction is symmetric with respect to this rotation. For a homonuclear diatomic molecule, half of the rotational states are symmetric and half are antisymmetric. For nuclei that are bosons, such as ^{16}O in dioxygen, only the former states are allowed, for nuclei that are fermions, only the latter are allowed. Hence, we need to divide the partition function, which counts accessible states, by two. In this example, we have deliberately chosen a case with only one nuclear spin state. If nuclear spin states with different symmetry exist, all rotational states are accessible, but correlated to the nuclear spin states, as we have seen before for dihydrogen. In the following we consider the situation with only one nuclear spin state or for a fixed nuclear spin state.

Group	σ	**Group**	σ	**Group**	σ	**Group**	σ
$C_1, C_i, C_s, C_{\infty v}$	1	$D_{\infty h}$	2	T, T_d	12	O_h	24
C_n, C_{nv}, C_{nh}	n	D_n, D_{nh}, D_{nd}	$2n$	S_n	$n/2$	I_h	60

Table 7.1: Symmetry numbers σ corresponding to symmetry point groups [Her45; Iri98].

Although we still did not discuss other complications for multi-atomic molecules, we generalize this concept of symmetry-accessible states by introducing a *symmetry number* σ. In general, σ is the number of distinct orientations of a rigid molecule that are distinguished only by interchange of identical atoms. For an NH_3 molecule, rotation about the C_3 axis by 120° generates one such orientation from the other. No other rotation axis exists. Hence, $\sigma = 3$ for NH_3. In general, the symmetry number can be obtained from the molecule's point group [Her45; Iri98], as shown in Table 7.1.

With the symmetry number, the continuum approximation (7.47) becomes

$$Z_{\mathrm{rot}} = \frac{T}{\sigma\Theta_{\mathrm{rot}}} , \tag{7.48}$$

where we still assume that symmetry is sufficiently high for assigning a single characteristic temperature.

We further find

$$\ln Z_{\mathrm{rot}} = \ln T + \ln \frac{1}{\sigma\Theta_{\mathrm{rot}}} , \tag{7.49}$$

and with this

$$\begin{aligned} U_{\mathrm{rot}} &= N_{\mathrm{Av}} k_{\mathrm{B}} T^2 \left(\frac{\partial \ln Z_{\mathrm{rot}}}{\partial T} \right)_V & (7.50) \\ &= RT^2 \frac{\partial}{\partial T} \ln T & (7.51) \\ &= RT \text{ (linear)} . & (7.52) \end{aligned}$$

On going from Eq. (7.50) to (7.51) we could drop the second term on the right-hand side of Eq. (7.49), as this term does not depend on temperature. This is a general principle: Constant factors in the partition function do not contribute to internal energy. The result can be generalized to multi-atomic linear molecules that also have two rotational degrees of freedom. This result is expected from the equipartition theorem, as each degree of freedom should contribute a term $k_{\mathrm{B}}T/2$ to the molecular or a term $RT/2$ to the molar internal energy. However, if we refrain from the continuum approximation in Eq. (7.47) and numerically evaluate Eq. (7.46) instead, we find a lower contribution for temperatures lower than or comparable to Θ_{rot}. This is also a general principle: Contributions of modes to internal energy and, by inference, to heat capacity, are fully realized only at temperatures much higher than their characteristic temperature and are negligible at temperatures much lower than their characteristic temperature.[2]

---

[2] The zero-point vibrational energy is an exception from this principle with respect to internal energy, but not heat capacity.

For the rotation contribution to molar heat capacity at constant volume of a linear molecule we obtain

$$C_{\mathrm{rot},V} = \frac{\partial}{\partial T} U_{\mathrm{rot}} = R \, . \tag{7.53}$$

A non-linear multi-atomic molecule has, in general, three independent rotational moments of inertia corresponding to three pairwise orthogonal directions a, b, c. With

$$\Theta_{\mathrm{rot},\alpha} = \frac{h^2}{8\pi^2 I_\alpha k_\mathrm{B}} \quad (\alpha = a \, , b \, , c) \tag{7.54}$$

one finds for the partition function

$$Z_{\mathrm{rot}} = \frac{\sqrt{\pi}}{\sigma} \left(\frac{T^3}{\Theta_{\mathrm{rot},a} \Theta_{\mathrm{rot},b} \Theta_{\mathrm{rot},c}} \right)^{1/2} \, . \tag{7.55}$$

For spherical-top molecules, all three moments of inertia are equal, $I_a = I_b = I_c$, and hence all three characteristic temperatures are equal. For symmetric-top molecules, $I_a = I_b \neq I_c$.

The general equations for U_{rot} and $C_{\mathrm{rot},V}$ at sufficiently high temperature $T \gg \Theta_{\mathrm{rot}}$ are

$$U_{\mathrm{rot}} = \frac{d}{2} RT \tag{7.56}$$

$$C_{\mathrm{rot},V} = \frac{d}{2} R \, , \tag{7.57}$$

where $d = 1$ for a free internal rotation (e.g., about a C≡C bond), $d = 2$ for linear, and $d = 3$ for non-linear molecules. We note that the contribution of a free internal rotation needs to be added to the contribution from total rotation of the molecule.

The expressions for the rotational contributions to molar entropy are a bit lengthy and do not provide additional physical insight. They can be easily derived from the appropriate expressions for the rotational partition function and internal energy using Eq. (5.11). We note, however, that the contribution from the u/T term in the entropy expression is $dR/2$ and the contribution from $\ln Z$ is positive. Hence, at temperatures where the continuum approximation is valid, the rotational contribution to entropy is larger than $dR/2$.

7.5 Vibrational Partition Function

7.5.1 The Harmonic Oscillator Extended

Vibration in a diatomic molecule can be described by the 1D harmonic oscillator that we have considered in Section 6.3.2. In a multi-atomic molecule the $3n-5$ (linear) or $3n-6$ (non-linear) normal modes can be treated independently,

$$Z_{\mathrm{vib}} = \prod_{i=1}^{3n-5 \vee 3n-6} Z_{\mathrm{vib},i} = \prod_{i=1}^{3n-5 \vee 3n-6} \frac{1}{1 - e^{-\Theta_{\mathrm{vib},i}/T}} \, . \tag{7.58}$$

Normal mode energies are no longer independent and the partition function is no longer factorisable if anharmonicity of the vibration needs to be included, which is the case only at very high temperatures. We ignore this and ask about the limiting behavior of Z_{vib} for a diatomic molecule or $Z_{\mathrm{vib},i}$ for an individual normal mode at high temperatures. In the denominator of Eq. (6.51) we can make the approximation $e^{-\Theta_{\mathrm{vib}}/T} = 1 - \Theta_{\mathrm{vib}}/T$, if $\Theta_{\mathrm{vib}}/T \ll 1$. We obtain

$$\lim_{T \to \infty} Z_{\mathrm{vib},i} = \frac{T}{\Theta_{\mathrm{vib},i}} \, . \tag{7.59}$$

Vibrational temperatures for most normal modes are much higher than ambient temperature. Hence, at 298 K we have often $Z_{\mathrm{vib},i} \approx 1$. Appreciable deviations are observed for vibrations that involve heavy atoms, for instance $Z_{\mathrm{vib}} = 1.556$ at $T = 300$ K for I_2 [Mac98].

7.5.2 Vibrational Contributions to U, C_V, and S

The vibrational partition function for a system consisting of N diatomic molecules is

$$z_{\mathrm{vib}} = Z_{\mathrm{vib}}^N = \left(\frac{1}{1-e^{-\Theta_{\mathrm{vib}}/T}}\right)^N . \tag{7.60}$$

With $N = N_{\mathrm{Av}}$ we obtain the vibrational contribution to the molar internal energy

$$U_{\mathrm{vib}} = k_{\mathrm{B}} T^2 \left(\frac{\partial \ln z_{\mathrm{vib}}}{\partial T}\right)_V = \frac{N_{\mathrm{Av}} k_{\mathrm{B}} \Theta_{\mathrm{vib}}}{e^{\Theta_{\mathrm{vib}}/T}-1} \tag{7.61}$$

$$= \frac{R\Theta_{\mathrm{vib}}}{e^{\Theta_{\mathrm{vib}}/T}-1} . \tag{7.62}$$

For multi-atomic molecules the contributions from the individual normal modes with characteristic vibrational temperatures $\Theta_{\mathrm{vib},i}$ must be summed. Eq. (7.62) neglects the zero-point energy, as we had defined the partition function for an energy scaled by the zero-point energy. On an absolute energy scale, where $U = 0$ corresponds to the minimum of the Born-Oppenheimer potential energy hypersurface, an additional term $U_{\mathrm{zp}} = N_{\mathrm{Av}} h\nu_i/2$ needs to be added for each normal mode, with ν_i being the frequency of the normal mode.

