

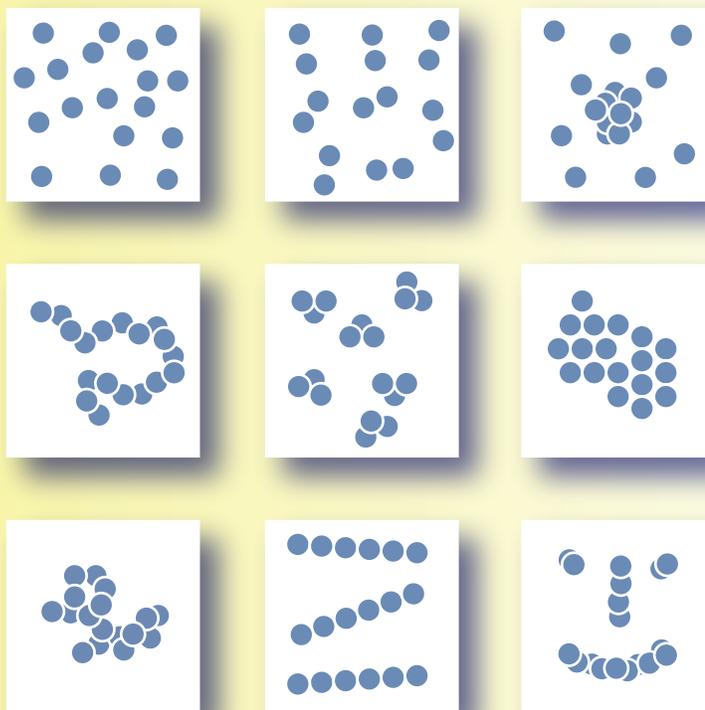
Themis Matsoukas

2019 Edition

CHE 524

*Lecture Notes on*

# Statistical Thermodynamics



1 1



[back side of cover page – leave blank]

LECTURE NOTES ON

# STATISTICAL THERMODYNAMICS

Themis Matsoukas

Department of Chemical Engineering  
Pennsylvania State University  
2019 Edition

Copyright © 2019 Themis Matsoukas

PUBLISHED BY ARACHTHOS PRESS



*This draft is dated 2019/08/23 at 08:29:00*

# Contents

<b>About this Book</b>	<b>i</b>
<b>1 The Fundamental Differential</b>	<b>1</b>
1.1 Scope and definitions	1
1.2 Internal energy	8
1.3 Entropy	9
1.4 The microcanonical differential	13
1.5 The calculation of entropy	15
1.6 Other inequalities involving entropy	16
1.7 Temperature, entropy and heat	19
1.8 Thermodynamic calculus	19
1.9 Additional reading	24
1.10 Problems	24
<b>2 Transformations of the Fundamental Differential</b>	<b>28</b>
2.1 The Legendre transformation	29
2.2 Partial molar properties	32
2.3 Jacobians	35
2.4 Applied thermodynamics: calculation of properties	37
2.5 Additional reading	44
2.6 Problems	44
<b>3 Equilibrium and Stability</b>	<b>51</b>
3.1 Conditions for equilibrium	52
3.2 Geometric properties of entropy	53
3.3 Equilibrium conditions under various constraints	54
3.4 Stability	58
3.5 Thermodynamics of phase equilibrium	59
3.6 Chemical potential and related properties	61
3.7 Additional reading	65
3.8 Problems	65
<b>4 Introduction to Stochastic Calculus</b>	<b>70</b>
4.1 Discrete frequency distributions	72
4.2 Continuous distributions	74
4.3 Some important distributions	76
4.4 Combinatorics	77
4.5 Central limit theorem	83

4.6	Stochastic process . . . . .	84
4.7	Monte Carlo sampling . . . . .	85
4.8	Additional reading . . . . .	90
4.9	Problems . . . . .	90
<b>5</b>	<b>Statistical Thermodynamics</b>	<b>94</b>
5.1	Microstates and ensembles . . . . .	95
5.2	Microcanonical ensemble . . . . .	98
5.3	Canonical ensemble . . . . .	104
5.4	Grand canonical ensemble . . . . .	117
5.5	Statistical Thermodynamics . . . . .	124
5.6	Connection to classical thermodynamics . . . . .	126
5.7	Fluctuations . . . . .	128
5.8	A combinatorial view of entropy . . . . .	131
5.9	A brief history of the entropy functional . . . . .	134
5.10	Additional reading . . . . .	135
5.11	Problems . . . . .	135
<b>6</b>	<b>Statistical Mechanics</b>	<b>140</b>
6.1	Newtonian mechanics . . . . .	140
6.2	Hamilton's equations . . . . .	141
6.3	Microstate and the continuous phase space . . . . .	143
6.4	Probability of microstate and Liouville's theorem . . . . .	144
6.5	Canonical distribution in continuous phase space . . . . .	147
6.6	Microcanonical ensemble . . . . .	148
6.7	Grand canonical ensemble . . . . .	151
6.8	The linear harmonic oscillator . . . . .	152
6.9	Thermodynamics and the equations of motion . . . . .	158
6.10	Additional reading . . . . .	162
6.11	Problems . . . . .	162
<b>7</b>	<b>Classical Molecular Systems</b>	<b>168</b>
7.1	Intermolecular potential . . . . .	169
7.2	Canonical partition function . . . . .	171
7.3	Configuration integral . . . . .	171
7.4	Molecular partition function . . . . .	173
7.5	Probability density of microstates . . . . .	175
7.6	Ideal gas . . . . .	177
7.7	Virial expansion . . . . .	181
7.8	van der Waals equation . . . . .	184
7.9	Molecular theory of corresponding states . . . . .	186
7.10	Additional reading . . . . .	188
7.11	Problems . . . . .	188
<b>8</b>	<b>Lattice Models</b>	<b>193</b>
8.1	Elementary lattice statistics . . . . .	193
8.2	Lattice gas of non-interacting particles . . . . .	194
8.3	Lattice gas with interactions . . . . .	196
8.4	Lattice Mixtures . . . . .	197

## Contents

---

8.5	Random solution	201
8.6	Regular solution	202
8.7	Quasichemical approximation (QCA)	204
8.8	Flory-Huggins theory for polymer solutions	208
8.9	Quasichemical treatment for molecules of dissimilar size	210
8.10	Additional reading	216
8.11	Problems	217
<b>9</b>	<b>Distribution Functions</b>	<b>222</b>
9.1	Probability of Configurational Microstate	222
9.2	Correlation functions	225
9.3	Radial distribution	227
9.4	Mechanical properties	227
9.5	Free energy and the coupling parameter	230
9.6	Additional reading	232
9.7	Problems	232
	<b>Bibliography</b>	<b>233</b>

## About this Book

This book is a compilation of my lecture notes for the first-year graduate course in chemical thermodynamics at Penn State, CHE 524. It represents an evolution over the years that has expanded to include material, commentary and problems with the intention to provide a compendium that is more or less self contained to serve as the main textbook for students taking the course but which can also serve as a resource for self study. The inevitable question is, *why another book on thermodynamics?* Although there is a large number of books on statistical mechanics, there is a dearth of books for chemical engineers, especially students at the beginning of their graduate career. Do chemical engineers need a *different* textbook? They do. The applications of thermodynamics are very diverse and cover an area of enormous expanse: classical molecules, quantum particles, polymers, chemical equilibrium, spectroscopy, cells . . . The standard pedagogy is to develop the principles of statistical thermodynamics alongside with the problems that thermodynamics is called to solve. Statistical thermodynamics is unfamiliar enough, students who approach it for the first time should not be burdened with physical details of unfamiliar fields. It only makes sense that the examples we draw ought to relate to the undergraduate background of the intended audience. For chemical engineers this background relies heavily on equations of state, property calculations and phase equilibrium. Van der Waals, UNIQUAC and Flory-Huggins are familiar terms –ferromagnets and quantum oscillators are not. The majority of textbooks are written with physicists and chemists in mind. Those aimed at engineers can be counted on the fingers of one hand.

The philosophy that guides this book and the approach I have chosen to fulfill it can be summarized as follows. The ultimate goal is to formulate statistical thermodynamics and to accomplish this we first need a firm mathematical understanding on classical thermodynamics. Chapters 1, 2 and 3 cover this material. This is not a rehash of the standard undergraduate fare –students are expected to go back their undergraduate texts to get up to speed with the background for advanced topics discussed here. It is a formulation of the network of Lagrange relationships among various properties that form the basis of the mathematical theory of equilibrium and stability. This effort is not wasted because these relationships will make a re-entry in the statistical development by an entirely different reasoning, where they will be recognized as familiar anchors to connect abstract statistical quantities to measurable properties.

Chapter 4 is a primer on probabilities and statistics. This material, which might more appropriately belong to an appendix, is important for the subsequent development and given that many students reach graduate school with

---

little or no familiarity with probabilities, especially continuous probability density functions, it deserves its place in the main text. This is also a place to introduce concepts of Metropolis sampling and Monte Carlo simulation of stochastic systems.

Chapter 5 presents an axiomatic development of ensemble thermodynamics based on the principle of *entropy maximization* and introduces the partition function and its properties. The formulation is generic and while energy plays a central role there is no explicit reference to the *physics* that give rise to this energy. To maintain clarity of concepts, the formulation remains entirely in the discrete domain with relatively simple examples of lattice systems, not necessarily realistic models of matter, that nonetheless demonstrate the emergence of the ideal-gas law and the entropy of mixing. This chapter makes passing references to Information Theory, if no other reason, to make interested students aware of connections and implications outside the realm of physical particles.

Chapter 6 draws the connection between statistical thermodynamics and statistical mechanics, using the Liouville equation as the starting point to demonstrate that the equations of motion, coupled with a statistical argument about our ignorance of the precise state, produces a direct connection to the formalism of the partition function in the previous chapter. This chapter takes the partition function into the continuous limit, and tackles some simple mechanical problems as a means of demonstrating the theory. This chapter takes a closer look at Gibbs's own approach and reasoning in developing statistical mechanics. A small number of problems at the end of the chapter are directly drawn from Gibbs's own derivations.

Chapter 7 formulates the statistical mechanics of classical particles interacting via a pairwise additive potential, and defines the configuration integral and the molecular partition functions. The topics covered come from the standard list of problems, ideal gas law, virial expansion, van der Waals equation. They are meant to provide a molecular interpretation to a number of familiar classical topics.

Chapter 8 discusses lattice models with special interest in problems in phase equilibrium. Topics include the Flory-Huggins model, Guggenheim's quasichemical approximation and the UNIQUA model of Prausnitz. The aim is to highlight lattice models that produce semi-quantitative agreement with experiment and especially to demonstrate the power of theory to guide the development of industrial-strength engineering tools.

Chapter 9 covers distribution functions, the pair correlation function and the calculation of properties by methods that are accessible to those who perform Molecular Dynamic Simulations. This is the most tentative chapter at the moment because by the time we reach this material the semester draws to an end.

## **Additional Reading**

Overall, the material covered is pretty much the syllabus of CHE 524, and this makes this book more tightly focused and less encyclopedic in the selection of topics than many other books. The references below cover many of the topics discussed in this book, in more detail.

- 
- T. L. Hill. *An Introduction to Statistical Thermodynamics*. Dover, New York, 1986. This inexpensive book is a reprint of the 1960 original. Its age notwithstanding, the book is a classic with respect to presenting the development of the core theory of statistical mechanics.
  - D. A. McQuarrie. *Statistical Mechanics*. University Science, Sausalito, CA, first edition, 2001. This book follows a theoretical approach similar to Hill with applications that have a chemistry focus. It is easier to read than Hill but derivations are not as detailed.
  - J. W. Tester and M. Modell. *Thermodynamics and Its Applications*. Prentice Hall, third edition, 1996. This book covers mostly classical thermodynamics and offers a rigorous and detailed mathematical treatment. A good reference for Legendre transformations and Jacobians.
  - J. M. Prausnitz, R. N. Lichtenhaler, and E. G. de Azevedo. *Molecular Thermodynamics of Fluid-Phase Equilibria*. Prentice Hall PTR, New Jersey, 3rd edition, 1999. The focus of the book is in the applications of thermodynamics to chemical engineering problems with lesser emphasis on derivations and basic theory. Its strong point is the chemical engineering focus unlike other books whose perspective is oriented towards either physics or chemistry. A good reference for lattice theories of solutions.
  - J. W. Gibbs. *Elementary Principles in Statistical Mechanics*. University Press, John Wilson and Son, Cambridge, USA, 1902. This is not an easy read, in part because the notation is very different than modern notation, and in part because its focus is on statistical *mechanics* (as opposed to thermodynamics) which requires a good background in physics. Nonetheless, anyone who dabbles in thermodynamics ought to at least flip through the pages of the book. A few problems in chapter 6 refer directly to Gibbs's book and the brave are given the challenge, with the guidance offered in the problem statements, to attempt the original.
  - E. A. Guggenheim. *Mixtures: The Theory of the Equilibrium Properties of Some Simple Classes of Mixtures Solutions and Alloys*. Oxford University Press, 1952. This is a monograph on lattice models of mixtures, and it is generally more accessible and easier to follow than one might think for a book of that nature.