The vibrational contribution to molar heat capacity at constant volume is

$$C_{\mathrm{vib},V} = \left(\frac{\partial U_{\mathrm{vib}}}{\partial T}\right)_V = R\left(\frac{\Theta_{\mathrm{vib}}}{T}\right)^2 \frac{e^{\Theta_{\mathrm{vib}}/T}}{\left(e^{\Theta_{\mathrm{vib}}/T}-1\right)^2} , \tag{7.63}$$

which is called the *Einstein equation*. With the Einstein function,

$$\mathcal{F}_{\mathrm{E}}(u) = \frac{u^2 e^u}{(e^u-1)^2} , \tag{7.64}$$

it can be written as

$$C_{\mathrm{vib},V} = RT\mathcal{F}_{\mathrm{E}}\left(\frac{\Theta_{\mathrm{vib}}}{T}\right) . \tag{7.65}$$

For computing the vibrational contribution to molar entropy we revert to the shifted energy scale. This is required, as inclusion of the zero-point contribution to u would leave us with an infinity. We find

$$S_{\mathrm{vib},i} = R\left[\frac{\Theta_{\mathrm{vib},i}}{T\left(e^{\Theta_{\mathrm{vib},i}/T}-1\right)} - \ln\left(1-e^{-\Theta_{\mathrm{vib},i}/T}\right)\right] . \tag{7.66}$$

Again contributions from individual normal modes add up. For $\Theta_{\mathrm{vib},i}/T \ll 1$, which is the usual case, both terms in the brackets are much smaller than unity, so that the contribution of any individual normal mode to entropy is much smaller than R. Hence, at ambient temperature the vibrational contribution to entropy is negligible compared to the rotational contribution unless the molecule contains heavy nuclei.

7.6 Electronic Partition Function

Atoms and molecules can also store energy by populating excited electronic states. For the hydrogen atom or any system that contains only a single electron, the energy levels can be given in closed form, based on the Bohr model,

$$\epsilon_{\mathrm{el},n} = -\frac{z_q^2 R_{\mathrm{E}}}{n^2} , \tag{7.67}$$

where $n = 0 \ldots \infty$ is the principal quantum number, z_q the nuclear charge, and R_E the Rydberg constant. However, this is an exception. For molecules and all other neutral atoms closed expressions for the energy levels cannot be found.

In most cases, the problem can be reduced to considering either only the electronic ground state with energy $\epsilon_{\mathrm{el},0}$ or to considering only excitation to the first excited state with energy $\epsilon_{\mathrm{el},1}$. If we use an energy shift to redefine $\epsilon_{\mathrm{el},0} = 0$, we can define a characteristic electronic temperature

$$\Theta_\mathrm{el} = \frac{\epsilon_{\mathrm{el},1}}{k_\mathrm{B}} \; . \tag{7.68}$$

Characteristic electronic temperatures are usually of the order of several thousand Kelvin. Hence, in most cases, $\Theta_\mathrm{el} \gg T$ applies, only the electronic ground state is accessible, and thus

$$Z_\mathrm{el} = g_{\mathrm{el},0} \; , \tag{7.69}$$

where $g_{\mathrm{el},0}$ is the degeneracy of the electronic ground state. We note that spatial degeneracy of the electronic ground state cannot exist in non-linear molecules, according to the *Jahn-Teller theorem*. However, a spatially non-degenerate ground state can still be spin-degenerate.

In molecules, total orbital angular momentum is usually quenched ($\Lambda = 0$, Σ ground state). In that case

$$Z_\mathrm{el}^{\{\Sigma\}} = 2S + 1 \; , \tag{7.70}$$

where S is the electron spin quantum number. For the singlet ground state of a closed-shell molecule ($S = 0$) we have $Z_\mathrm{el}^{\{\Sigma\}} = 1$, which means that the electronic contribution to the partition function is negligible. The contribution to internal energy and heat capacity is generally negligible for $\Theta_\mathrm{el} \gg T$. The electronic contribution to molar entropy,

$$S_\mathrm{el}^{\{\Sigma\}} = R \ln \left(2S + 1\right) \; , \tag{7.71}$$

is not negligible for open-shell molecules or atoms with $S > 0$. At high magnetic fields and low temperatures, e.g. at $T < 4.2$ K and $B_0 = 3.5$ T, where the high-temperature approximation for electron spin states does no longer apply, the electronic partition function and corresponding energy contribution are smaller than given in Eq. (7.70). For a doublet ground state ($S = 1/2$) the problem can be solved with the treatment that we have given in Section 6.3.1. For $S > 1/2$ the spin substates of the electronic ground state are not strictly degenerate even at zero magnetic field, but split by the zero-field splitting, which may exceed thermal energy in some cases. In that case Eq. (7.70) does not apply and the electronic contribution to the partition function depends on temperature. Accordingly, there is a contribution to internal energy and to heat capacity.

For a $\Lambda > 0$ species with term symbol ${}^{2S+1}\Lambda_\Omega$, each Ω component is doubly degenerate [SHM12]. For instance, for NO with a Π ground state ($\Lambda = 1$), both the ${}^2\Pi_{1/2}$ and the ${}^2\Pi_{3/2}$ state are doubly degenerate. As the ${}^2\Pi_{3/2}$ state is only 125 cm^{-1} above the ground state, the characteristic temperature for electronic excitation is $\Theta_\mathrm{el} = 178$ K. In this situation, Eq. (7.69) does not apply at ambient temperature. The energy gap to the next excited state, on the other hand, is very large. Thus, we have

$$Z_\mathrm{el} = g_{\mathrm{el},0} + g_{\mathrm{el},1} e^{-\Theta_\mathrm{el}/T} \; . \tag{7.72}$$

This equation is fairly general, higher excited states rarely need to be considered. The electronic contribution to the heat capacity of NO derived from Eq. (7.72) is in good agreement with experimental data from Eucken and d'Or [Mac98].

7.7 Equilibrium Constant for Gas Reactions

For clarity, we use an example reaction [WF12]

$$|\nu_\mathrm{A}|\ \mathrm{A} + |\nu_\mathrm{B}|\ \mathrm{B} \leftrightharpoons |\nu_\mathrm{C}|\ \mathrm{C} + |\nu_\mathrm{D}|\ \mathrm{D} \tag{7.73}$$

with adaptation to other reactions being straightforward. At equilibrium we must have

$$\Delta G = 0\ , \tag{7.74}$$

hence

$$\sum_i \nu_i \mu_i = \nu_\mathrm{A}\mu_\mathrm{A} + \nu_\mathrm{B}\mu_\mathrm{B} + \nu_\mathrm{C}\mu_\mathrm{C} + \nu_\mathrm{D}\mu_\mathrm{D} = 0\ , \tag{7.75}$$

where the μ_i are *molar* chemical potentials. To solve this problem, we do not need to explicitly work with the grand canonical ensemble, as we can compute the μ_i from the results that we have already obtained for the canonical ensemble. According to one of Gibbs' fundamental equations, which we derived in the lecture course on phenomenological thermodynamics, we have

$$\mathrm{d}f = -s\mathrm{d}T - p\mathrm{d}V + \sum_i \mu_i \mathrm{d}n_i\ . \tag{7.76}$$

Comparison of coefficients with the total differential of $f(T, V, n_i)$ reveals that

$$\mu_i = \left(\frac{\partial f}{\partial n_i}\right)_{T,V,n_{j\neq i}}\ , \tag{7.77}$$

a result that we had also obtained in the lecture course on phenomenological thermodynamics. Using Eq. (5.21), Eq. (5.18), and Stirling's formula, we obtain for the contribution f_i of an individual chemical species to Helmholtz free energy

$$f_i = -k_\mathrm{B}T\ln z_i \tag{7.78}$$

$$= -k_\mathrm{B}T\ln\frac{1}{N_i!}Z_i^{N_i} \tag{7.79}$$

$$= -k_\mathrm{B}T\left(N_i\ln Z_i - N_i\ln N_i + N_i\right) \tag{7.80}$$

$$= -n_iRT\ln\frac{Z_i}{n_iN_\mathrm{Av}} - n_iRT\ , \tag{7.81}$$

where n_i is the amount of substance (mol). Eq. (7.77) then gives

$$\mu_i = n_iRT\cdot\frac{1}{n_i} - RT\ln\frac{Z_i}{n_iN_\mathrm{Av}} - RT \tag{7.82}$$

$$= -RT\ln\frac{Z_i}{n_iN_\mathrm{Av}} \tag{7.83}$$

$$= -RT\ln\frac{Z_i}{N_i}\ . \tag{7.84}$$

Eq. (7.84) expresses the dependence of the chemical potential, a *molar* property, on the *molecular* partition function. It may appear odd that this property depends on the *absolute* number of molecules N_i, but exactly this introduces the contribution of mixing entropy that counterbalances the differences in *standard* chemical potentials $\mu_i^\ominus$. Because of our habit of shifting energies by $\epsilon_\mathrm{el,0}$ and by the zero-point vibration energies, we cannot directly apply Eq.

(7.84). We can avoid explicit dependence on the $\epsilon_{\mathrm{el},0,i}$ and the zero-point vibrational energies by relying on *Hess' law* and referencing energies of all molecules to the state where they are fully dissociated into atoms. The energies $\epsilon_{i,\mathrm{diss}}$ for the dissociated states can be defined at 0 K. We find

$$
\begin{aligned}
Z_{i,\mathrm{corr}} &= \sum_j e^{-\left(\epsilon_{ij}-\epsilon_{i,\mathrm{diss}}\right)/k_\mathrm{B}T} && (7.85)\\
&= e^{\epsilon_{i,\mathrm{diss}}/k_\mathrm{B}T} \sum_j e^{-\epsilon_{ij}/k_\mathrm{B}T} && (7.86)\\
&= Z_i e^{\epsilon_{i,\mathrm{diss}}/k_\mathrm{B}T} \ , && (7.87)
\end{aligned}
$$

where index j runs over the states of molecule i.