Several other references are cited throughout these notes but these mentioned here are the ones more closely related to the lectures in this class.

In its present form this version of the Notes is still work in progress. Read *critically*, by which I mean “not passively.” Compare what you read with what we said in class, check the material for consistency with previous knowledge, and if you catch something that’s obviously wrong or doesn’t make sense, please [email me](#). Any typos or mistakes that are found during the semester will be posted with corrections on the web site of the course.

Themis Matsoukas  
University Park  
August 2019

## The Fundamental Differential

For historical reasons we distinguish thermodynamics into *classical* and *statistical*. Classical refers to the development that precedes the atomic theory of matter. It is based on the work of Carnot, Clapeyron, Lord Kelvin (William Thomson) and Clausius and represents a triumph of scientific reasoning, whose crown achievement is the formulation of the second law and the discovery of entropy. Statistical thermodynamics was formulated by J. Willard Gibbs.<sup>1</sup> It supersedes the classical treatment because not only does it reproduce all the relationships of classical thermodynamics, it connects macroscopic behavior (experimentally measured properties) to the microscopic level (molecular properties). Another way to organize thermodynamics is as “macroscopic” and “molecular.” By macroscopic we are referring to a mathematical development that reproduces the calculus of thermodynamics without specific reference to molecular or statistical concepts, based on reasoning that involves only our macroscopic experience with the behavior of matter. This is essentially the same as “classical,” except that we are not be obliged to follow the strict historical order, thus avoiding the detours and circuitous paths that marked the historical development of thermodynamics.

The statistical treatment is complete and self-contained. Why bother with macroscopic thermodynamics first? The reason is pedagogical. Thermodynamics is abstract and requires the skill to visualize mathematical concepts of fleeting physical interpretation –the reason being that the fundamental concepts are not quite “physical” but rather probabilistic. Statistical thermodynamics will make more sense if we know already what the results of the theory must look like. Since macroscopic thermodynamics lacks the connection to the molecular level, we will not be able to give convincing answers to the various “why’s” that we will encounter<sup>2</sup> – we will defer these discussion until the statistical part. Nonetheless, macroscopic thermodynamics provides a complete *phenomenology* of equilibrium thermodynamics that allows us to derive exact relationships among various thermodynamic quantities and describe the phase behavior of matter.

### 1.1 Scope and definitions

What is *thermodynamics*? When we ask this question we get answers that seem quite unrelated to each other. Here are a few.

<sup>1</sup>Between classical and statistical thermodynamic lies the kinetic theory of Boltzmann which attempts to arrive at thermodynamics based on a pure mechanical model of matter applied to dilute gases. Boltzmann's pioneering work led for the first time to the interpretation of entropy as a property of a *probability distribution*. Kinetic theory was superseded by Gibbs's statistical mechanics, which provides a universal theory for all matter, not only dilute gases.

<sup>2</sup>For example, why do all gases go over to the ideal gas law at low pressures? What gives rise to the the law of corresponding states? ...

*Thermodynamics is:*

- *the study of energy conversion between heat and work*
- *the study of the properties of matter as a function of temperature, pressure and composition*
- *a theory for the behavior of systems composed of an enormously large number of unobservable internal degrees of freedom*

These statements and many others are true on their face but merely state some particular *application* we have in mind. They do not tell us what what “thermodynamics is,” but rather what it “is *used* for.” Indeed the scope of thermodynamics is very broad and covers a wide range of subjects with apparently little in common, such as power plants and distillation columns, electrochemical cells and heat exchangers, foams and magnets. This enormity of scope is in part responsible for the confusion as to the true nature of thermodynamics, because a common thread among these areas is not immediately obvious. A common thread exists nonetheless. In all of these examples we are dealing with matter, more specifically, matter in *equilibrium*. Thermodynamics is the *mathematical calculus of the equilibrium state*. When we say thermodynamics we mean “equilibrium thermodynamics.” In actual processes such as power plants, distillation etc., where the working fluid is not in equilibrium, we apply thermodynamics on the *assumption* that the actual state is not very far from equilibrium. A general theory of non equilibrium thermodynamics is not available and current non equilibrium theories lack the rigor and universality of equilibrium thermodynamics.

We say that thermodynamics is the “mathematical calculus of the *equilibrium state*” and are careful not to say “equilibrium state of *matter*.” We will be applying thermodynamics to matter, of course, but the mathematical calculus is independent of the physical theories of matter. Thermodynamics survived the quantum revolution precisely because it is not dependent on classical or quantum assumptions about the nature of matter. This distinction is not well recognized in the literature in large part because the theory of thermodynamics was developed alongside with the theories of matter, first classical atomic in the late 1800’s, and then quantum in the early 1900’s. The input we require from *physics* to build the mathematical structure of thermodynamics is minimal and this is true both in classical and in statistical thermodynamics. There is a point of contact between *molecular* physics and thermodynamics and it is made in *statistical mechanics*. By being aware of which part is physics and which is math we will be in better position to understand thermodynamics and appreciate its universality.

We will begin our discussion with *macroscopic* thermodynamics, by which we mean that all physical input to the theory of thermodynamics will be via *macroscopic* observations of the behavior of matter. There will be no substantive reference to molecular except in qualitative ways to provide physical clarity to concepts such as the ideal gas state, which we describe as a system whose molecules are on average at distances much larger than the range of interactions. The precise microscopic connection will be given later when we develop *microscopic* thermodynamics in Chap. 5 and beyond. We can summarize the

## 1.1. Scope and definitions

required physical input in macroscopic thermodynamics in two fundamental empirical observations that form its foundational principles:

1. Matter has the capacity to store energy *internally* (first law).
2. The approach to equilibrium is *spontaneous*; a system cannot escape from a state of equilibrium without interference from the surroundings (second law).

In addition to these two principles we will be using the experimental finding that all gases at low pressure obey a universal equation of state regardless of chemical nature or composition:

$$\lim_{\substack{P \rightarrow 0 \\ T = \text{const.}}} \frac{Pv}{RT} = 1, \quad (1.1)$$

where  $P$  is pressure (Pa),  $v$  is molar volume ( $\text{m}^3/\text{mol}$ ),  $T$  is absolute temperature (K) and  $R = 8.314 \text{ J/mol K}$  is a universal constant. These principles form the basis of macroscopic thermodynamics. Our first goal is to express them in mathematical form as a single equation. Once we have this fundamental equation, all else in macroscopic thermodynamics is simply calculus. That's *good* news – it means that all thermodynamics can be developed using standard mathematical tools, without reference to *physics*, that is, without reference to physical models of matter. Macroscopic thermodynamics answers two questions: *How* are properties related to each other; and *what* are the conditions of *equilibrium*. Statistical mechanics will answer one additional question: *how* do we calculate the equilibrium properties from a microscopic model of matter.

Our goal in macroscopic thermodynamics is to *obtain the equilibrium properties of matter as a function of temperature, pressure and composition*. This statement contains two undefined terms, “equilibrium” and “temperature”. Clarifying these terms by providing definitions for them is part of the project we have undertaken. There is a logical inconsistency in this: we are starting by presuming concepts that will emerge later. A strictly logical development ought to make use only of concepts that have been defined already. In the literature of thermodynamics this is known as the *postulatory* approach. Those who subscribe to this view begin with a set of axiomatic statements that provide all necessary definitions for the subsequent developments. We will not follow this approach for two reasons. What is logically consistent is usually not quite intuitive, much less pedagogical. Logical consistency is an *emergent* quality, it develops over time as one gains more experience with the subject; it is not good pedagogy on which to base the exposition of a complex topic. The second and more serious reason is that the postulatory development implies, incorrectly, that thermodynamics is a *physical* theory that applies to physical systems only, specifically, physical systems composed of *interacting particles*, whether these are molecules, spins, photons, etc. Such narrow view obscures the nature of thermodynamics and leaves us puzzled over the breadth of areas that connect to the theory of thermodynamics.<sup>3</sup>

Bypassing the question of logical consistency our goal is to study the mathematical dependence of properties as a function of temperature, pressure and composition. Mathematically, we will be studying functions of the form<sup>4</sup>

<sup>3</sup>The tools of statistical mechanics are used today to study such diverse subjects as information theory, ecology, and financial markets. It would be absurd to suggest that a physical theory of matter has such wide reach. Only the language of *mathematics* has such universal scope.

<sup>4</sup>We will abbreviate this equation as

$$F = F(T, P, n_i),$$

where  $n_i$  stands for the complete set  $(n_1, n_2, \dots, n_K)$  of components present. The number of components  $K$  will not be specified – the system may contain any number of components.

## 1.1. Scope and definitions

$$F = F(T, P, n_1, n_2, \dots) \quad (1.2)$$

where  $F$  refers to some property, for example, volume,  $V$ , and  $n_i$  refers to the number of moles of component  $i$ . In writing such an equation we are limiting ourselves to systems whose properties are affected *only* by temperature, pressure and composition. There are several examples of systems whose properties depend on other variables as well. The behavior of magnetic systems depends, among other things, on the magnetic susceptibility of the material and the applied magnetic fields; the properties of a rubber band depend on its elastic modulus and the applied tension; and those of a small droplet depend on the exposed surface. It is by *choice* of the scope of our inquiry that we focus on those systems whose properties are completely determined by temperature, pressure and composition. The narrow justification for this choice is that systems of interest to chemical engineers are predominantly of this type. The broader reasoning is that the formalism of thermodynamics can be extended fairly easily to systems that are described by different sets of variables. Had we adopted a more mathematical approach, we would be speaking of  $F$  as a function of  $x, y, \dots$ , and we would leave it to the particular problem to assign common names to these variables, e.g., temperature, surface energy, and so on.

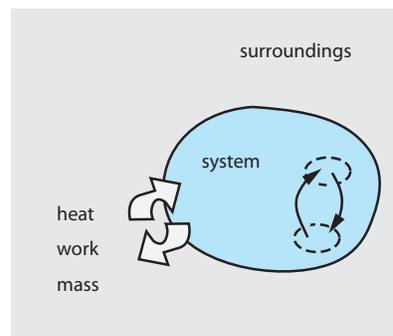
A mathematical equation of the form in Eq. (1.2) can be solved, say for  $P$ , to produce a function with independent variables  $F, T$  and  $n_i$ . In general, we may express a thermodynamic property in terms of number of other combinations of variables, for example, we may write

$$F = F(X, Y, n_i) \quad (1.3)$$

where  $X$  and  $Y$  may be other combinations of variables.<sup>5</sup> The set  $(T, P, n_i)$  is very convenient from a practical perspective since temperature, pressure and composition can be easily measured. They can just as easily be controlled in experiment or an industrial process. For this reason, the tabulation of properties is usually given in terms of  $(T, P, n_i)$ . From a theoretical perspective, Eq. (1.3) is a more general expression for the generic property  $F$ .

**System and surroundings** To connect thermodynamics to the physical world we introduce the “system” as an *amount of matter* whose properties we wish to study under changes of temperature, pressure, and composition. To impose such changes we must introduce interactions between the system and the outside world which we call *surroundings*. To change the temperature we may add or remove heat; to change pressure we may apply mechanical work to compress or expand; to change composition we may add or remove material. The exchange of heat, work and mass represent the only possible interactions between a system and its surroundings. These exchanges are under our control, we may construct systems that permit or forbid certain exchanges. With respect to the transfer of heat a system is classified as *adiabatic* (no heat transfer allowed) or *diathermal* (heat transfer allowed); with respect to the transfer of mass it is classified as *closed* (no mass transfer into or out of the system) or *open* (mass transfer is allowed). For example, a system surrounded by an impermeable wall cannot exchange mass and is closed; if the walls are movable, then the system can exchange work by expanding or contracting against the surroundings; and depending on whether

<sup>5</sup>Variables  $X$  and  $Y$  will normally be selected among the set  $T, P, V, S$ , though other choices are possible. We may also replace the  $n_i$  with other variables, however, it makes practical sense to retain the compositional dependence of properties in terms of the moles of each component.



**Figure 1.1:** The system interacts with the surroundings via the exchange of heat, mechanical work, and mass. In a simple system, any subregion can exchange heat, work and mass with any other subregion of the system.

the walls are conducting or not, the system will be diathermal or adiabatic. The system is *isolated* if no transfer of any kind is allowed. For the purposes of thermodynamic calculations, the system and its surroundings constitute the *universe*. A system is *simple* if it contains no internal walls or other barriers to the transport of heat, work, and mass. Any region of a simple system can exchange heat, work and mass with any other region of the system.

**Equilibrium** It is a common observation that when physical systems are left undisturbed they eventually reach a state of quiescence that we call *equilibrium*. *Undisturbed* means that there is no interference from the surroundings, i.e., there are no exchanges of any kind. This implies that the system is closed (no exchange of mass), is adiabatic (no exchange of heat) and its volume is constant (no  $PV$  work). Such a system is called *isolated* and is defined by the condition

$$(U, V, n_i) = \text{fixed},$$

which fixes the total energy, volume and moles in the system. The adiabatic condition is satisfied implicitly by the condition  $dQ = dU - dW$  since both  $dU$  and  $dW$  are zero.