With this correction we have

$$\Delta G = -RT \sum_i \nu_i \ln \frac{Z_i e^{\epsilon_{i,\mathrm{diss}}/k_\mathrm{B}T}}{N_i} \ . \tag{7.88}$$

For our example reaction, the equilibrium condition is

$$\nu_\mathrm{A}\mu_\mathrm{A} + \nu_\mathrm{B}\mu_\mathrm{B} = -\nu_\mathrm{C}\mu_\mathrm{C} - \nu_\mathrm{D}\mu_\mathrm{D} \ , \tag{7.89}$$

which gives

$$
\begin{aligned}
&- RT\nu_\mathrm{A} \ln \frac{Z_\mathrm{A} e^{\epsilon_{\mathrm{A,diss}}/k_\mathrm{B}T}}{N_\mathrm{A}} - RT\nu_\mathrm{B} \ln \frac{Z_\mathrm{B} e^{\epsilon_{\mathrm{B,diss}}/k_\mathrm{B}T}}{N_\mathrm{B}}\\
=&RT\nu_\mathrm{C} \ln \frac{Z_\mathrm{C} e^{\epsilon_{\mathrm{C,diss}}/k_\mathrm{B}T}}{N_\mathrm{C}} + RT\nu_\mathrm{D} \ln \frac{Z_\mathrm{D} e^{\epsilon_{\mathrm{D,diss}}/k_\mathrm{B}T}}{N_\mathrm{D}} && (7.90)
\end{aligned}
$$

and can be rearranged to

$$\ln \frac{Z_\mathrm{A}^{-\nu_A} \cdot e^{-\nu_\mathrm{A}\epsilon_{\mathrm{A,diss}}/k_\mathrm{B}T} \cdot Z_\mathrm{B}^{-\nu_B} \cdot e^{-\nu_\mathrm{B}\epsilon_{\mathrm{B,diss}}/k_\mathrm{B}T}}{N_\mathrm{A}^{-\nu_\mathrm{A}} \cdot N_\mathrm{B}^{-\nu_\mathrm{B}}} = \ln \frac{Z_\mathrm{C}^{\nu_C} \cdot e^{\nu_\mathrm{C}\epsilon_{\mathrm{C,diss}}/k_\mathrm{B}T} \cdot Z_\mathrm{D}^{\nu_D} \cdot e^{\nu_\mathrm{D}\epsilon_{\mathrm{D,diss}}/k_\mathrm{B}T}}{N_\mathrm{C}^{\nu_\mathrm{C}} \cdot N_\mathrm{D}^{\nu_\mathrm{D}}} \tag{7.91}$$

and further rearranged to

$$\frac{N_\mathrm{C}^{|\nu_\mathrm{C}|} \cdot N_\mathrm{D}^{|\nu_\mathrm{D}|}}{N_\mathrm{A}^{|\nu_\mathrm{A}|} \cdot N_\mathrm{B}^{|\nu_\mathrm{B}|}} = \frac{Z_\mathrm{C}^{|\nu_\mathrm{C}|} \cdot Z_\mathrm{D}^{|\nu_\mathrm{D}|}}{Z_\mathrm{A}^{|\nu_\mathrm{A}|} \cdot Z_\mathrm{B}^{|\nu_\mathrm{B}|}} \cdot e^{\left(\nu_\mathrm{A}\epsilon_{\mathrm{A,diss}}+\nu_\mathrm{B}\epsilon_{\mathrm{B,diss}}+\nu_\mathrm{C}\epsilon_{\mathrm{C,diss}}+\nu_\mathrm{D}\epsilon_{\mathrm{D,diss}}\right)/k_\mathrm{B}T} \ . \tag{7.92}$$

In Eq. (7.92) we can make the identifications

$$K_N\left(V,T\right) = \frac{N_\mathrm{C}^{|\nu_\mathrm{C}|} \cdot N_\mathrm{D}^{|\nu_\mathrm{D}|}}{N_\mathrm{A}^{|\nu_\mathrm{A}|} \cdot N_\mathrm{B}^{|\nu_\mathrm{B}|}} \ , \tag{7.93}$$

where $K_N\left(V,T\right)$ is a volume-dependent equilibrium constant expressed with particle numbers, and, since dissociation energies are negative energies of formation,

$$\Delta U_0 = -N_\mathrm{Av}\left(\nu_\mathrm{A}\epsilon_{\mathrm{A,diss}} + \nu_\mathrm{B}\epsilon_{\mathrm{B,diss}} + \nu_\mathrm{C}\epsilon_{\mathrm{C,diss}} + \nu_\mathrm{D}\epsilon_{\mathrm{D,diss}}\right) \ , \tag{7.94}$$

where ΔU_0 is the molar reaction energy at 0 K. Hence, we have

$$K_N\left(V,T\right) = \frac{Z_\mathrm{C}^{|\nu_\mathrm{C}|} \cdot Z_\mathrm{D}^{|\nu_\mathrm{D}|}}{Z_\mathrm{A}^{|\nu_\mathrm{A}|} \cdot Z_\mathrm{B}^{|\nu_\mathrm{B}|}} e^{-\Delta U_0/RT} \ . \tag{7.95}$$

The dependence on volume arises from the dependence of the canonical partition functions on volume.

By dividing all particle numbers by $N_{\mathrm{Av}}^{\nu_i}$ and volume V^{ν_i}, we obtain the equilibrium constant $K_c(T)$ in molar concentrations

$$K_c\left(T\right) = \frac{Z_{\mathrm{C}}^{|\nu_{\mathrm{C}}|} \cdot Z_{\mathrm{D}}^{|\nu_{\mathrm{D}}|}}{Z_{\mathrm{A}}^{|\nu_{\mathrm{A}}|} \cdot Z_{\mathrm{B}}^{|\nu_{\mathrm{B}}|}} \cdot \left(N_{\mathrm{Av}} V\right)^{-\sum_i \nu_i} \cdot e^{-\Delta U_0/RT} \ . \tag{7.96}$$

By dividing them by the total particle number $N = \sum_i N_i$ to the power of ν_i we obtain

$$K_x(V,T) = \frac{Z_{\mathrm{C}}^{|\nu_{\mathrm{C}}|} \cdot Z_{\mathrm{D}}^{|\nu_{\mathrm{D}}|}}{Z_{\mathrm{A}}^{|\nu_{\mathrm{A}}|} \cdot Z_{\mathrm{B}}^{|\nu_{\mathrm{B}}|}} \cdot N^{-\sum_i \nu_i} \cdot e^{-\Delta U_0/RT} \ , \tag{7.97}$$

which coincides with the thermodynamical equilibrium constant $K^\dagger$ at the standard pressure $p^\ominus$. The most useful equilibrium constant for gas-phase reactions is obtained by inserting $p_i = c_i RT$ into Eq. (7.96)[3]:

$$K_p\left(T\right) = \frac{Z_{\mathrm{C}}^{|\nu_{\mathrm{C}}|} \cdot Z_{\mathrm{D}}^{|\nu_{\mathrm{D}}|}}{Z_{\mathrm{A}}^{|\nu_{\mathrm{A}}|} \cdot Z_{\mathrm{B}}^{|\nu_{\mathrm{B}}|}} \cdot \left(\frac{RT}{N_{\mathrm{Av}} V}\right)^{\sum_i \nu_i} \cdot e^{-\Delta U_0/RT} \ . \tag{7.98}$$

For each molecular species, the molecular partition function is a product of the contributions from individual modes, Eq. (7.7), that we have discussed above. In the expression for equilibrium constants, the nuclear-spin contribution cancels out since the number of nuclei and their spins are the same on both sides of the reaction equation. Symmetry requirements on the nuclear wavefunction are considered in the symmetry numbers σ_i for the rotational partition function. The electronic contribution often reduces to the degeneracy of the electronic ground state and in the vibrational contribution, normal modes with $\Theta_{\mathrm{vib},i} > 5T$ can be neglected.

---

[3] There is a misprint in [WF12]

Thermodynamics of Mixing
- Entropy of Binary Mixing
- Energy of Binary Mixing

Entropic Elasticity
- Ideal Chain Model
- Random Walk
- Conformational Entropy and Free Energy

8 — Macromolecules

8.1 Thermodynamics of Mixing

The formalism introduced in Chapter 7 is suitable for small molecules in the gas phase, but does not easily extend to condensed phases or to larger molecules with several rotameric states, let alone to macromolecules, such as synthetic polymers, proteins, nucleic acids, and carbohydrates. Nevertheless, statistical thermodynamics is an important theory for understanding such systems. In this Chapter we introduce some of the concepts of statistical thermodynamics that do not depend on explicit computation of the partition function. We start with the entropy of mixing and, for simplicity, restrain the discussion to binary mixtures [RC03].

8.1.1 Entropy of Binary Mixing

We consider mixing of two species A with volume V_A and B with volume V_B and neglect volume change, so that the total volume is $V_\mathrm{A} + V_\mathrm{B}$. The volume fractions of the two components in the mixture are thus given by

$$\phi_\mathrm{A} = \frac{V_\mathrm{A}}{V_\mathrm{A} + V_\mathrm{B}} \tag{8.1}$$

$$\phi_\mathrm{B} = \frac{V_\mathrm{B}}{V_\mathrm{A} + V_\mathrm{B}} = 1 - \phi_\mathrm{A} \ . \tag{8.2}$$

To consider the statistics of the problem we use a lattice model.