A system with fixed energy, volume and number of particles can exist in various combinations of internal states. Suppose the system is a gas mixture of components A and B and it is prepared so that all A is in one half of the box and all B in the other half; or that half of the box is in higher pressure than the other half; or hotter than the other half, or any combination of internal partitioning. Such arbitrary preparations of the system are generally not stable and the system will be observed to transition through a series of internal states until it reaches one that is stable and time invariant. This ultimate state of equilibrium is independent of the initial preparation and depends solely on the total energy  $U$ , volume  $V$  and number of particles  $n_i$  of all species. We conclude that while a system of fixed  $(U, V, n_i)$  can exist in an infinite number of internal states, only one of these states has the property that is time invariant in the sense that once reached, the system never escapes from it without action from the surroundings. In other words *the equilibrium state is uniquely defined by  $(U, V, n_i)$* . We will say that the equilibrium state is known (or fixed) if  $U$ ,  $V$  and  $n_i$  are known. More colloquially we will refer to  $(U, V, n_i)$  simply as the state of the system with the understanding that we are referring to the *equilibrium* state.

The variable set  $(U, V, n_i)$  plays a fundamental role in thermodynamics. It is a set of mechanical variables and defines a system that is in no communication with its surroundings. In anticipation of the statistical treatment we will call this set *microcanonical*.

**Properties** The equilibrium state is uniquely determined by the microcanonical set. Anything we can measure about a system in equilibrium, e.g., density, volume, pressure, refractive index, etc., is a *property* whose value is uniquely determined by  $(U, V, n_i)$ . In thermodynamic language we say *properties are functions of state* and mathematically we write

$$F = F(U, V, n_i). \quad (1.4)$$

This implies that a function  $F(U, V, n_i)$  exists such that its value at given  $(U, V, n_i)$  is equal to the measured value of property  $F$ .

Properties are of two types: extensive and intensive. Extensive properties are proportional to the size (*extent*) of the system, which we take to be expressed by the total mass contained in the system, either in kg or in number of moles. The total volume  $V$  or the total number of moles  $n_i$  of component  $i$  are examples of extensive properties. Intensive properties are independent of the size of the system and for a single-phase system they have the same value everywhere inside the system. Pressure and temperature are intensive properties. The ratio of any two extensive properties is an intensive property. This follows from the fact that extensive properties are proportional to size, hence their ratio is not. Density, for example, is mass by volume, and is an intensive property. It is practical to express extensive properties in molar or specific form. A molar property is the ratio of an extensive property to the total number of moles; a specific property is the ratio to the total mass. For every extensive property  $F$  there is an intensive form  $f = F/n$  or  $f = F/m$ . We will use upper case symbols for the extensive variant and lower case for the intensive; for simplicity we use the same symbol for the molar and specific variants, but most of the time we will be working with the *molar* form.

Extensive properties are *additive*: the total property of a system is equal to the sum of the parts:

$$F(1,2) = F(1) + F(2).$$

Additivity applies whether 1 and 2 are internal parts of a system (for example, the left-half and right-half of a box) or two separate systems (two different boxes). This should not be confused with what happens we we *combine* two separate systems. For example, the total volume occupied by two separate liquids is the sum of the individual volumes. However, if the two liquids are mixed the total volume could be more or less than the sum of the parts, depending on how the materials interact. In general, if  $F(1+2)$  is the extensive property of the combined system, we have

$$F(1+2) \neq F(1) + F(2),$$

unless  $F$  is a conserved quantity (mass or energy).

**Quasistatic or reversible process** A thermodynamic *process* is action that results in change of the state of a system, schematically,

$$(U, V, n_i) \rightarrow (U', V', n'_i).$$

The notions of *equilibrium* and *process* are mutually exclusive: to cause an equilibrium system to change we must apply forces, forces create *gradients* and gradients produce non-equilibrium internal states. We overcome this problem by making use of an idealization that we call *quasistatic* or *reversible* process. We apply a differential change of state  $(\delta U, \delta V, \delta n_i)$  to an equilibrium state  $(U, V, n_i)$  in such way that the end state  $(U + dU, V + dV, n_i + dn_i)$  is an equilibrium state, such that all phases remain internally homogeneous. For example, if we transfer  $\delta n_i$  moles of species  $i$  into the system we assume that this amount is instantaneously distributed uniformly throughout the entire system. To implement

## 1.1. Scope and definitions

such process in the real world, all changes would have to be small and should be applied slowly in order to minimize deviations from equilibrium. We call this process *quasistatic*, or *reversible*. These two terms are equivalent but each refers to a different aspect of the process: Quasi-static means that the gradients required to produce the change of state are infinitesimal; reversible means that if a step is reversed, the state will reverse to its previous condition. The entire path then can be reversed back to the initial state by retracing its evolution backwards.<sup>6</sup> The idealization of the quasistatic/reversible process allows us to apply equilibrium thermodynamics not only to static equilibrium states, but also to *sequences* of equilibrium states, which we may view as a limiting form of a real physical process that can be conducted experimentally.

**The reversible process as an integration path:** The reversible process is a mathematical construction for the integration of thermodynamic differentials. We imagine the plane defined by the state variables  $(U, V, n_i)$  such that a point on that plane represents an equilibrium state. A continuous line between any two states  $(U, V, n_i)$  and  $(U', V', n'_i)$  represents a path along which we can compute the integral

$$\int_{U, V, n_i}^{U', V', n'_i} dF$$

for any property  $F$ . A path on the  $(U, V, n_i)$  plane is a sequence of equilibrium states that can be represented by a real process but only in a limiting sense: to produce a sequence of equilibrium states through a dynamic process as a function of time, care must be taken to upset equilibrium as little as possible. This generally means taking *very slow* steps, which is of course an impractical way to run a process. But we are *not* trying to run a process; we are merely *visualizing* a line on the thermodynamic plane as a thought experiment that *could* be conducted in the lab.

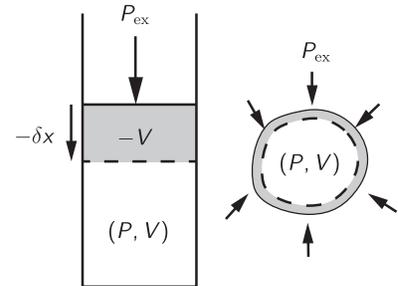
**Work** Exchange of energy is one of the two possible interactions between system and surroundings (exchange of mass is the other one). Energy may be exchanged in the form of mechanical work or heat. Work has the familiar definition from mechanics and is associated with the displacement of a force. In thermodynamic systems this is the force due to the external pressure and is exerted on the boundary between system and surroundings. This force moves and produces work when the boundary moves, as when the system expands or contracts (compression). We call this  $PV$  work because it is associated with pressure and volume. Other forms of mechanical work are possible, for example via moving shafts, springs or other mechanical devices. If these are present in a particular situation they must be included in the analysis. For our purposes, however, both “system” and surroundings will be “matter,” for example a fixed amount of matter enclosed via a boundary that allows the exchange of mass and energy. In such system the only possible form of mechanical work is  $PV$  work.

The  $PV$  work associated with a change  $dV$  in the volume of the system against external pressure  $P_{\text{ex}}$  is (see Fig. 1.2)

$$dW = -P_{\text{ex}}dV. \quad (1.5)$$

The sign is chosen by convention so that in compression ( $dV < 0$ ) the work is counted as positive (enters the system) and in expansion ( $dV > 0$ ) as negative

<sup>6</sup>Some processes are clearly irreversible. If we puncture a balloon its contents escape. This only takes a pinch with a needle. We cannot reverse the process by “un-puncturing” the balloon. In other processes the distinction is less stark. We may expand a gas in a piston-and-cylinder arrangement from initial state  $(P_1, T_1)$  to final state  $(P_2, T_2)$ , and then reverse the process until we are back to initial pressure  $P_1$ . If we fail to do this in a quasistatic manner we will discover that temperature at the end state is *not* the initial temperature but a bit higher. In this case irreversibility is demonstrated by the inability of the system to reach the initial state upon reversing all actions (i.e., switching heating to cooling, expansion to compression etc.). Quasistatic conditions ensure that the process is reversible. The two terms are equivalent but we prefer to use “reversible” to refer to this special process.



**Figure 1.2:** The work associated with a change of volume by  $dV$  under external pressure  $P_{\text{ex}}$  is  $(P_{\text{ex}}A)(-dx) = -P_{\text{ex}}dV$ , where  $A$  is the area exposed to pressure  $P_{\text{ext}}$ . The result is independent of the geometry of the system.

## 1.2. Internal energy

because it is done by the system on the surroundings.<sup>7</sup>

Under quasistatic conditions the external pressure is infinitesimally different from the pressure of the system,

$$P_{\text{ex}} = P + \delta P$$

with  $\delta P > 0$  for compression and  $\delta P < 0$  for expansion. Letting  $\delta P \rightarrow 0$  we have  $P_{\text{ex}} \rightarrow P$  and the work in Eq. (1.5) becomes

$$dW_{\text{rev}} = -PdV. \quad (1.6)$$

The subscript <sub>rev</sub> is a reminder that this expression for the  $PV$  work is valid only for reversible process.<sup>8</sup> Equation (1.6) allows us to express the reversible work entirely in terms of properties of the *system*, its volume and pressure.

**Heat** Heat is the form of energy that is exchanged between two systems held at different temperatures and always passes from higher to the lower temperature. As a non mechanical quantity, heat cannot be defined in a simple way as work. In fact the concepts of *heat* and *temperature* are both related to entropy and we cannot define them until we have a definition of entropy. Until then we assume the intuitive meaning of “temperature” and “heat.” Heat and work share common characteristics: they both represent transfers between a system and its surroundings and both result in a change of state. the sign convention for heat is the same as for work: positive if it enters the system, negative if it exits to the surroundings.

## 1.2 Internal energy

If a system is capable of exchanging energy with its surroundings, whether in the form of mechanical work, heat or any other form, it must be capable of *storing* energy. This is required by the principle of energy conservation. We call the energy that is stored within matter *internal energy*,  $U$ . If a system of fixed mass receives an amount of work  $dW$  and heat  $dQ$  during a differential process, internal energy must increase by the sum of these inputs:

$$dU = dQ + dW, \quad (\text{constant } n_i). \quad (1.7)$$

This is the mathematical statement of the first law and defines internal energy, an “invisible” form of energy, in terms of measurable quantities, mechanical work and heat. In writing this form we tacitly assume that no other forms of energy are exchanged between the system and its surroundings. If the entire system moves inside an external field then its bulk kinetic and potential energy as well as any other energy inputs (electric, shaft work etc.) should be included in the balance. We will restrict our attention to situations where heat and  $PV$  work are the only energy exchanges with the surroundings. As a statement of energy conservation, Eq. (1.7) is general and applies to any change of state, i.e., any process, reversible or not. In the special case that the process is reversible the  $PV$  work is given by Eq. (1.6) and the differential of internal energy becomes

$$dU = dQ - PdV, \quad (\text{constant } n_i). \quad (1.8)$$

<sup>7</sup>For both heat and work the sign convention is the same: positive means that the transfer is to the system from the surroundings; negative means that the transfer is out of the system to the surroundings.

<sup>8</sup>The distinction between reversible and non-reversible work is important only when we apply thermodynamics to real processes where  $W$  and  $W_{\text{rev}}$  may be different. In these notes we will be dealing almost exclusively with reversible paths. Still, to avoid confusion we will use  $W_{\text{rev}}$  to indicate beyond doubt that the path is reversible.

This is a more restrictive form of the first law because it is limited to a mechanically reversible processes.

**Microscopic interpretation of internal energy** Within the context of macroscopic thermodynamics the existence of internal energy is a consequence of energy conservation. The name refers collectively to all microscopic forms in which matter can store energy. Generally speaking this energy is of two types, kinetic and potential. Kinetic refers to the motion of all movable constituents of matter, including center-of-mass motion, rotation, vibration and so on. Potential refers to all types of interactions between the movable constituents and includes molecular interactions, chemical bonds, as well as electronic and nuclear effects. In the macroscopic treatment there is no need to identify these forms of energy individually and instead we lump them together under the term internal energy.

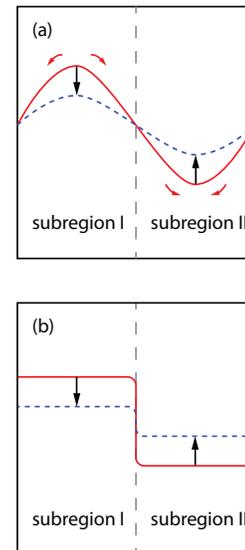
### 1.3 Entropy

The second law articulates the spontaneous tendency of all systems to reach equilibrium. Examples abound: if we mix cold and warm water the final state reaches a temperature in between; if we open the valve of a tank with pressurized gas, the gas will escape until the pressure in the tank matches the ambient pressure; if we place a drop of ink in water, eventually it will disperse throughout the entire liquid. These processes are spontaneous, which is to say, they do not require intervention from the outside (to perform these experiments in a controlled manner we should place the system inside a rigid box and insulated it from the surroundings to prevent any exchange with the outside work). Spontaneity is unidirectional: if the process  $A \rightarrow B$  is spontaneous, the process  $B \rightarrow A$  is *not*. To go from  $B$  to  $A$  we must intervene and expend some energy, e.g., put the molten ice cube in the freezer and warm up the rest of the water a bit; pump gas into the empty tank and seal it; extract the ink from the water.<sup>9</sup> The approach to equilibrium is mediated by kinetic processes that set up transfers which in turn redistribute the extensive quantities within the system. The nature of the physical laws that govern rate processes is such that transfer always takes place in the direction of decreasing gradients. By doing so the gradient becomes more shallow. This sets up a process by which gradients decrease as a result of transfer until gradients are erased and no further transfer takes place. The ultimate state that is reached is stationary (invariant in time) and is the what we call equilibrium state. We may view a non-equilibrium system as a system with internal gradients, and the approach to equilibrium as a process that eliminates of gradients. We introduce entropy as a measure of dissipation<sup>10</sup> of gradients.

A system far from equilibrium is characterized by gradients as schematically shown in Fig. (1.3). A gradient is a non-uniform spacial distribution of an intensive property, for example density, pressure, concentration, etc. We are primarily concerned with gradients associated with the transport of heat,  $PV$  work, and mass, which comprise the possible exchanges between any two systems. Heat, for example, is transferred under a gradient of temperature from high temperature to low. As a result of this transfer the gradient decreases and is eventually eliminated completely. The gradient in Fig. (1.3a) cannot be analyzed with the tools of equilibrium thermodynamics. We work instead with the idealization in Fig. (1.3b). The gradient is represented as two non-interacting subsystems,

<sup>9</sup>The fact that a particular process is not spontaneous does not mean it cannot take place; it means it cannot take place without outside intervention.

<sup>10</sup>By *dissipation* we mean the gradual elimination of gradients.



**Figure 1.3:** Real (top) and idealized (bottom) view of gradients in approach to equilibrium.

each in *internal* equilibrium, but *not* in equilibrium with each other. A physical partition would be necessary to maintain the gradient between two subregions. The entire system is in equilibrium if –and only if– the removal of the partition causes no macroscopic change in any of the subregions, i.e., if and only if the partition is unnecessary. This idealization allows us to apply equilibrium thermodynamics within each subregion, while allowing for the presence of a finite gradient between them. This idealized picture is not a realistic model for the transient behavior of the system as it approaches equilibrium, but this is not what we are after. We are after the condition of equilibrium and this is defined rigorously, uniquely and unambiguously but the requirement that the presence of the physical partition become unnecessary.

**Microcanonical partitioning & thermodynamics** The partitioning of an isolated system into a set of noninteracting subsystems is an important thought experiment in the formulation of equilibrium thermodynamics (Fig. 1.3b). *Partitioning* refers to the splitting of a system  $(U, V, n_i)$  into  $\mathcal{K}$  parts that do not interact, each with equilibrium state  $(U^{(k)}, V^{(k)}, n_i^{(k)})$ ,  $k = 1, \dots, \mathcal{K}$ . This amounts to partitioning the energy, volume and number of particles of the system such that their sum remains fixed and equal to that of the total.

$$U = \sum_k U^{(k)}, \quad V = \sum_k V^{(k)}, \quad n_i = \sum_k n_i^{(k)}.$$

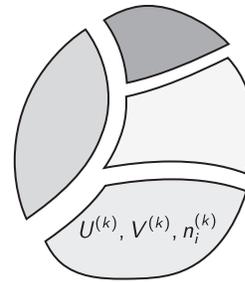
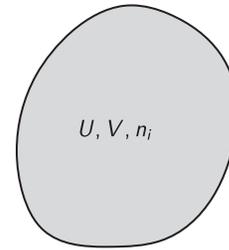
Because the parts do not interact, any extensive property of the the partitioned state is the sum of the parts:

$$F\{\mathcal{K}\} = F^{(1)} + F^{(2)} \dots + F^{(\mathcal{K})}. \quad (1.9)$$

Here,  $F\{\mathcal{K}\}$  refers to the total property of the partitioned system (not to be confused with  $F$ , which is the total property of the overall system with the partitions removed).

We may implement partitioning as follows: first place rigid adiabatic walls to subdivide the system into  $\mathcal{K}$  subsystems of arbitrary volume, then exchange arbitrary amounts of mass energy between parts to create the final partitioning. If upon removal of the partitions the macroscopic state is observed to evolve, the partitioning is a non equilibrium state. The partitioning is an equilibrium state only if the removal of the partitions leaves all subsystems unchanged. This serves as a simple test of equilibrium that can be expressed in precise mathematical form. The number of subsystems does not matter and for simplicity we will be using only two parts.

A non-equilibrium partition is an idealized model for a non-equilibrium state, namely, a state that is characterized internally by inhomogeneous distributions and gradients. It is idealized because the compartments that constitute the larger system do not interact. Even though gradients exist across the boundaries that separates the subsystems, there are no transfers of any kind, either within subsystems or across their boundaries. This idealization restricts the use of the theory to *equilibrium states*. This restriction is not as bad as it seems though. In a real system the removal of the boundaries between the parts of the system sets up an irreversible process that ends with the establishment of a new equilibrium state in the entire system. The partitioned state represents the initial condition and the non partitioned state represents the final state. Equilibrium thermodynamics applies to both; it does *not* apply to the transient states that are observed while the system is transitioning from the initial state to the final. When we integrate thermodynamic differentials we always integrate along a *reversible path*. This path does not match the actual transient between the end states but as long as the end points represent equilibrium states the result of the integration is correct.



$$\begin{aligned} \sum_k U^{(k)} &= U \\ \sum_k V^{(k)} &= V \\ \sum_k n_i^{(k)} &= n_i \end{aligned}$$

**Figure 1.4:** In microcanonical partitioning the energy, volume and mass of the system is distributed over an arbitrary number of non-interacting parts.

**Entropy and the second law** The mathematical problem ahead of us is the formulation of a quantitative criterion that distinguishes between equilibrium and non-equilibrium states and expresses the unidirectional approach towards the equilibrium state. The model we use to develop this criterion is the partitioning of a microcanonical system  $(U, V, n_i)$  into two parts,  $(U^{(1)}, V^{(1)}, n_i^{(1)})$  and  $(U^{(2)}, V^{(2)}, n_i^{(2)})$ . If the removal of the partition causes the states of the two parts to transition, the partitioning represents a non-equilibrium state. We introduce *entropy*,  $S(U, V, n_i)$ , as a property of microcanonical state in equilibrium that has the following properties:

1. It is homogeneous function of its variables with degree 1:

$$S(\lambda U, \lambda V, \lambda n_i) = \lambda S(U, V, n_i). \quad (1.10)$$

2. Given system 1 at  $(U^{(1)}, V^{(1)}, n_i^{(1)})$  and system 2 at  $(U^{(2)}, V^{(2)}, n_i^{(2)})$  it satisfies the inequality

$$S(1+2) \geq S(1) + S(2) \quad (1.11)$$

where 1 + 2 refers to the sum-state

$$(U^{(1)} + U^{(2)}, V^{(1)} + V^{(2)} + n_i^{(1)} + n_i^{(2)}).$$

The equality applies if and only if the two systems are in equilibrium with respect to each other.

Equation (1.10) states that entropy is extensive. As such, the total entropy of a collection of systems is the sum of the entropies of the individual systems. The inequality in Eq. (1.11) states that given a partitioning of a microcanonical system, the entropy of the combined system, i.e., of the system with the partitions removed, is higher than or equal to the entropy of the partitioned system. This provides us with a criterion to distinguish equilibrium from non-equilibrium states: if the partitioned system has lower entropy than the combined system, then it is a non-equilibrium partition; if the partitioned system has the same entropy as the combined system then and only then it represents an equilibrium partitioning.

We consider Eqs. (1.10) and (1.11) as axiomatic definitions of entropy and we will show later that that the definition is fully consistent with the behavior of physical systems. Nonetheless it is helpful to provide an argument to motivate (not prove) the conditions that define entropy. Directionality cannot be expressed via an equality between the “before” and “after” states because equalities treat both directions (from the left-hand side to the right-hand side and vice versa) in an equivalent manner; an inequality is needed to distinguish the direction of a change of state. To calculate this postulated property in each compartment, it must be a function of  $U, V, n_i$ , it must be then a *state* function. It must also be extensive, because all regions of the system must contribute into determining whether a particular transfer is in the direction of equilibrium or not. These arguments do not constitute a proof and we must admit that entropy as defined here is not intuitive.<sup>11</sup> It will make more sense as we apply it to problems and find out that it produces results in full agreement with observation. Ultimately, the proper way to think about entropy is probabilistic, but for this we will have to wait until chapter 5.

<sup>11</sup>The mathematical statement of the second law in classical thermodynamics is a triumph of inductive reasoning. A universal principle that eludes intuition and direct measurement *must* nonetheless *be true*, if everything else we can observe and measure around it is to be believed.

**Entropy and the approach to equilibrium** The inequality in Eq. (1.11) is between an arbitrary partitioning of the microcanonical system (right-hand side) and the equilibrium state (left-hand side). If we remove the partitions the system will transition to its equilibrium state. For this process we have

$$\Delta S = S_{\text{equ}} - S \geq 0, \quad (1.12)$$

where  $S$  refers to the entropy of the partitioned system and  $S_{\text{equ}}$  to the entropy of the final equilibrium state. This process can be conducted experimentally: prepare the two parts inside rigid boxes so that they are each in internal equilibrium, then remove the boundary that separates them and allow the system to reach equilibrium in isolation from their surroundings while keeping the total volume constant. During the transition the system is characterized by internal gradients that set up transfers of heat due to temperature differences, convective motion due pressure gradients and mass transfer by molecular diffusion. These are non-equilibrium states in dynamic evolution for which entropy is undefined. Nonetheless the inequality in Eq. (1.12) holds because it involves equilibrium states. This highlights the scope and limitation of equilibrium thermodynamics:

- we *can* calculate the net change in entropy between the end states of an irreversible transition as long as these states are in equilibrium;
- we *cannot* calculate the entropy as a function of time during a real irreversible transition.<sup>12</sup>

The inequality in Eq. (1.12) is between an arbitrary partition of the microcanonical system and the final equilibrium state, and states that among all microcanonical partitions the equilibrium state has the maximum entropy. We may use this inequality to compare two non equilibrium partitions. If  $A$  and  $B$  are two different partitions of the same microcanonical system  $(U, V, n_i)$ , the one with the higher entropy is closer to the equilibrium state; the one with the *highest* entropy *is* the equilibrium state. We can now construct an entire series of partitions and order them in increasing entropy. Along this path

$$dS \Big|_{U, V, n_i} \geq 0, \quad (1.13)$$

where  $dS$  is the entropy difference between adjacent partitions. When this inequality reduces to an equality we have reached the maximum entropy, i.e., we have identified the equilibrium state. Equation (1.13) is the differential equivalent of (1.11). It applies to a continuous path of internally equilibrated partitions that ends at the equilibrium state.

**Entropy and real processes** The differential form in Eq. (1.13) might give the wrong impression that it is unconditionally true for any real system approaching equilibrium under isolation. Entropy is defined on equilibrium states only. Non equilibrium states can be treated only in the idealized form of Fig. 1.3 as a set of non interacting subsystems in internal equilibrium. For small gradients this idealization is an acceptable approximation but it breaks down when gradients are strong. Therefore the application of the second law to the path of a real process must be done cautiously and with the understanding that it represents an approximation.

<sup>12</sup>To perform this calculation we must make assumptions about the *kinetics* of the transition. Such assumptions are external to equilibrium thermodynamics and represent an approximation, not a rigorous calculation.

**Reversibility and the quasistatic process** A practical reading of Eq. (1.13) is that no process can decrease the entropy of an isolated system.<sup>13</sup> This is not a controversial statement as some have claimed: the only possible process in a system that is isolated from its surroundings is a *spontaneous process*—any other process requires intervention. Entropy can only increase because a spontaneous process is always in the direction of equilibrium.<sup>14</sup> The equal sign applies to equilibrium. It also applies to another special case, the quasistatic process. We defined *quasistatic* as a process that proceeds in the absence of gradients, i.e., by maintaining equilibrium throughout the process. Along this path Eq. (1.13) applies as an exact equality. A process that is conducted such that the total entropy of an isolated system is constant is called *reversible* to emphasize the fact that the system is always at equilibrium such that if the steps of the process are reversed the system will retrace the same path in the reverse direction. The terms *quasistatic process* and *reversible process* are equivalent but not identical. *Quasistatic* refers to the mode of operation, specifically, to the absence of gradients during the process. Reversible refers to the condition  $dS = 0$ . A quasistatic process is always reversible; a reversible process can only be implemented in a quasistatic manner. We will be using the two terms interchangeably but we should remember that each refers to a different aspect of a process.

<sup>14</sup>Think of entropy as a counter that counts the number of steps the universe takes during a process: this number can only go up; if it stops changing, it means that the equilibrium state has been reached.