Concept 8.1.1 — Lattice model. A *lattice model* is a discrete representation of a system as opposed to a continuum representation. A three-dimensional lattice model is a regular arrangement of sites in Cartesian space, such as a crystal lattice is a regular arrangement of atoms in Cartesian space. The state of the model is defined by the distribution of units of matter, for instance molecules or the repeat units of a polymer (short: monomers), on the lattice sites. In statistical thermodynamics, one particular arrangement of the units on the lattice is a microstate. Energy of the microstate depends on interactions of units between lattice sites, in the simplest case only between direct neighbor sites. By considering the statistical distribution of microstates, thermodynamic state functions of the macrostate of the system can be obtained.

In our example we assign the lattice site a volume v_0, which cannot be larger than the volume required for one molecule of the smaller component in the mixture. The other component may then also occupy a single site (similarly sized components) or several lattice sites. A macromolecule with a large degree of polymerization consists of a large number of monomers and will thus occupy a large number of lattice sites. The molecular volumes of the species are

$$v_\mathrm{A} = N_\mathrm{A} v_0 \tag{8.3}$$

$$v_\mathrm{B} = N_\mathrm{B} v_0 \ , \tag{8.4}$$

where N_A and N_B are the number of sites occupied by one molecule of species A and B, respectively. We consider the three simple cases listed in Table 8.1. *Regular solutions* are mixtures of two low molecular weight species with $N_\mathrm{A} = N_\mathrm{B} = 1$. *Polymer solutions* are mixtures of one type of macromolecules ($N_\mathrm{A} = N \gg 1$) with a solvent, whose molecular volume defines the lattice site volume v_0 ($N_\mathrm{B} = 1$). *Polymer blends* correspond to the general case $1 \neq N_\mathrm{A} \neq N_\mathrm{B} \neq 1$. They are mixtures of two different species of macromolecules, so that $N_\mathrm{A}, N_\mathrm{B} \gg 1$.

	N_A	N_B
Regular solutions	1	1
Polymer solutions	N	1
Polymer blends	N_A	N_B

Table 8.1: Number of lattice sites occupied per molecule in different types of mixtures [RC03].

The mixture occupies

$$n = \frac{V_\mathrm{A} + V_\mathrm{B}}{v_0} \tag{8.5}$$

lattice sites, whereas component A occupies $V_\mathrm{A}/v_0 = n\phi_\mathrm{A}$ of these sites. We consider a microcanonical ensemble and can thus express entropy as

$$s = k_\mathrm{B} \ln \Omega \ , \tag{8.6}$$

where Ω is the number of ways in which the molecules can be arranged on the lattice (number of microstates). In a homogeneous mixture, a molecule or monomer of component A can occupy any of the n lattice sites. Before mixing, it can occupy only one of the lattice sites in volume V_A. Hence, the entropy change for one molecule of species A is

$$\begin{aligned} \Delta S_\mathrm{A} &= k_\mathrm{B} \ln n - k_\mathrm{B} \ln \phi_\mathrm{A} n && (8.7)\\ &= k_\mathrm{B} \ln \frac{n}{\phi_\mathrm{A} n} && (8.8)\\ &= -k_\mathrm{B} \ln \phi_\mathrm{A} \ . && (8.9) \end{aligned}$$

The total mixing entropy for both species is

$$\Delta s_\mathrm{mix} = -k_\mathrm{B} \left(n_\mathrm{A} \ln \phi_\mathrm{A} + n_\mathrm{B} \ln \phi_\mathrm{B} \right) \ . \tag{8.10}$$

We note the analogy with the expression that we had obtained in phenomenological thermodynamics for an ideal mixture of ideal gases, were we had used the molar fraction x_i instead of the volume fraction ϕ_i. For ideal gases, $V_i \propto n_i$ and thus $\phi_i = x_i$. Eq. (8.10) generalizes the result to any ideal mixture in condensed phase. The mixture is ideal because we did not yet consider energy of mixing and thus could get away with using a microcanonical ensemble.

For discussion it is useful to convert the extensive quantity Δs_{mix} to the intensive entropy of mixing per lattice site,

$$\Delta \overline{S}_{\mathrm{mix}} = -k_{\mathrm{B}} \left(\frac{\phi_{\mathrm{A}}}{N_{\mathrm{A}}} \ln \phi_{\mathrm{A}} + \frac{\phi_{\mathrm{B}}}{N_{\mathrm{B}}} \ln \phi_{\mathrm{B}} \right) , \tag{8.11}$$

where we have used the number of molecules per species $n_i = n\phi_i/N_i$ and normalized by the total number n of lattice sites.

For a regular solution with $N_{\mathrm{A}} = N_{\mathrm{B}} = 1$ we obtain the largest entropy of mixing at given volume fractions of the components,

$$\Delta \overline{S}_{\mathrm{mix}} = -k_{\mathrm{B}} \left(\phi_{\mathrm{A}} \ln \phi_{\mathrm{A}} + \phi_{\mathrm{B}} \ln \phi_{\mathrm{B}} \right) \ \text{(regular solutions)} . \tag{8.12}$$

For a polymer solution with $N_{\mathrm{A}} = N \gg 1$ and $N_{\mathrm{B}} = 1$ we have

$$\Delta \overline{S}_{\mathrm{mix}} = -k_{\mathrm{B}} \left(\frac{\phi_{\mathrm{A}}}{N} \ln \phi_{\mathrm{A}} + \phi_{\mathrm{B}} \ln \phi_{\mathrm{B}} \right) \tag{8.13}$$

$$\approx -k_{\mathrm{B}} \phi_{\mathrm{B}} \ln \phi_{\mathrm{B}} , \tag{8.14}$$

where the approximation by Eq. (8.14) holds for $\phi_{\mathrm{B}} \gg 1/N$, i.e. for solving a polymer and even for any appreciable swelling of a high-molecular weight polymer by a solvent. For polymer blends, Eq. (8.11) holds with $N_{\mathrm{A}}, N_{\mathrm{B}} \gg 1$. Compared to formation of a regular solution or a polymer solution, mixing entropy for a polymer blend is negligibly small, which qualitatively explains the difficulty of producing such polymer blends. Nevertheless, the entropy of mixing is always positive, and thus the Helmholtz free energy $\Delta \overline{F}_{\mathrm{mix}} = -T\Delta \overline{S}_{\mathrm{mix}}$ always negative, so that an *ideal* mixture of two polymers should form spontaneously. To see what happens in real mixtures, we have to consider the energetics of mixing.

Before doing so, we note the limitations of the simple lattice model. We have neglected conformational entropy of the polymer, which will be discussed in Section 8.2.3. This amounts to the assumption that conformational entropy does not change on mixing. For blends of polymers, this is a very good assumption, whereas in polymers solutions there is often an excluded volume that reduces conformational space. We have also neglected the small volume change that occurs on mixing, most notably for regular solutions. For polymer solutions and blends this volume change is very small.

8.1.2 Energy of Binary Mixing

To discuss the internal energy contribution to the free energy of mixing, we continue using the simplified lattice model. In particular, we consider mixing at constant volume and we assume that attractive or repulsive interactions between lattice sites are sufficiently small to not perturb random distributions of solvent molecules and monomers on lattice sites. We also ignore that the polymer chain is connected, as this would exclude random distribution of the monomers to the lattice sites. Regular solution theory, as we consider it here, is a mean-field approach where the interaction at a given lattice site is approximated by a mean interaction with the other lattice sites. This neglects correlations. Although the model may appear crude (as many models in polymer physics), it provides substantial insight and an expression that fits experimental data surprisingly well (as is the case for many crude models in polymer physics).

We start by defining three pairwise interaction energies u_{AA}, u_{AB}, and u_{BB} between adjacent sites of the lattice. For random distribution, the probability that a molecule or monomer A has a neighbor A is ϕ_{A} and the probability that it has a neighbor B is $1 - \phi_{\mathrm{A}}$. We neglect boundary effects, as the ratio between the number of surface sites and inner sites is very small for a macroscopic system. The mean-field interaction energy per lattice site occupied by an A unit is thus

$$U_\mathrm{A} = \phi_\mathrm{A} u_\mathrm{AA} + (1-\phi_\mathrm{A})\, u_\mathrm{AB} \tag{8.15}$$

and the corresponding expression for a lattice site occupied by a B unit is

$$U_\mathrm{B} = \phi_\mathrm{A} u_\mathrm{AB} + (1-\phi_\mathrm{A})\, u_\mathrm{BB}\ . \tag{8.16}$$