#### 1.4 The microcanonical differential

We are now ready to develop a mathematical differential that consists entirely of thermodynamic quantities. We begin with the fact that as a state function  $S$  is a function of  $U$ ,  $V$  and  $n_i$ . We express its differential as

$$dS = \left( \frac{\partial S}{\partial U} \right)_{V, n_i} dU + \left( \frac{\partial S}{\partial V} \right)_{U, n_i} dV + \sum \left( \frac{\partial S}{\partial n_i} \right)_{U, V, n_{j \neq i}} dn_i. \quad (1.14)$$

This can be expressed in the equivalent form:<sup>15</sup>

$$dU = \left( \frac{\partial U}{\partial S} \right)_{V, n_i} dS + \left( \frac{\partial U}{\partial V} \right)_{S, n_i} dV + \sum_i \left( \frac{\partial U}{\partial n_i} \right)_{S, V, n_{j \neq i}} dn_i. \quad (1.15)$$

For reversible adiabatic process in closed system the first law says,

$$dU = -PdV \Big|_{\text{closed \& reversible adiabatic}}. \quad (1.16)$$

This states that the differential of internal energy along a path specified by the conditions “closed” ( $dn_i = 0$ ) and “reversible adiabatic” depends only on  $dV$ . We apply the same conditions to Eq. (1.15). All terms involving  $dn_i$  drop out; for the rest of the result to depend only  $dV$  we must have  $dS = 0$ . That is, the condition *closed system + reversible adiabatic* is equivalent to *isentropic*. We then have

$$dU = -PdV \Big|_{S, n_i}. \quad (1.17)$$

This is the differential of  $U(S, V, n_i)$  at fixed  $S$  and  $n_i$ . Then we identify  $-P$  as the derivative of internal energy with respect to volume at fixed  $S$ ,  $n_i$ :

$$P = - \left( \frac{\partial U}{\partial V} \right)_{S, n_i}. \quad (1.18)$$

<sup>15</sup>We have used

$$\left( \frac{\partial U}{\partial S} \right)_{V, n_i} \left( \frac{\partial S}{\partial V} \right)_{U, n_i} = \left( \frac{\partial U}{\partial V} \right)_{n_i},$$

and

$$\left( \frac{\partial U}{\partial S} \right)_{V, n_i} \left( \frac{\partial S}{\partial n_i} \right)_{U, V, n_{j \neq i}} = \left( \frac{\partial U}{\partial n_i} \right)_{V, n_{j \neq i}},$$

both of which are obtained by application of the chain rule. Alternatively, Eq. (1.15) may be written immediately as the differential of  $U(S, V, n_i)$ .

#### 1.4. The microcanonical differential

---

For the general case of reversible but not necessarily adiabatic process in closed system, the first law gives

$$dU = dQ_{\text{rev}} - PdV \quad (\text{const. } n_i). \quad (1.19)$$

By comparison with Eq. (1.15) and taking into consideration (1.18) we make the identification

$$dQ_{\text{rev}} = \left( \frac{\partial U}{\partial S} \right)_{V, n_i} dS. \quad (1.20)$$

We now *define* absolute temperature as the derivative of internal energy with respect to entropy at fixed  $V$  and  $n_i$ :

$$T = \left( \frac{\partial U}{\partial S} \right)_{V, n_i}. \quad (1.21)$$

Using this definition, Eq. (1.20) gives,

$$dS = \frac{dQ_{\text{rev}}}{T}. \quad (1.22)$$

The last result relates entropy to heat and gives the basic recipe to calculate entropy changes in closed system: connect the end states with a reversible path and integrate the quantity  $dQ_{\text{rev}}/T$ . Since entropy is a state function, *any* reversible path is acceptable and must yield the same answer.

We introduce one last definition, the *chemical potential*  $\mu_i$  of component  $i$ , as the partial derivative of internal energy with respect to  $n_i$  with all other microcanonical variables fixed:

$$\mu_i = \left( \frac{\partial U}{\partial n_i} \right)_{S, V, n_{j \neq i}}. \quad (1.23)$$

Using these definitions, the differential of entropy can be written in the more compact form,

$$dU = TdS - PdV + \sum_i \mu_i dn_i. \quad (1.24)$$

Solving this for  $dS$  we obtain

$$dS = \frac{dU}{T} + \frac{PdV}{T} - \sum_i \frac{\mu_i dn_i}{T}. \quad (1.25)$$

Equations (1.24) and (1.25) are alternative expressions of the fundamental differential of the thermodynamics and expresses the relationship between the variables,  $U$ ,  $S$ ,  $V$ , and  $n_i$ . Implicitly, these differentials give the corresponding partial derivatives of  $U(S, V, n_i)$  and  $S(U, V, n_i)$ . The derivatives of  $U$  are given in Eqs. (1.18), (1.21) and (1.23). The derivatives of  $S$  follow directly from Eq. (1.25):

$$\frac{1}{T} = \left( \frac{\partial S}{\partial U} \right)_{V, n_i}, \quad \frac{P}{T} = \left( \frac{\partial S}{\partial V} \right)_{U, n_i}, \quad \frac{\mu}{T} = - \left( \frac{\partial S}{\partial n_i} \right)_{U, n_j}.$$

## 1.5. The calculation of entropy

Both differentials have the same structure: the independent variables are extensive and all derivatives are intensive. By Euler integration of Eq. (1.24) we obtain

$$U = TS - PV + \sum n_i \mu_i. \quad (1.26)$$

Similarly, Euler integration of Eq. (1.25) gives

$$S = \frac{U}{T} + \frac{PV}{T} - \frac{n_i \mu_i}{T}, \quad (1.27)$$

a result that may also be obtained by solving Eq. (1.26) for  $S$ . While the integral and differential forms of  $U$  and  $S$  are mathematically equivalent, it is the differential form that will serve as the starting point of most derivations because the differentials govern the evolution of thermodynamic properties on quasistatic transitions between equilibrium states.

### 1.5 The calculation of entropy

Equation (1.22) relates entropy to heat and provides a direct connection between this postulated property and measurable quantities. The general methodology for the calculation of entropy changes between two states is to connect the end states with a reversible path and perform the integration of the quantity  $dQ_{\text{rev}}/T$ :

$$\Delta S_{12} = \int_1^2 \frac{dQ_{\text{rev}}}{T}. \quad (1.28)$$

Since entropy is a state function the result of this integration is independent of the chosen path. Any reversible path will yield the same answer.

Solving Eq. (1.22) for  $Q_{\text{rev}}$  we obtain

$$dQ_{\text{rev}} = TdS; \quad \text{and} \quad Q_{\text{rev}} = \int_1^2 TdS. \quad (1.29)$$

These equations for heat are entirely analogous to those for reversible work,

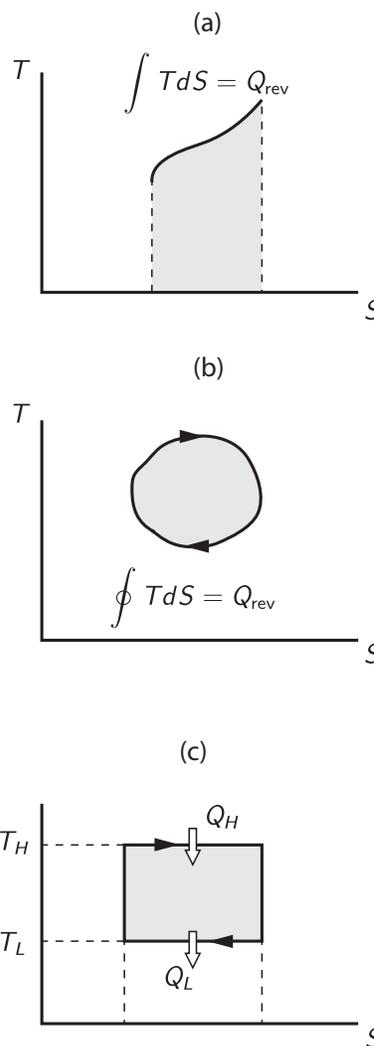
$$dW_{\text{rev}} = -PdV; \quad \text{and} \quad W_{\text{rev}} = - \int_1^2 PdV, \quad [??]$$

and demonstrate that heat, just like work, is a *path* function (Fig. 1.5a).

Consider a reversible cycle on the  $TS$  graph (Fig. 1.5b). The change in internal energy is  $\Delta U = 0$  and the first law gives

$$\Delta U = Q_{\text{rev}} + W_{\text{rev}} = \oint TdS - W_{\text{rev}} = 0.$$

The integral of  $TdS$  over a closed path represents the area enclosed by the path. If the process moves in clockwise direction, this area is positive, i.e., the cycle absorbs a net amount of heat and produces an equivalent amount of work. If the process tracks the path in counter clockwise direction, the cycle absorbs a net amount of work and rejects a net amount of heat to a higher temperature, i.e., the cycle operates as a heat pump. These effects can be followed more easily on the Carnot cycle whose path on the  $TS$  graph is a rectangle. This cycle



**Figure 1.5:** (a) Graphical interpretation of  $Q_{\text{rev}}$  on the  $TS$  graph. The area of a cycle on the  $TS$  graph is equal to the net amount of heat, also equal to  $-W_{\text{rev}}$ . (c) The Carnot cycle on the  $TS$  graph.

## 1.6. Other inequalities involving entropy

contains two isothermal branches at temperatures  $T_H$  and  $T_L$ , respectively, and two isentropic branches (Fig. 1.5c). When this cycle operates in power-plant mode (clockwise) it absorbs heat at  $T_H$  ( $Q_H > 0$ ) and rejects heat at  $T_L < T_H$  ( $Q_L < 0$ ) to produce a net amount of work ( $W < 0$ ). By first and second law we have<sup>16</sup>

$$|Q_H| - |Q_L| + |W| = 0, \quad -\frac{|Q_H|}{T_H} + \frac{|Q_L|}{T_L} = 0.$$

Eliminating  $|Q_L|$  between the two equations we obtain

$$\frac{|W|}{|Q_H|} = 1 - \frac{T_L}{T_H}. \quad (1.30)$$

This result gives the net amount of work as a ratio of the heat that is absorbed at the high-temperature branch of the process. In a power plant we pay for the input energy (the fuel that generates a heat source at  $T_H$ ) and profit from the work that is produced. In this sense the above result defines the efficiency of this conversion. If the cycle operates with irreversibilities, the actual work is less than the reversible amount. Therefore the general result is

$$\frac{|W|}{|Q_H|} \leq 1 - \frac{T_L}{T_H},$$

with the equal sign referring to reversible operation. What limits the efficiency is the leftover amount of heat  $Q_L$ . This heat, which is now at lower temperature, is of lesser value because to take advantage of it we need to couple it with an even lower temperature. The efficiency would be 1 only if  $T_H = 0$ . If the Carnot cycle operates in reverse (counter-clockwise), all energy flows are reversed: this cycle absorbs heat at  $T_L$  and rejects heat at  $T_H$ . The net effect is to transfer heat from a reservoir at a lower temperature to one at higher temperature, a task that requires work. In this mode the cycle operates as a heat pump.

The Carnot cycle demonstrates that a temperature gradient is capable of producing work. In this sense temperature is similar to the water level in a hydroelectric plant or the voltage of a battery pole. In all cases the spontaneous tendency of gradients to dissipate can be converted into work, and conversely, work can be used to generate and maintain gradients.

### 1.6 Other inequalities involving entropy

The inequality in Eq. (1.13) is the basic statement of the second law and says that the entropy of an isolated system can only increase (during the approach to equilibrium) or at best remain constant (when equilibrium is reached). The entropy of an isolated system cannot decrease. If we consider subregion I in Fig. 1.3 to be the “system,” and subregion II the “surroundings,” the combined system constitutes the “universe.” The universe is an isolated system and the inequality of Eq. (1.13) applies. We usually restate the second law in this case to say that in any process that involves interaction between a system and the surroundings, the total entropy change is positive or zero:

$$dS_{\text{sys}} + dS_{\text{sur}} \geq 0. \quad (1.31)$$

<sup>16</sup>To present the balances more clearly we use the absolute values for heat and work.

## 1.6. Other inequalities involving entropy

The equality applies either for a state of equilibrium between system and surroundings, or for *reversible* process. The individual terms in the above expression may be either positive or negative, but their sum must be positive. That is, the entropy of any part of the universe can be made to decrease; however, this must come at the expense of an increase that is *at least* of the same magnitude as the decrease, or bigger. If the process is conducted reversibly, the entropy of the universe is conserved; if the process is conducted irreversibly, there is a net generation of entropy. We define entropy generation as the net change in the entropy of the universe as a result of a process:

$$S_{\text{gen}} = \Delta S_{\text{sys}} + \Delta S_{\text{sur}}. \quad (1.32)$$

The second law may then be expressed as<sup>17</sup>

$$S_{\text{gen}} \geq 0, \quad (1.33)$$

<sup>17</sup>We do not write  $\Delta S_{\text{gen}}$  for the same reason we do not write  $\Delta Q$  or  $\Delta W$ : entropy generation is *not* an exact differential; its value depends on the path of the process.

which states that entropy generation cannot be negative.