To continue, we need to specify the lattice, as the number of sites a adjacent to the site under consideration depends on that. For a cubic lattice we would have $a = 6$. We keep a as a parameter in the hope that we can eliminate it again at a later stage. If we compute a weighted sum of the expressions (8.15) and (8.16) we double count each pairwise interaction, as we will encounter it twice. Hence, total interaction energy of the mixture is

$$u = \frac{an}{2}\left[\phi_\mathrm{A} U_\mathrm{A} + (1-\phi_\mathrm{A})\, U_\mathrm{B}\right]\ , \tag{8.17}$$

where we have used the probability ϕ_A of encountering a site occupied by a unit A and $(1-\phi_\mathrm{A})$ of encountering a site occupied by a unit B. By inserting Eqs. (8.15) and (8.16) into Eq. (8.17) and abbreviating $\phi_\mathrm{A} = \phi$, we obtain

$$\begin{aligned} u &= \frac{an}{2}\left\{\phi\left[\phi u_\mathrm{AA} + (1-\phi)\, u_\mathrm{AB}\right] + (1-\phi)\left[\phi u_\mathrm{AB} + (1-\phi)\, u_\mathrm{BB}\right]\right\} && (8.18)\\ &= \frac{an}{2}\left[\phi^2 u_\mathrm{AA} + 2\phi\,(1-\phi)\, u_\mathrm{AB} + (1-\phi)^2\, u_\mathrm{BB}\right]\ . && (8.19)\end{aligned}$$

Before mixing the interaction energy per site in pure A is $a u_\mathrm{AA}/2$ and in B $a u_\mathrm{BB}/2$. Hence, the total interaction energy before mixing is

$$u_0 = \frac{an}{2}\left[\phi u_\mathrm{AA} + (1-\phi)\, u_\mathrm{BB}\right]\ , \tag{8.20}$$

so that we obtain for the energy change $\Delta u = u - u_0$ on mixing

$$\begin{aligned} \Delta u &= \frac{an}{2}\left[\phi^2 u_\mathrm{AA} + 2\phi\,(1-\phi)\, u_\mathrm{AB} + (1-\phi)^2\, u_\mathrm{BB} - \phi u_\mathrm{AA} - (1-\phi)\, u_\mathrm{BB}\right] && (8.21)\\ &= \frac{an}{2}\left[\left(\phi^2-\phi\right) u_\mathrm{AA} + 2\phi\,(1-\phi)\, u_\mathrm{AB} + \left(1-2\phi+\phi^2-1+\phi\right) u_\mathrm{BB}\right] && (8.22)\\ &= \frac{an}{2}\left[\phi\,(\phi-1)\, u_\mathrm{AA} + 2\phi\,(1-\phi)\, u_\mathrm{AB} + \phi\,(\phi-1)\, u_\mathrm{BB}\right] && (8.23)\\ &= \frac{an}{2}\phi\,(1-\phi)\,(2u_\mathrm{AB} - u_\mathrm{AA} - u_\mathrm{BB})\ . && (8.24)\end{aligned}$$

We again normalize by the number n of lattice sites to arrive at the energy change per site on mixing:

$$\Delta\overline{U}_\mathrm{mix} = \frac{a}{2}\phi\,(1-\phi)\,(2u_\mathrm{AB} - u_\mathrm{AA} - u_\mathrm{BB})\ . \tag{8.25}$$

For discussion we need an expression that characterizes the mixing energy per lattice site as a function of composition ϕ and that can be easily combined with the mixing entropy to free energy. The *Flory interaction parameter*,

$$\chi = \frac{a}{2}\cdot\frac{2u_\mathrm{AB} - u_\mathrm{AA} - u_\mathrm{BB}}{k_\mathrm{B}T}\ , \tag{8.26}$$

elegantly eliminates the number of adjacent lattice sites and provides just such an expression:

$$\Delta\overline{U}_\mathrm{mix} = \chi\phi\,(1-\phi)\, k_\mathrm{B}T\ . \tag{8.27}$$

Introducing such a parameter is an often-used trick when working with crude models. If the parameter is determined experimentally, the expression may fit data quite well, because part of the deviations of reality from the model can be absorbed by the parameter and its dependence on state variables. We finally obtain the *Flory-Huggins equation* for the Helmholtz free energy of mixing, $\Delta\overline{F}_{\mathrm{mix}} = \Delta\overline{U}_{\mathrm{mix}} - T\Delta\overline{S}_{\mathrm{mix}}$,

$$\Delta\overline{F}_{\mathrm{mix}} = k_{\mathrm{B}}T\left[\frac{\phi}{N_{\mathrm{A}}}\ln\phi + \frac{1-\phi}{N_{\mathrm{B}}}\ln\left(1-\phi\right) + \chi\phi\left(1-\phi\right)\right] . \tag{8.28}$$

As the entropy contribution (first two terms in the brackets on the right-hand side of Eq. (8.28)) to $\Delta\overline{F}_{\mathrm{mix}}$ is always negative, entropy always favors mixing. The sign of $\Delta\overline{F}_{\mathrm{mix}}$ depends on the sign of the Flory parameter χ and the ratio between the energy and entropy. The Flory parameter is negative and thus favors mixing, if $2u_{\mathrm{AB}} < u_{\mathrm{AA}} + u_{\mathrm{BB}}$, i.e., if the interaction in AB pairs is more attractive than the mean interaction in AA and and BB pairs. Such cases occur, but are rare. In most cases, the Flory parameter is positive. Since the entropy terms are very small for polymer blends, such blends tend to phase separate. In fact, high molecular weight poly(styrene) with natural isotope abundance phase separates from deuterated poly(styrene).

8.2 Entropic Elasticity

8.2.1 Ideal Chain Model

Most polymer chains have rotatable bonds as well as bond angles along the polymer backbone that differ from 180°. This leads to flexibility of the chain. Even if the rotations are not free, but give rise to only n_{rot} rotameric states per rotatable bond, the number of possible chain conformations becomes vast. For N_{rot} rotatable bonds, the number of distinct conformations is $n_{\mathrm{rot}}^{N_{\mathrm{rot}}}$. The simplest useful model for such a flexible chain is the *freely jointed chain model* [RC03]. Here we assume bond vectors that all have the same length $l = |\vec{r}_i|$, where $\vec{r}_i$ is the bond vector of the i^{th} bond. If we further assume an angle θ_{ij} between consecutive bond vectors, we can write the scalar product of consecutive bond vectors as

$$\vec{r}_i \cdot \vec{r}_j = l^2 \cos\theta_{ij} . \tag{8.29}$$

This scalar product is of interest, as we can use it to compute the mean-square end-to-end distance $\langle R^2 \rangle$ of an ensemble of chains, which is the simplest parameter that characterizes the spatial dimension of the chain. With the end-to-end distance vector of a chain with n bonds,

$$\vec{R}_n = \sum_{i=1}^{n} \vec{r}_i , \tag{8.30}$$

we have

$$\begin{aligned}
\langle R^2 \rangle &= \langle \vec{R}_n^2 \rangle && (8.31)\\
&= \langle \vec{R}_n \cdot \vec{R}_n \rangle && (8.32)\\
&= \left\langle \left(\sum_{i=1}^{n} \vec{r}_i\right) \cdot \left(\sum_{j=1}^{n} \vec{r}_j\right) \right\rangle && (8.33)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \langle \vec{r}_i \cdot \vec{r}_j \rangle . && (8.34)
\end{aligned}$$

By using Eq. (8.29) we find

$$\langle R^2 \rangle = l^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \langle \cos\theta_{ij} \rangle \ . \tag{8.35}$$

In the freely jointed chain model we further assume that there are no correlations between the directions of different bond vectors, $\langle \cos\theta_{ij} \rangle = 0$ for $i \neq j$. Then, the double sum in Eq. (8.35) has only n non-zero terms for $i = j$ with $\cos\theta_{ij} = 1$. Hence,

$$\langle R^2 \rangle = nl^2 \ . \tag{8.36}$$

This again appears to be a crude model, but we shall now rescue it by redefining l. In an ideal polymer chain we can at least assume that there is no interaction between monomers that are separated by many other monomers,

$$\lim_{|i-j| \to \infty} \langle \cos\theta_{ij} \rangle = 0 \ . \tag{8.37}$$

Furthermore, for a given bond vector $\vec{r}_i$ the sum over all correlations with other bond vectors converges to some finite number that depends on i,

$$\sum_{j=1}^{n} \langle \cos\theta_{ij} \rangle = C'(i) \ . \tag{8.38}$$

Therefore, when including the correlations, Eq. (8.35) can still be simplified to

$$\langle R^2 \rangle = l^2 \sum_{i=1}^{n} C'(i) = C_n n l^2 \ , \tag{8.39}$$

where *Flory's characteristic ratio* C_n is the average value of $C'(i)$ over all backbone bonds of the chain.

In general, C_n depends on n, but for very long chains it converges to a value C_∞. For sufficiently long chains, we can thus approximate

$$\langle R^2 \rangle \approx n C_\infty l^2 \ , \tag{8.40}$$

which has the same dependence on n and l as the crude model of the freely jointed chain, Eq. (8.36). Hence, we can define an *equivalent freely jointed chain* with N Kuhn segments of length b. From

$$\langle R^2 \rangle = N b^2 \approx n C_\infty l^2 \tag{8.41}$$

and the length of the maximally stretched equivalent chain, the *contour length* R_max,

$$R_\mathrm{max} = Nb \ , \tag{8.42}$$

we obtain

$$N = \frac{R_\mathrm{max}^2}{C_\infty n l^2} \tag{8.43}$$

and the *Kuhn length*

$$b = \frac{\langle R^2 \rangle}{R_\mathrm{max}} = \frac{C_\infty n l^2}{R_\mathrm{max}} \ . \tag{8.44}$$

Typical values of C_∞ for synthetic polymers range from 4.6 for 1,4-poly(isoprene) to 9.5 for atactic poly(styrene) with corresponding Kuhn lengths of 8.2 Å to 18 Å, respectively.