### 1.6.1 Heat and work in arbitrary process

Consider a closed system undergoing a differential change of state. According to the second law we must have

$$dS_{\text{gen}} = dS - \frac{dQ}{T_{\text{sur}}} \geq 0,$$

where  $dS$  is the entropy change of the system and  $-dQ/T_{\text{sur}}$  is the entropy change of the surroundings, which are assumed to be at temperature  $T_{\text{sur}}$ . If the same change of state is accomplished reversibly, then

$$dS_{\text{gen}} = dS - \frac{dQ_{\text{rev}}}{T_{\text{sur}}} = 0.$$

Comparing the two results we conclude,

$$dQ \leq dQ_{\text{rev}} \quad \text{or} \quad Q \leq Q_{\text{rev}} \quad (\text{closed system}). \quad (1.34)$$

Whether the change of state is conducted irreversibly or reversibly, the first law gives

$$dU = dQ + dW = dQ_{\text{rev}} + dW_{\text{rev}}.$$

The inequality between actual and reversible heat implies a corresponding inequality for work:

$$dW \geq dW_{\text{rev}} \quad \text{or} \quad W \geq W_{\text{rev}} \quad (\text{closed system}). \quad (1.35)$$

If the work is positive (done by the surroundings on the system), the actual work is more than the reversible. If it is negative (done by the system on the surroundings), the absolute value of the work produced is less than the reversible case. In this view, irreversibilities during a process represent a penalty in terms of work: the work extracted is less but the work consumed is more than that for reversible operation.

---

## 1.6. Other inequalities involving entropy

The asymmetry between heat and work that is expressed by the opposite inequalities in Eqs. (1.34) and (1.35), implies that while the two forms of energy are equivalent with respect to the first law, with respect to the second law they are *not*—except under reversible conditions. For a process that consumes work, the actual work is more and the actual heat is less than the corresponding reversible amounts. Aspects of the process that could have been accomplished by heat are actually accomplished by work. The physical mechanism that gives rise to this behavior is transport under finite gradients. Finite gradients produce a net “force” that initiates transport at the expense of dissipating the gradient. The reversible process is the idealization that conserves gradients during a process at the expense of bringing the rate of transport to zero.

### 1.6.2 Adiabatic process in closed system

We return to the inequality in Eq. (1.34) and use Eq. (1.29) to write in the form

$$dS \geq \frac{dQ}{T}.$$

For adiabatic process this becomes

$$\left[ dS \geq 0 \right]_{\text{closed, adiabatic}}. \quad (1.36)$$

This is a less restrictive and thus even more general inequality than the original statement of the second law. Whereas the inequality in Eq. (1.13) applies to *isolated* system, Eq. (1.36) allows the system to exchange mechanical work (though not heat). Even if the condition of isolation is relaxed to allow for the exchange of mechanical work, the fundamental inequality of the second law remains intact. This is yet another way to say that work and heat are not equivalent in the eyes of the second law.

**Exercise 1.1.** *Lost work* Suppose a process is conducted in arbitrary manner such that all heat exchange is with a single reservoir at temperature  $T'$ . Show that the actual work is

$$W = W_{\text{rev}} + T' S_{\text{gen}},$$

where  $S_{\text{gen}}$  is the entropy generation for the process. The difference  $W - W_{\text{rev}}$  is called *lost work* and it is always positive. In what sense is this work “lost?”

**Exercise 1.2.** *Equivalent statements of second law* The inequality in Eq. (1.36) is often used as the statement of the second law. Starting with this inequality, obtain Eqs. (1.13), (1.18) and (1.22).

**Exercise 1.3.** *Adiabatic process and entropy* The inequality in Eq. (1.36) implies that it is not possible to decrease the entropy of an adiabatic system using work. In our macroscopic view, entropy is associated with gradients. It is surely possible to expend work to create a gradient inside a system, for example, to separate a gas into a region of high pressure and one of low pressure. The inequality in Eq. (1.36) states that such process will either *increase* the entropy of the system, or at best it will leave it unchanged. Why the creation of a gradient cannot *decrease* the entropy?

## 1.7 Temperature, entropy and heat

The definition of temperature in Eq. (1.21) may seem forced and arbitrary, but the fact is that until this point we have not defined temperature in any quantitative manner. The thermodynamic temperature is intimately related to entropy and heat. Variables  $T$  and  $S$  form a conjugate pair in the microcanonical differential, and the product  $TdS$  represents the amount of heat that is exchanged with the surroundings. To convince ourselves that temperature as defined in Eq. (1.21) is consistent with the familiar notion, consider the following problem. Two closed systems of fixed volume, one at temperature  $T_1$  and the other at  $T_2$ , make thermal contact, while they remain isolated from all other surroundings. Suppose the amount  $|dQ|$  of heat is transferred between the two systems. Let us determine the direction of this transfer. Since the two systems together are isolated, the second law gives

$$dS_1 + dS_2 \geq 0.$$

Let us define the sign of  $dQ$  based on system 1 such that positive  $dQ$  refers to transfer from 2 to 1 while negative  $dQ$  refers to the reverse direction. The entropy change in system 1 is  $\Delta S_1 = dQ/T_1$  and in system 2,  $\Delta S_2 = -dQ/T_2$ . Then,

$$\frac{dQ}{T_1} - \frac{dQ}{T_2} \geq 0 \Rightarrow dQ \left( \frac{1}{T_1} - \frac{1}{T_2} \right) \geq 0.$$

If  $T_1 > T_2$  then  $dQ < 0$  and the heat passes from system 1 to system 2; if  $T_2 > T_1$  then  $dQ > 0$  and the heat passes from system 2 to system 1. In both cases heat passes from higher to lower temperature. The result confirms that the definitions of entropy and temperature given earlier are consistent with physical observation.

**Absolute temperature** A closed system under the condition  $dV = 0$  can only exchange heat with its surroundings. This heat is given by Eq. (1.29):

$$dQ_{\text{rev}} = TdS.$$

When heat is removed at constant volume the temperature of the system decreases.<sup>18</sup> Since a finite amount of matter must contain a finite amount of energy it is not possible to extract unlimited amounts of heat from it. When all possible energy has been extracted we must reach the point such that  $dQ_{\text{rev}} = 0$ . This condition is met at  $T = 0$ . Since no amount of heat can be removed from a system at zero temperature it is not possible to have negative temperatures. The temperature as defined in Eq. (1.21) is the absolute temperature.

## 1.8 Thermodynamic calculus

### 1.8.1 Homogeneity

All *single-phase* systems at equilibrium exhibit remarkably simple behavior with respect to the extent (size) of the system. If we collect a sample whose mass is a fraction  $\lambda$  of the total mass of the system, then all extensive properties in the sample, volume, energy, etc., are the same fraction  $\lambda$  of the property in the big

<sup>18</sup>Recall

$$C_V = \left. \frac{dQ_{\text{rev}}}{dT} \right|_{V, n_i},$$

which shows that  $dQ_{\text{rev}}$  and  $dT$  have the same sign provided that  $C_V$  is positive (see Section 3.4).

system, and all intensive properties (pressure, density, molar heat capacity etc.) are the same. Mathematically, if  $F$  is extensive and  $f$  is intensive, then:

$$F(\lambda U, \lambda V, \lambda n_i) = \lambda F(U, V, n_i), \quad (1.37)$$

$$f(\lambda U, \lambda V, \lambda n_i) = f(U, V, n_i). \quad (1.38)$$

These relationships express the mathematical property of *homogeneity*. A function  $f(x_1, x_2, \dots)$  is homogeneous in  $x_1, x_2, \dots$  with degree  $\nu$ , where  $\nu$  is constant, if

$$f(\lambda x_1, \lambda x_2, \dots) = \lambda^\nu f(x_1, x_2, \dots)$$

for all  $x_i$  and all  $\lambda$ . Extensive properties are homogeneous with degree 1 with respect to extensive independent variables; intensive properties are homogeneous with degree 0 with respect to extensive independent variables.

**Euler's theorem for homogeneous functions** An important mathematical property of homogeneous functions is given by Euler's theorem, which states,

$$\nu f(x_1, x_2, \dots) = x_1 \left( \frac{\partial f}{\partial x_1} \right) + x_2 \left( \frac{\partial f}{\partial x_2} \right) + \dots \quad (1.39)$$

where  $\nu$  is the degree of homogeneity. For  $\nu = 1$ , this becomes

$$f(x_1, x_2, \dots) = x_1 \left( \frac{\partial f}{\partial x_1} \right) + x_2 \left( \frac{\partial f}{\partial x_2} \right) + \dots \quad (1.40)$$

The theorem states that an extensive property can be expressed as a linear combination of its partial derivatives.<sup>19</sup>

**Euler integration** Given a function  $f = f(x_1, x_2, \dots)$ , its differential is

$$df = \left( \frac{\partial f}{\partial x_1} \right) dx_1 + \left( \frac{\partial f}{\partial x_2} \right) dx_2 + \dots \quad (1.41)$$

If  $f$  is also homogeneous in all  $x_i$  with degree 1, then by Euler's theorem,

$$f = x_1 \left( \frac{\partial f}{\partial x_1} \right) + x_2 \left( \frac{\partial f}{\partial x_2} \right) + \dots, \quad (1.42)$$

which shows that the differentials in Eq. (1.41) may simply be replaced by the corresponding full variables. Indeed we may go from Eq. (1.41) to (1.42) by formal integration along the following path: increase all extensive properties by the same factor, then integrate with respect to all  $x_i$  from 0 until final state  $(x_1, x_2, \dots)$ . This integration represents a simple physical process that can be conducted experimentally: imagining the state to be liquid (for pure convenience) in container A, pour its contents into container B. As the material accumulates in container B all extensive properties increase simultaneously by the same factor. Accordingly, all intensive properties remain constant. All partial derivatives in Eq. (1.41) are intensive,<sup>20</sup> and since they remain constant, integration with respect to  $x_i$  is trivial and immediately leads to Eq. (1.42). We call this an Euler path to indicate that all extensive properties increase simultaneously by the same amount while all intensive properties remain fixed.

<sup>19</sup>To prove Euler's theorem, start with

$$f(\nu x_1, \nu x_2, \dots) = \lambda^\nu f(x_1, x_2, \dots)$$

and differentiate both sides with respect to  $\lambda$ . For the left-hand side we have

$$\begin{aligned} \frac{\partial f}{\partial \lambda} &= \left( \frac{\partial f}{\partial \lambda x_1} \right) \left( \frac{\partial \lambda x_1}{\partial \lambda} \right) + \dots \\ &= \left( \frac{\partial f}{\partial x_1} \right) x_1 + \left( \frac{\partial f}{\partial x_2} \right) x_2 + \dots \end{aligned}$$

The right-hand side gives

$$\frac{\partial (\lambda^\nu f)}{\partial \lambda} = \nu \lambda^{\nu-1} f$$

Equating these results we obtain

$$\nu \lambda^{\nu-1} f = x_1 \left( \frac{\partial f}{\partial x_1} \right) + x_2 \left( \frac{\partial f}{\partial x_2} \right) + \dots$$

This applies for any  $\lambda$ . Setting  $\lambda = 1$  we obtain

$$\nu \lambda^{\nu-1} f = x_1 \left( \frac{\partial f}{\partial x_1} \right) + x_2 \left( \frac{\partial f}{\partial x_2} \right) + \dots,$$

which proves the theorem.

<sup>20</sup>Problem 1.6.

**Gibbs-Duhem equation** We write Eqs. (1.40) and (1.41) in the simpler form

$$f = x_1 f_1 + x_2 f_2 + \dots \quad (1.43)$$

$$df = f_1 dx_1 + f_2 dx_2 + \dots, \quad (1.44)$$

where  $f_i$  is the partial derivative of  $f$  with respect to  $x_i$ . We take the differential of (1.43) with respect to all  $x_i$  and all  $f_i$  on the right-hand side:

$$df = (f_1 dx_1 + f_2 dx_2 \dots) + (x_1 df_1 + x_2 df_2 \dots).$$

The quantity in the first parenthesis is the differential  $df$ , then we must have

$$x_1 df_1 + x_2 df_2 \dots = 0. \quad (1.45)$$

This is the *Gibbs-Duhem* equation and expresses the fact that the derivatives  $f_i$  of a homogeneous function are not independent from each other. Indeed, Eq. (1.43) establishes a constraint on  $f_i$  and Eq. (1.45) expresses a mathematical consequence of that constraint.

**Application to thermodynamics** Extensive properties are homogeneous functions of  $(U, S, n_i)$ . If  $F$  is such property, then Euler's theorem states<sup>21</sup>

$$F(U, V, n_i) = U \left( \frac{\partial F}{\partial U} \right)_{V, n_i} + V \left( \frac{\partial F}{\partial V} \right)_{U, n_i} + \sum_i n_i \left( \frac{\partial F}{\partial n_i} \right)_{U, V, n_i}. \quad (1.46)$$

<sup>21</sup>When  $n_j$  is held constant during differentiation with respect to  $n_i$ , we mean to say that all  $n_{j \neq i}$  are held constant.

This we write more compactly as

$$F = UF_U + VF_V + \sum n_i F_{n_i} \quad (1.47)$$

where  $F_X$  is the derivative with respect to independent variable  $X = \{U, V, n_i\}$  keeping all other variables constant. Euler theorem allows us to write any extensive property as a linear combination of  $U$ ,  $V$  and  $n_i$ , with factors equal to the corresponding partial derivatives.

### 1.8.2 Partial molar properties

Of the many sets of independent variables that can be used to describe a property  $F$ , the set  $(T, P, n_i)$  is the most practical because it involves variables (temperature, pressure, composition) that are straightforward to measure and control in an experimental setting. We will define temperature later, but for our purposes here it suffices to say that  $T$  is *intensive*, as is  $P$ .

The differential of extensive property  $F(T, P, n_i)$  is

$$dF = \left( \frac{\partial F}{\partial T} \right)_{P, n_i} dT + \left( \frac{\partial F}{\partial P} \right)_{T, n_i} dP + \sum_i \left( \frac{\partial F}{\partial n_i} \right)_{T, P, n_j} dn_i.$$

The derivatives with respect to  $n_i$  at fixed  $T$  and  $P$  are called *partial molar*<sup>22</sup> and given the symbol  $\bar{F}_i$ :

$$\bar{F}_i = \left( \frac{\partial F}{\partial n_i} \right)_{T, P, n_j}. \quad (1.48)$$

<sup>22</sup>The name *partial molar* is reserved for the derivative with respect to  $n_i$  at fixed temperature and pressure. The derivative

$$\left( \frac{\partial F}{\partial n_i} \right)_{UV, n_{i \neq j}},$$

encountered earlier, is not a partial molar derivative.

The differential of  $F$  can now be written in the simpler form,

$$dF = \left(\frac{\partial F}{\partial T}\right)_{P,n_i} dT + \left(\frac{\partial F}{\partial P}\right)_{T,n_i} dP + \sum_i \bar{F}_i dn_i. \quad (1.49)$$

As an extensive property,  $F$  is homogeneous with degree 1 in  $n_i$  provided that  $T$  and  $P$  are held constant.<sup>23</sup> Euler's theorem applied to all  $n_i$ , which are the extensive independent variables of  $F(T, P, n_i)$ , gives

$$F = \sum n_i \bar{F}_i. \quad (1.50)$$

We may derive the same result by a thermodynamic argument. Partial molar properties are intensive (why?), which means they are fixed if  $T$ ,  $P$  and all mol fractions are constant. By Euler integration of Eq. (1.49) at fixed  $T$ ,  $P$ , and constant  $\bar{F}_i$  gives Eq. (1.50). This equation provides a simple interpretation for the partial molar property: it is the contribution of component  $i$  to extensive property  $F$  of the system.

**Gibbs-Duhem equation** We differentiate Eq. (1.50) with respect to all variables on the right hand side,

$$dF = \sum_i \bar{F}_i dn_i + \sum_i n_i d\bar{F}_i, \quad (1.51)$$

then combine with Eq. (1.49) to obtain

$$\left(\frac{\partial F}{\partial T}\right)_{P,n_i} dT + \left(\frac{\partial F}{\partial P}\right)_{T,n_i} dP = \sum_i n_i d\bar{F}_i. \quad (1.52)$$

At constant temperature and pressure this reduces to

$$\sum n_i d\bar{F}_i = 0, \quad (\text{const. } T, P). \quad (1.53)$$

Dividing both side by the total number of moles, the above result can be expressed in terms of mole fractions in the form,

$$\sum x_i d\bar{F}_i = 0, \quad (\text{const. } T, P). \quad (1.54)$$

This is a form of the Gibbs-Duhem equation and states that the variations with composition of the partial molar properties of mixture are not independent of each other. The name *Gibbs-Duhem* usually refers to Eq. (1.54) applied to the partial molar Gibbs energy (chemical potential). Nonetheless, Eq. (1.54) is general and applies to any partial molar property.

The differential with respect to  $n_i$  at fixed  $(T, P, n_j)$  may be viewed as an operator that applies to a function  $F$ . This operator is linear, i.e., it has the property

$$\overline{aF + bG} = a\bar{F}_i + b\bar{G}_i,$$

where  $a$  and  $b$  may be functions of  $T$  and  $P$  but not of  $n_i$ . Moreover, since the order of differentiation does not matter, we also have<sup>24</sup>

<sup>23</sup>An extensive property may be a function of a combination of variables, some intensive and some extensive. This means that if all intensive variables are held constant and all extensive variables are multiplied by  $\lambda$ ,  $f$  is multiplied by  $\lambda$  and Euler's theorem applies with respect to the extensive variables only. For a mathematical example see Problem 1.7.

<sup>24</sup>This equation reads

$$\frac{\partial}{\partial n_i} \left(\frac{\partial F}{\partial T}\right)_{P,n_i} \Big|_{T,P,n_j} = \frac{\partial}{\partial T} \left(\frac{\partial F}{\partial n_i}\right)_{P,n_j} \Big|_{P,n_i}.$$

$$\left(\frac{\partial F}{\partial T}\right)_{P,n_j} = \left(\frac{\partial \bar{F}_i}{\partial T}\right)_{P,n_i};$$

and similarly for the derivative with respect to pressure:

$$\left(\frac{\partial F}{\partial P}\right)_{T,n_j} = \left(\frac{\partial \bar{F}_i}{\partial P}\right)_{T,n_i}.$$

The results amount to distributing the partial molar operator inside the derivative; this can only be done with the derivative with respect to  $T$  at fixed  $P$  and  $n_j$ , or with respect to  $P$  at fixed  $T$  and  $n_j$ . We will encounter applications of these results later, but here we give two examples to demonstrate.

**Example 1.1.**

Obtain the partial molar property of a pure component.

*Solution* For a pure component,  $n_1 = n$ ,  $n_j = 0$  and  $F_i(T, P, n) = n f_i(T, P)$  where  $n$  is the total number of moles and  $f_i = F_i/n$  is the molar property of the pure component. We apply the definition of the partial molar property and noting that the specification  $n_j = \text{const.}$  is not needed since  $n_j = 0$ , we have

$$\bar{F}_i = \left(\frac{\partial n f_i(T, P)}{\partial n}\right)_{T, P} = f_i. \quad (1.55)$$

The partial molar property of a pure component is the same as the molar property of the component.

The result also applies to mixtures in the limit  $x_i \rightarrow 1$ ,

$$\lim_{x_i \rightarrow 1} \bar{F}_i = f_i, \quad (1.56)$$

and states that the partial molar property of a component whose mole fraction in the mixture is close to unity is asymptotically equal to the molar property of the pure component. On the other hand, the partial molar property of a component in the limit  $x_i \rightarrow 0$  is not zero but a value characteristic of the component in the infinite dilution limit.

**Example 1.2.**

Obtain the relationship between  $\bar{U}_i$ ,  $\bar{H}_i$  and  $\bar{V}_i$ .

*Solution* Starting with  $H = U + PV$  we have

$$\bar{H}_i = \overline{U + PV} = \bar{U}_i + P\bar{V}_i.$$

The result amounts to replacing all extensive properties with their partial molar counterparts.

**Example 1.3.**

As we will see later, the derivative of enthalpy with respect to pressure at fixed temperature and composition is given by

$$\left(\frac{\partial H}{\partial P}\right)_{T,n_i} = V - T \left(\frac{\partial V}{\partial T}\right)_{P,n_i},$$

Write the corresponding expression for the partial molar enthalpy  $\bar{H}_i$ .

*Solution* Using the linear property of the partial molar operator we obtain

$$\left(\frac{\partial \bar{H}}{\partial P}\right)_{T, n_i} = \bar{V} - T \left(\frac{\partial \bar{V}}{\partial T}\right)_{P, n_i}.$$

Again result amounts to replacing all extensive properties with their partial molar counterparts.

## 1.9 Additional reading

The material of this chapter (first and second law) is covered in most undergraduate textbooks. See for example [Matsoukas \(2013\)](#), chapters 3 and 4.

## 1.10 Problems

**Problem 1.1.** (Homogeneity) a) We can construct an unlimited number of multivariate homogeneous functions of any degree. For example, all of the functions below (and their linear combinations) are homogeneous with degree  $\nu$  ( $a$  and  $b$  are constants):

$$(ax_1 + bx_2)^\nu, \quad ax_1^\nu + bx_2^\nu, \quad a(x_1 x_2)^{\nu/2} \left(\frac{x_1}{x_1 + x_2}\right)^b.$$

If however  $f(x)$  is a function of a single variable, there is only one such function. What is that function? (Hint: set  $x_2 = x_1$  in any of the above examples, but you must show it is the only one.)

b) Use the previous result to show that any extensive property  $F(T, P, n)$  of pure component  $F(T, P, n)$  can be expressed as

$$F(T, P, n) = n f(T, P),$$

where  $f$  depends only on  $T$  and  $P$ . Obviously  $f$  is the molar form of  $F$  and as an intensive property it depends on  $T$  and  $P$ . The problem however asks you to reach this conclusion using the homogeneity properties of  $F$  (recall that  $F$  is extensive in  $n$ ).

**Problem 1.2.** (Homogeneity) Show that if  $f(x, y)$  is a homogeneous function of degree  $\nu$ , then:

- The first derivatives,  $f_x$  and  $f_y$ , are homogeneous with degree  $\nu - 1$ .
- The second order derivatives,  $f_{xx}$ ,  $f_{yy}$  and  $f_{xy}$  are homogeneous with degree  $\nu - 2$ .
- The derivatives of order  $k$  are homogeneous with degree  $\nu - k$ .
- Use these results to show that partial molar properties are intensive.

**Problem 1.3.** (Homogeneity) Function  $f(x, y)$  is homogeneous in  $x$  and  $y$  with degree  $\nu$ . Show that  $h(x, y)$ , defined as

$$h(x, y) = \int_{ay}^x f(z, y) dz$$

with  $a$  a constant is homogeneous and determine the degree of homogeneity.

**Problem 1.4.** (Homogeneity and curvature) If  $f(x, y)$  is homogeneous in  $x$  and  $y$  with degree 1, show that:

- $x f_{xx} + y f_{xy} = 0$ .
- $x^2 f_{xx} - y^2 f_{xy} = 0$ .

## 1.10. Problems

c)  $f$  has the same curvature with respect to both  $x$  and  $y$ , i.e., if it is concave (convex) in  $x$  it is also concave (convex) in  $y$  and vice versa.

**Thermodynamic relevance:** If  $S(U, V, N)$  is homogeneous in  $U, V, N$ , and concave in one of these variables, it is concave in all of them.

**Problem 1.5.** (Homogeneity) Show that if  $f(x, y)$  satisfies the Euler theorem for all  $x$  and  $y$ , i.e.,

$$f(x, y) = xf_x(x, y) + yf_y(x, y)$$

then it is homogeneous in  $x$  and  $y$  with degree 1.

**Problem 1.6.** (Homogeneity) Show that if  $f(x, y)$  is a homogeneous function of degree  $v$ , then:

- a) The first derivatives,  $f_x$  and  $f_y$ , are homogeneous with degree  $v - 1$ .
- b) The second order derivatives,  $f_{xx}$ ,  $f_{yy}$  and  $f_{xy}$  are homogeneous with degree  $v - 2$ .
- c) The derivatives of order  $k$  are homogeneous with degree  $v - k$ .
- d) Use these results to show that partial molar properties are intensive.

**Problem 1.7.** (Homogeneity) A function may be homogeneous in some but not all of its independent variables. Euler's theorem then applies to those variables that satisfy homogeneity. Consider the function  $f(x_1, x_2, x_3)$  defined below:

$$f(x_1, x_2, x_3) = \frac{x_1 x_2 + x_2^2 x_3}{x_1 + x_2}$$

- a) Show that  $f$  is homogeneous in  $x_1$  and  $x_2$  with degree 1 but is not homogeneous in  $x_3$ .
- b) Conform that  $f$  satisfies

$$f = x_1 \frac{\partial f}{\partial x_1} + x_2 \frac{\partial f}{\partial x_2}$$

This is Euler's theorem applied only to  $x_1$  and  $x_2$ .

- c) Confirm that

$$x_1 \frac{\partial^2 f}{\partial x_1^2} + x_2 \frac{\partial^2 f}{\partial x_1 \partial x_2} = 0.$$

This is the Gibbs-Duhem equation divided by  $dx_1$  (you can also confirm the corresponding result for division by  $dx_2$ ).

**Thermodynamic Relevance:** Here  $x_3$  is analogous to an intensive property while  $f, x_1$  and  $x_2$  are extensive. Euler's theorem and the Gibbs-Duhem equation are satisfied with respect to all extensive variables. The Eq. in part (c) is analogous to Eq. (1.50), which expresses  $F(P, T, n_i)$  in Euler form with respect to  $n_i$ .