At this point we have found the mean-square end-to-end distance as a parameter of an equilibrium macrostate. If we stretch the chain to a longer end-to-end distance, it is no longer at equilibrium and must have larger free energy. Part of this increase in free energy must come from a decrease in entropy that stretching induces by reducing the number of accessible chain conformations. It turns out that this entropic contribution is the major part of the increase in free energy, typically 90%. The tendency of polymer chains to contract after they have been stretched is thus mainly an entropic effect. To quantify it, we need a probability distribution for the end-to-end vectors and to that end, we introduce a concept that is widely used in natural sciences.

8.2.2 Random Walk

The freely jointed chain model explicitly assumes that the direction of the next Kuhn segment is uncorrelated to the directions of all previous Kuhn segments. Where the chain end will be located after the next step that prolongs the chain by one segment depends only on the location of the current chain end. The freely jointed chain thus has aspects of a Markov chain. Each prolongation step is a random event and the trajectory of the chain in space a random walk.

Concept 8.2.1 — Random walk. Many processes can be discretized into individual steps. What happens in the next step may depend on only the current state or also on what happened in earlier steps. If it depends only on the current state, the process is memoryless and fits the definition of a Markov chain. A Markov chain where the events are analogous steps in some parameter space can be modeled as a *random walk*. A random walk is a mathematically formalized succession of random steps. A random walk on a lattice, where each step can only lead from a lattice point to a directly neighboring lattice point is a particularly simple model.

We can use the concept of a random walk in combination with the concepts of statistical thermodynamics in order to solve the problem of polymer chain stretching and contraction. The problem is solved if we know the dependence of Helmholtz free energy on the length of the end-to-end vector. This, in turn, requires that we know the entropy and thus the probability distribution of the length of the end-to-end vector. This probability distribution is given by the number of possible random walks (trajectories) that lead to a particular end-to-end distance $\sqrt{\vec{R}^2}$.

For simplicity we start with a simpler example in one dimension that we can later extend to three dimensions. We consider the standard example in this field, a drunkard who has just left a pub. We assume that, starting at the pub door, he makes random steps forth and back along the road. What is the probability $P(N,x)$ that after N steps he is at a distance of x steps up the road from the pub door? The problem is equivalent to finding the number $W(N,x)$ of trajectories of length N that end up x steps from the pub door and dividing it by the total number of trajectories.

Any such trajectory consists of N_+ steps up the road and N_- steps down the road, with the final position being $x = N_+ - N_-$ [RC03]. The number of such trajectories is, again, given by a binomial distribution (see Section 3.1.6)

$$W(N,x) = \frac{(N_+ + N_-)!}{N_+! N_-!} = \frac{N!}{[(N+x)/2]!\,[(N-x)/2]!} , \tag{8.45}$$

whereas the total number of trajectories is 2^N, as the drunkard has two possibilities at each step.

Hence,

$$P(N,x) = \frac{1}{2^N} \cdot \frac{N!}{[(N+x)/2]!\,[(N-x)/2]!}\,, \tag{8.46}$$

leading to

$$\ln P(N,x) = -N\ln 2 + \ln(N!) - \ln\left(\frac{N+x}{2}\right)! - \ln\left(\frac{N-x}{2}\right)!\,. \tag{8.47}$$

The last two terms on the right-hand side can be rewritten as

$$\ln\left(\frac{N+x}{2}\right)! = \ln\left(\frac{N}{2}\right)! + \sum_{s=1}^{x/2} \ln\left(\frac{N}{2}+s\right) \text{ and} \tag{8.48}$$

$$\ln\left(\frac{N-x}{2}\right)! = \ln\left(\frac{N}{2}\right)! - \sum_{s=1}^{x/2} \ln\left(\frac{N}{2}+1-s\right)\,, \tag{8.49}$$

which leads to

$$\ln P(N,x) = -N\ln 2 + \ln(N!) - 2\ln\left(\frac{N}{2}\right)! - \sum_{s=1}^{x/2} \ln\left(\frac{N/2+s}{N/2+1-s}\right)\,. \tag{8.50}$$

We now assume a long trajectory. In the range where $x \ll N$, which is realized in an overwhelming fraction of all trajectories, the numerator and denominator logarithms in the last term on the right-hand side of Eq. (8.50) can be approximated by series expansion, $\ln(1+y) \approx y$ for $|y| \ll 1$, which gives

$$\ln\left(\frac{N/2+s}{N/2+1-s}\right) = \ln\left(\frac{1+2s/N}{1-2s/N+2/N}\right) \tag{8.51}$$

$$= \ln\left(1+\frac{2s}{N}\right) - \ln\left(1-\frac{2s}{N}+2/N\right) \tag{8.52}$$

$$\approx \frac{4s}{N} - \frac{2}{N}\,. \tag{8.53}$$

Hence,

$$\sum_{s=1}^{x/2} \ln\left(\frac{N/2+s}{N/2+1-s}\right) = \sum_{s=1}^{x/2}\left(\frac{4s}{N}-\frac{2}{N}\right) \tag{8.54}$$

$$= \frac{4}{N}\sum_{s=1}^{x/2} s - \frac{2}{N}\sum_{s=1}^{x/2} 1 \tag{8.55}$$

$$= \frac{4}{N} \cdot \frac{(x/2)(x/2+1)}{2} - \frac{x}{N} \tag{8.56}$$

$$= \frac{x^2}{2N}\,. \tag{8.57}$$

Inserting Eq. (8.57) into Eq. (8.50) provides,

$$P(N,x) \approx \frac{1}{2^N} \cdot \frac{N!}{(N/2)!(N/2)!} \cdot \exp\left(-\frac{x^2}{2N}\right)\,, \tag{8.58}$$

where we recognize, in the last factor on the right-hand side, the approximation of the binomial distribution by a Gaussian distribution that we discussed in Section 3.1.6. Using the improved formula of Stirling, Eq. (3.31), for expressing the factorials, we have

$$\frac{1}{2^N}\cdot\frac{N!}{(N/2)!(N/2)!}=\frac{1}{2^N}\frac{\sqrt{2\pi N}N^N\exp(-N)}{\left(\sqrt{\pi N}(N/2)^{N/2}\exp(-N/2)\right)^2}=\sqrt{\frac{2}{\pi N}}\,, \tag{8.59}$$

which leads to the exceedingly simple result:

$$P(N,x)=\sqrt{\frac{2}{\pi N}}\exp\left(-\frac{x^2}{2N}\right)\,. \tag{8.60}$$

The drunkard, if given enough time and not falling into sleep, perfectly simulates a Gaussian distribution.

We may even further simplify this result by asking about the mean square displacement $\langle x^2\rangle$, which is given by

$$\langle x^2\rangle=\int_{-\infty}^{\infty}x^2P(N,x)\mathrm{d}x=\sqrt{\frac{2}{\pi N}}\int_{-\infty}^{\infty}x^2\exp\left(-\frac{x^2}{2N}\right)\mathrm{d}x=N\,. \tag{8.61}$$

Before we go on, we need to fix a problem that occurs when we interpret the discrete probabilities computed at this point as a continuous probability density distribution of x. In the discrete case, $W(N,x)$ can be non-zero only for either even or odd x, depending on whether N is even or odd. Thus, to arrive at the proper probability distribution we need to divide by 2. Hence, we can express the probability density distribution for a one-dimensional random walk as

$$\rho_{\mathrm{1d}}(x)=\frac{1}{\sqrt{2\pi\langle x^2\rangle}}\exp\left(-\frac{x^2}{2\langle x^2\rangle}\right)\,. \tag{8.62}$$

This result does no longer depend on step size, not even implicitly, because we have removed the dependence on step number N. Therefore, it can be generalized to three dimensions. Since the random walks along the three pairwise orthogonal directions in Cartesian space are independent of each other, we have

$$\rho_{\mathrm{3d}}(x,y,z)\mathrm{d}x\mathrm{d}y\mathrm{d}z=\rho_{\mathrm{1d}}(x)\mathrm{d}x\cdot\rho_{\mathrm{1d}}(y)\mathrm{d}y\cdot\rho_{\mathrm{1d}}(z)\mathrm{d}z\,. \tag{8.63}$$

At this point we relate the result to the conformational ensemble of an ideal polymer chain, using the Kuhn model discussed in Section 8.2.1. We pose the question of the distribution of mean-square end-to-end distances $\left\langle\vec{R}^2\right\rangle$ with the Cartesian components of the end-to-end vector $\vec{R}$ being $x=R_x$, $y=R_y$, and $z=R_z$. According to Eq. (8.41), we have

$$\left\langle\vec{R}^2\right\rangle=\langle R_x^2\rangle+\langle R_y^2\rangle+\langle R_z^2\rangle \tag{8.64}$$

$$=Nb^2\,. \tag{8.65}$$

For symmetry reasons we have,

$$\langle R_x^2\rangle=\langle R_y^2\rangle=\langle R_z^2\rangle=\frac{Nb^2}{3}\,, \tag{8.66}$$

leading to

$$\rho_{\mathrm{1d}}(N,x)=\sqrt{\frac{3}{2\pi Nb^2}}\exp\left(-\frac{3R_x^2}{2Nb^2}\right) \tag{8.67}$$

and analogous expressions for $\rho_{1\mathrm{d}}(y)$ and $\rho_{1\mathrm{d}}(z)$. We have reintroduced parameter N, which is now the number of Kuhn segments. Yet, by discussing a continuous probability density distribution, we have removed dependence on a lattice model. This is necessary since the steps along dimensions x, y, and z differ for each Kuhn segment. By using Eq. (8.64), we find

$$\rho_{3\mathrm{d}}(N, \vec{R}) = \left(\frac{3}{2\pi N b^2}\right)^{3/2} \exp\left(-\frac{3\vec{R}^2}{2Nb^2}\right) . \tag{8.68}$$

The probability density attains a maximum at zero end-to-end vector.

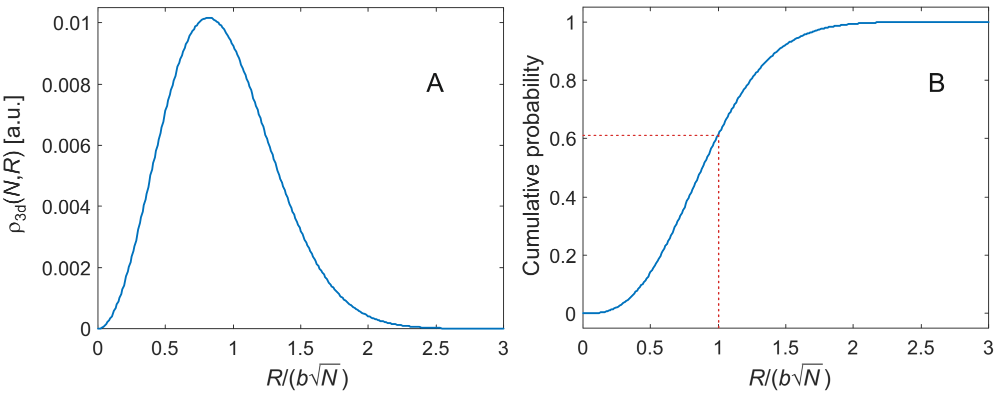


Figure 8.1: End-to-end distance distribution of an ideal chain. (A) Probability density distribution $\rho_{3\mathrm{d}}(N, R)$ of normalized en-to-end distances $R/(b\sqrt{N})$. (B) Cumulative probability distribution indicating the probability to find a chain with end-to-end-distance smaller than $R/(b\sqrt{B})$. Slightly more than 60% of all chains (red dashed horizontal line) have a normalized end-to-end distance $R/(b\sqrt{N}) \leq 1$.

Finally, we can pose the following question: If we let all chains of the ensemble start at the same point, how are the chain ends distributed in space? This is best pictured in a spherical coordinate system. Symmetry dictates that the distribution is uniform with respect to polar angles θ and ϕ. The polar coordinate R is equivalent to the end-to-end distance of the chain. To find the probability distribution for this end-to-end distance we need to include area $4\pi R^2$ of the spherical shells. Hence,

$$\rho_{3\mathrm{d}}(N, R) \cdot 4\pi R^2 \mathrm{d}R = 4\pi \left(\frac{3}{2\pi N b^2}\right)^{3/2} \exp\left(-\frac{3R^2}{2Nb^2}\right) R^2 \mathrm{d}R \, . \tag{8.69}$$

Because of this scaling with the volume of an infinitesimally thin spherical shell, the probability density distribution (Figure 8.1A) for the end-to-end distance does not peak at zero distance. As seen in Figure 8.1B it is very unlikely to encounter a chain with $R > 2b\sqrt{N}$. Since the contour length is $R_{\max} = Nb$, we can conclude that at equilibrium almost all chains have end-to-end distances shorter than $2R_{\max}/\sqrt{N}$.

We need to discuss validity of the result, because in approximating the discrete binomial distribution by a continuous Gaussian probability distribution we had made the assumption $x \ll N$. Within the ideal chain model, this assumption corresponds to an end-to-end distance that is much shorter than the contour length Nb. If R approaches Nb, the Gaussian distribution overestimates true probability density. In fact, the Gaussian distribution predicts a small, but finite probability for the chain to be longer than its contour length, which is unphysical. The model can be refined to include cases of such strong stretching of the chain [RC03]. For our qualitative discussion of entropic elasticity not too far from equilibrium, we can be content with Eq. (8.68).

8.2.3 Conformational Entropy and Free Energy

We may now ask the question of the dependence of free energy on chain extension $\vec{R}$. With the definition of Boltzmann entropy, Eq. (5.3), and the usual identification $k = k_\mathrm{B}$ we have

$$S(N, \vec{R}) = k_\mathrm{B} \ln \Omega(N, \vec{R}) \, . \tag{8.70}$$

The probability density distribution in Eq. (8.68) is related to the statistical weight Ω by

$$\rho_{3\mathrm{d}}(N, \vec{R}) = \frac{\Omega(N, \vec{R})}{\int \Omega(N, \vec{R}) \mathrm{d}\vec{R}} \, , \tag{8.71}$$

because $\rho_{3\mathrm{d}}$ is the fraction of all conformations that have an end-to-end vector in the infinitesimally small interval between $\vec{R}$ and $\vec{R} + \mathrm{d}\vec{R}$. Hence,[1]

$$S(N, \vec{R}) = k_\mathrm{B} \ln \rho_{3\mathrm{d}}(N, \vec{R}) + k_\mathrm{B} \ln \left[\int \Omega(N, \vec{R}) \mathrm{d}\vec{R} \right] \tag{8.72}$$

$$= -\frac{3}{2} k_\mathrm{B} \frac{\vec{R}^2}{N b^2} + \frac{3}{2} k_\mathrm{B} \ln \left(\frac{3}{2\pi N b^2} \right) + k_\mathrm{B} \ln \left[\int \Omega(N, \vec{R}) \mathrm{d}\vec{R} \right] \, . \tag{8.73}$$

The last two terms do not depend on $\vec{R}$ and thus constitute an entropy contribution $S(N, 0)$ that is the same for all end-to-end distances, but depends on the number of monomers N,

$$S(N, \vec{R}) = -\frac{3}{2} k_\mathrm{B} \frac{\vec{R}^2}{N b^2} + S(N, 0) \, . \tag{8.74}$$

Since by definition the Kuhn segments of an ideal chain do not interact with each other, the internal energy is independent of $\vec{R}$. The Helmholtz free energy $F(N, \vec{R}) = U(N, \vec{R}) - TS(N, \vec{R})$ can thus be written as

$$F(N, \vec{R}) = \frac{3}{2} k_\mathrm{B} T \frac{\vec{R}^2}{N b^2} + F(N, 0) \, . \tag{8.75}$$

It follows that the free energy of an individual chain attains a minimum at zero end-to-end vector, in agreement with our conclusion in Section 8.2.2 that the probability density is maximal for a zero end-to-end vector. At longer end-to-end vectors, chain entropy decreases quadratically with vector length. Hence, the chain can be considered as an *entropic spring*. Elongation of the spring corresponds to separating the chain ends by a distance $R \ll Nb$. The force required for this elongation is the derivative of Helmholtz free energy with respect to distance. For one dimension, we obtain

$$f_x = -\frac{\partial F\left(N, \vec{R}\right)}{\partial R_x} = -\frac{3 k_\mathrm{B} T}{N b^2} \cdot R_x \, . \tag{8.76}$$

For the three-dimensional case, the force is a vector that is linear in $\vec{R}$,

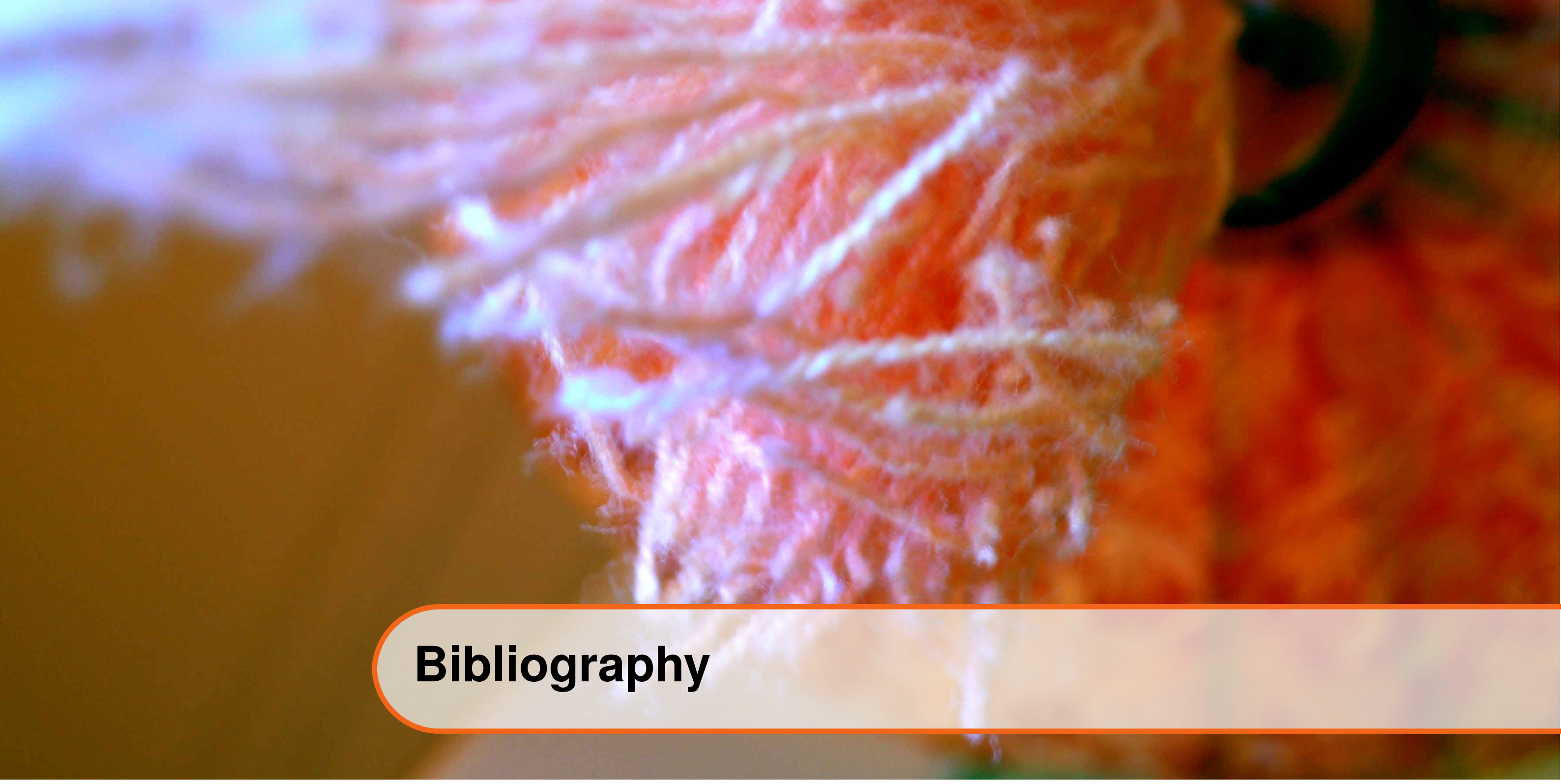
$$\vec{f} = -\frac{3 k_\mathrm{B} T}{N b^2} \cdot \vec{R} \, , \tag{8.77}$$

i.e., the entropic spring satisfies Hooke's law. The entropic spring constant is $3 k_\mathrm{B} T / (N b^2)$.

Polymers are thus the easier to stretch the larger their degree of polymerization (proportional to N), the longer the Kuhn segment b and the lower the temperature T. In particular, the temperature dependence is counterintuitive. A polymer chain under strain will contract if temperature is raised, since the entropic contribution to Helmholtz free energy, which counteracts the strain, then increases.

---

[1] This separation of the terms is mathematically somewhat awkward, since in the last two terms the argument of the logarithm has a unit. However, if the two terms are combined the logarithm of the unit cancels.



Bibliography

Books

[AP13] P. W. Atkins and J. de Paula. *Physikalische Chemie*. 5th Ed. Weinheim: WILEY-VCH, 2013 (cited on page 8).

[ER06] T. Engel and P. Reid. *Physikalische Chemie*. 1st Ed. München: Pearson Studium, 2006 (cited on page 8).

[Her45] G. Herzberg. *Molecular Spectra and Molecular Structure: II. Infrared and Raman Spectra of Polyatomic Molecules*. New York: van Nostrand, 1945 (cited on page 73).

[Mac98] A. Maczek. *Statistical Thermodynamics*. First. Oxford: Oxford University Press, 1998 (cited on pages 8, 66, 68, 74, 76).

[Pen70] O. Penrose. *Foundations of Statistical Mechanics*. Oxford: Pergamon Press, 1970 (cited on pages 8, 12, 15, 16).

[Pre+97] W. H. Press et al. *Numerical Recipes in C*. Second. Cambridge: Cambridge University Press, 1997 (cited on page 25).

[RC03] M. Rubinstein and R. H. Colby. *Polymer Physics*. First. Oxford: Oxford University Press, 2003 (cited on pages 81, 82, 85, 87, 90).

[SHM12] M. Schäfer, U. Hollenstein, and F. Merkt. *Advance Physical Chemistry. Statistical Thermodynamics.* Autumn Semester 2012. Zürich: ETH Zürich, 2012 (cited on page 76).

[Sch06] F. Schwabl. *Statistische Mechanik*. 3rd Ed. Berlin: Springer, 2006 (cited on pages 8, 14, 49, 54).

[SJ01] A. Schweiger and G. Jeschke. *Principles of Pulse Electron Paramagnetic Resonance*. Oxford: Oxford University Press, 2001 (cited on page 52).

[Swe12] R. H. Swendsen. *An Introduction to Statistical Mechanics and Thermodynamics*. 1st Ed. Oxford: Oxford University Press, 2012 (cited on pages 8, 18, 20, 22, 24, 25, 41, 52–54).

[WF12] G. Wedler and H.-J. Freund. *Lehrbuch der Physikalischen Chemie*. 6th Ed. Weinheim: WILEY-VCH, 2012 (cited on pages 8, 33, 37, 43, 44, 55–58, 62, 63, 67, 77, 79).

Articles

[Coh11] D. Cohen. “Lecture Notes in Statistical Mechanics and Mesoscopics”. In: *arXiv:1107.0568* (2011), pages 1–121 (cited on page 8).

[FR03] S. R. D. French and D. P. Rickles. “Understanding Permutation Symmetry”. In: *arXiv:quant-ph/0301020* (2003), pages 1– (cited on page 54).

[Kub63] R. Kubo. In: *J. Math. Phys.* 4 (1963), pages 174–183 (cited on page 49).

[Tan06] Y. Tanimura. In: *J. Phys. Soc. Jap.* 75 (2006), page 082001 (cited on page 49).

[VF74] A. J. Vega and D. Fiat. In: *J. Chem. Phys.* 60 (1974), pages 579–583 (cited on page 49).

Web Pages

[Iri98] K. K. Irikura. *Essential Statistical Thermodynamics*. `http://cccbdb.nist.gov/thermo.asp`. [Online; accessed 4-August-2015]. 1998 (cited on pages 9, 73).

[UNM12] UNM. *Laminar flow*. `https://www.youtube.com/watch?v=_dbnH-BBSNo`. [Online; accessed 28-July-2015]. 2012 (cited on page 47).

Index

A

accessible states 68

B

Bayes' theorem 18
binomial distribution 22
    Gaussian approximation 23
Boltzmann distribution 33, **36**, 54
Born-Oppenheimer approximation 66
Bose-Einstein statistics 55
boson 54

C

canonical ensemble 33
central limit theorem 23
central paradox 47
characteristic temperature
    electronic 76
    rotational 70
    vibrational 61
coherence 52
configuration 34
contour length 86
correlation function 21

D

degeneracy 58, 76
    degree of 58
density matrix
    equilibrium 53
    quantum mechanical 51
density of states 56, 63, 67
density operator 51
Dirac delta function 27
distinguishable variables 19
dynamical space 11
dynamical state 11

E

Einstein equation 75
ensemble 31
    canonical 33
    grand canonical 39
    microcanonical 32
entropy 15, 41, **45**
    Boltzmann 43
    Clausius 42
    Gibbs 43, **49**
    mixing 81
    Shannon 50
    von Neumann 49
equipartition theorem **37**, 63
ergodicity 31

F

Fermi-Dirac statistics 56
fermion 54
Flory characteristic ratio 86
Flory interaction parameter 84
Flory-Huggins equation 85
freely jointed chain 85

G

Gibbs paradox 46
grand canonical ensemble 39
ground state 54

H

Hamiltonian equations of motion 13
Heisenberg's uncertainty principle 60
Hess' law 78

I

ideal chain 85
independent events 21
independent variables 19

J

Jahn-Teller theorem 76

K

Kronecker delta 20
Kuhn length 86

L

Lagrange multipliers 35
lattice model 81
Liouville equation 14
  stochastic 49
Liouville theorem 14
Liouville-von-Neumann equation 15
Loschmidt paradox 47

M

macrostate 12
Markov chain 13, 87
Markovian postulate 13
Maxwell-Boltzmann statistics 57
mean field approach 62, **83**
mean value 20
microcanonical ensemble 32
microstate 11
mixed state 53
mixing entropy 81
moment analysis 25
moment of inertia 71

N

Newtonian equations of motion 11

O

observational state 12

P

para-hydrogen induced polarization 71
partition function
  canonical 36, 46
  grand canonical 39
  molecular 45
  quantum mechanical 54
  system 45
Pauli exclusion principle 54
Pauli principle 70
Penrose postulates 15
phase space 13, **13**
Poisson brackets 14
population
  quantum mechanical 52
probability 12
  conditional 18, 27
  density 25
  joint 18
  marginal 18, 26

probability distribution 20
  moments 21
pure state 53

Q

quantum number
  rotational 70, 71
  translational 67
  vibrational 60

R

random number
  continuous 25
  discrete 17
  independent 27
random walk 87
residual zero-point vibration 60
rotamers 71

S

Sackur-Tetrode equation 69
standard deviation 20
state space 13
statistical weight 32
statistics
  Bose-Einstein 55
  Fermi-Dirac 56
  Maxwell-Boltzmann 57
Stirling’s formula 24
  Gosper’s approximation 24
stochastic Liouville equation 49
symmetry number 73

T

trajectory 11

U

Umkehreinwand 47

V

variance 20

W

Wiederkehreinwand 48

Z

Zermelo pardox 48