**Problem 1.8.** (Undergraduate review–Energy balance in open system) Consider an open system that exchanges mass with the surroundings via inlet and outlet streams. The mass passing through a stream is  $\pm dm$  (+ if it enters the system, – if it exits) with specific internal energy  $u$ , pressure  $P$ , density  $\rho$ , velocity  $v$  and elevation  $z$ . Derive the energy balance given below

$$dU + \Delta \left\{ \left( u + \frac{P}{\rho} + \frac{v^2}{2} + gz \right) dm \right\} = dQ + dW \quad (1.57)$$

where the operator  $\Delta[\dots]$  is defined as a difference between all inlet and outlet streams:

$$\Delta\{\dots\} = \sum_{\text{all out}} \{\dots\} - \sum_{\text{all in}} \{\dots\}.$$

## 1.10. Problems

**Problem 1.9.** (Undergraduate review–Macroscopic energy and entropy balance) Assuming that all the heat in problem 1.8 is exchanged with a bath at  $T_{\text{bath}}$ , show that the macroscopic entropy balance is

$$dS + \Delta(sdm) = \frac{dQ}{T_{\text{bath}}} + dS_{\text{gen}}, \quad (1.58)$$

or

$$\frac{dS}{dt} + \Delta(\dot{s}m) = \frac{\dot{Q}}{T_{\text{bath}}} + \dot{S}_{\text{gen}}$$

where  $S$  is the entropy of the system,  $s$  is the specific entropy of stream at the point it crosses the system boundary, and  $S_{\text{gen}}$  is the entropy generation.

**Problem 1.10.** (Undergraduate review–Bernoulli equation) The mechanical energy balance for steady-state isothermal flow through a pipe is given by the Bernoulli equation:

$$\frac{\dot{W}}{\dot{m}} = \int \frac{dP}{\rho} + \Delta\left(\frac{v^2}{2}\right) + g\Delta z + \epsilon,$$

where  $\Delta$  refers to the change from inlet to outlet, and  $\epsilon$  is the loss due to wall friction and viscous effects. Derive this equation by applying the first and second law to this process.

**Problem 1.11.** (Exergy) Think of the surroundings (temperature  $T_0$ , pressure  $P_0$ ) as a default source and sink for all energy exchanges. In order to remove the system from ambient conditions we must consume work. Similarly, if the system starts at other than ambient conditions, it has the ability to produce work in its tendency to equilibrate with the surroundings. Show that for a closed system the maximum work to change the state from ambient conditions is

$$W_{\text{max}} = (U - U_0) + P_0(V - V_0) - T_0(S - S_0),$$

where  $U$ ,  $V$ ,  $S$ , are the properties of the system at the final state and  $U_0$ ,  $V_0$ ,  $S_0$  are the same properties at the conditions of the environment. This motivates the definition of a new function, *exergy*, or *availability*,

$$\mathcal{E} = U + P_0V - T_0S,$$

such that the maximum work for a change of state is the difference in exergy between the two states,  $\Delta\mathcal{E}$ . *Hint:* First show that the reversible work for a process in a closed system that exchanges all its energy with a single bath at  $T_0$  is

$$W_{\text{ideal}} = \Delta U - T_0\Delta S,$$

(this is also known as the *ideal work* for the process). Then think of a reason to argue that the amount  $-P_0(V - V_0)$  ought to be subtracted from this result in order to obtain exergy.

As a numerical example, calculate the maximum work that can be extracted from a tank that contains pressurized air ( $C_p = 30 \text{ J/mol K}$ ) at 10 bar,  $200^\circ\text{C}$ . Take the surroundings to be at 1 bar,  $20^\circ\text{C}$ .

**Problem 1.12.** (Exact differential) Which of the two differentials below is exact?

$$dA = 6xe^{-2y} dx - 6x^2e^{-2y} dy$$

$$dB = 6xe^{-2y} dx + 6x^2e^{-2y} dy$$

For each differential, calculate the integral along the paths  $(x_0, y_0) \rightarrow (x_0, y) \rightarrow (x, y)$ , and  $(x_0, y_0) \rightarrow (x, y) \rightarrow (x, y)$  and identify the function that is represented by the exact differential.

## 1.10. Problems

**Problem 1.13.** (Exact differential) With each of the differentials below, first determine if they satisfy the criterion of exactness and then integrate to obtain the equation of state:

$$\begin{aligned} dV &= \frac{R}{P} dT - \frac{RT}{P^2} dP \\ dV &= \frac{V}{T} dT - \frac{V}{P} dP \\ dV &= -\frac{RT}{P^2} dP \end{aligned}$$

**Problem 1.14.** (Exact differential) a) Show that the differential below is exact, then obtain its integrated form starting from point  $f_0 = f(x_0, y_0)$ :

$$df = \left( \frac{b}{x^2 y} - \frac{a}{(a-x)^2} \right) dx + \frac{b}{x y^2} dy$$

b) Evaluate the integration constants and determine the function  $f(x, y)$  if

$$\lim_{x \rightarrow \infty} f(x, y) = 1$$

c) If  $f$  stands for the compressibility factor  $Z$ ,  $x$  stands for the molar volume  $V$ , and  $y$  stands for temperature, which equation of state is represented by this function?

**Problem 1.15.** (Ideal gas) In the derivation of entropy we used

$$P = \left( \frac{\partial S}{\partial V} \right)_{U, n_i} \left( \frac{\partial U}{\partial S} \right)_{V, n_i}$$

Confirm that this result holds true for an ideal gas.

**Problem 1.16.** (Review questions) a) True or false and why: equation  $dU = T dS - P dV$  is valid for pure components only.

b) True or false and why: if a process is sufficiently slow it may be assumed to be reversible.

c) True or false and why: it is possible to have  $S_{\text{gen}} < 0$  in some sub-region of the universe as long as  $S_{\text{gen}}$  for the entire universe is positive.

d) True or false and why: it is possible to have  $dS_{\text{gen}} < 0$  during a portion of a process as long as  $S_{\text{gen}}$  for the entire process is positive.

e) True or false and why: To decrease the entropy of a system, it is necessary and sufficient to supply work from the surroundings.

f) What is the thermodynamic difference between a system at equilibrium and a steady-state process?

## Transformations of the Fundamental Differential

The microcanonical differential is the basis for the mathematical development of thermodynamics. Practically every relationship in thermodynamics follows by applying standard mathematical calculus on this differential.<sup>1</sup> The general form of thermodynamic differentials is

$$df = a_1 dx_1 + a_2 dx_2 + \dots \quad (2.1)$$

where  $f$ ,  $x_i$  and  $a_i$  refer to common thermodynamic properties ( $U$ ,  $T$ ,  $\dots$ ). Given a generic differential of the form in Eq. (2.1) we may immediately identify the multipliers of each differential as partial derivatives,

$$a_i = \left( \frac{\partial f}{\partial x_i} \right)_{x_{j \neq i}}. \quad (2.2)$$

These derivatives are always understood to be taken within the variable set ( $x_1, x_2, \dots$ ) that comprises the set of independent variables on the right-hand side of the differential. Since this set can be transformed, it is important to remember which set corresponds to a function – this question is easily answered by looking at the differential of  $f$ . To avoid ambiguity we always indicate the variables that remain constant as subscripts in the partial derivative. Given the generic differential  $df$  we also have

$$\left( \frac{\partial a_2}{\partial x_1} \right)_{x_{i \neq 1}} = \left( \frac{\partial a_1}{\partial x_2} \right)_{x_{i \neq 2}}, \quad (2.3)$$

a result that is easily confirmed since both sides express the mixed derivative ( $\partial^2 f / \partial x_1 \partial x_2$ ) (this of course can be written between any two variables). Applying this relationship to variables ( $S, V$ ) of the microcanonical differential we obtain

$$\left( \frac{\partial T}{\partial V} \right)_{S, n_i} = \left( \frac{\partial P}{\partial S} \right)_{V, n_i}. \quad (2.4)$$

This equation is known as a *Maxwell* relationship and is part of a set of similar equations obtained by the above method, as we will see later in this chapter.

Equation (2.3) ensures that the generic differential in Eq. (2.1) is exact, i.e., it represents the differential of a quantity  $f$  that is an expressed function of  $x_i$ .

<sup>1</sup>For equilibrium and stability we also need the inequality of the second law. A few other relationships such as the ideal gas law and equations of state in general, are not connected to this differential.

## 2.1. The Legendre transformation

When an exact differential is integrated between two states  $A = (x_1, x_2, \dots)^A$  and  $B = (x_1, x_2, \dots)^B$ , the result depends only on the coordinates of the end states and not on the path that is used for the integration. All thermodynamic functions are exact differentials. By contrast, heat and work are *path* functions. While the differential of path functions has the form of Eq. (2.1), it is not possible to find a function  $f(x_1, x_2, \dots)$  that has the given differential. When an inexact differential is integrated between two points, the result depends on the path.

### Example 2.1. Exact and inexact differentials

Determine which of the two differentials below is exact and which is not:

$$df = ydx + xdy$$

$$dg = ydx - xdy.$$

Integrate both differentials from  $(x_1, y_1) = (0, 0)$  to  $(x_2, y_2) = (1, 1)$  along the following paths: (a)  $y = x$  and (b)  $y = x^2$ . Notice that only the exact differential gives the same answer for both paths.

## 2.1 The Legendre transformation

The Legendre transform is a useful mathematical tool that transforms the micro-canonical differential into mathematically equivalent forms, each with its own thermodynamic interpretation. We start with a function  $f = f(x_1, x_2, \dots)$ , whose differential is

$$df = a_1 dx_1 + a_2 dx_2 + \dots \quad 2.1$$

The Legendre transformation of  $f$  with respect to  $x_1$  is a new function  $h$ , defined as

$$h = f - a_1 x_1 \quad (2.5)$$

whose independent variables are  $(a_1, x_2, \dots)$ . This operation transforms both the function and the set of its independent variables. The graphical interpretation of the Legendre transformation is demonstrated in Fig. 2.1. The derivative  $a_1$  at  $x_1 = x_1^*$  is the slope of  $f$  at that point. The equation of the tangent line is

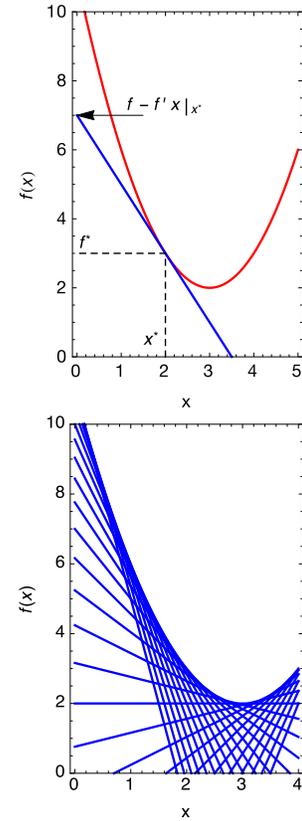
$$y = f^* + a_1^*(x_1 - x_1^*),$$

whose  $x_1$  intercept is  $f^* - a_1^* x_1^*$ . We recognize the intercept of the tangent line as the Legendre transformation of  $f$  evaluated at  $x_1^*$ . The intercept/slope pair represents a transformation of the original function. If we plot a series of straight lines with slope  $a_1$  and corresponding intercept  $f - a_1 x_1$ , the result is a graphical reconstruction of the original function  $f$ , as shown at the bottom of Fig. 2.1).

The graphical construction shows that the Legendre transformation of  $f$  with respect to variable  $x_1$  should be viewed as a function of  $a_1$  rather than of  $x_1$ ; schematically,

$$f(x_1, x_2, \dots) \longleftrightarrow h(a_1, x_2, \dots)$$

We now construct the differential of  $h$  with respect to its proper set of variables  $(a_1, x_2, \dots)$ . First we calculate the partial derivative of the transform with respect



**Figure 2.1:** (Top) The slope of the tangent line is  $a_1$  and the intercept is the Legendre transformation of  $f$  evaluated at the tangent point. (Bottom) Reconstruction of the original function from a series of tangent lines. The Legendre transformation gives the intercept of these lines as a function of their slope.

to  $a_1$ :

$$\begin{aligned} \left(\frac{\partial h}{\partial a_1}\right)_{x_i \neq 1} &= \frac{\partial}{\partial a_1} (f - a_1 x_1)_{x_i \neq 1} = \\ &= \left(\frac{\partial f}{\partial x_1}\right)_{x_i \neq 1} \left(\frac{\partial x_1}{\partial a_1}\right)_{x_i \neq 1} - x_1 - a_1 \left(\frac{\partial x_1}{\partial a_1}\right)_{x_i \neq 1} = -x_1. \end{aligned}$$

The derivative with respect to an untransformed variable, say  $x_2$ , is

$$\left(\frac{\partial h}{\partial x_2}\right)_{a_1, x_i \neq 2} = \frac{\partial}{\partial x_2} (f - a_1 x_1)_{x_i \neq 1} = a_2.$$

Now that we have all the partial derivatives we construct the differential of the transformed function with respect to its proper set of variables  $(a_1, x_2, \dots)$ :

$$dh = -x_1 da_1 + a_2 dx_2 + \dots \quad (2.6)$$

Comparing this with the original differential, we come to the following conclusion: by Legendre-transforming  $f$  with respect to  $x_1$  we obtain a new function whose differential is identical to that of the original function except that the term  $a_1 dx_1$  of the original function is now replaced by  $-x_1 da_1$ . Essentially, the Legendre transformation swaps the role of  $x_1$  and  $a_1$  in the differential. The transformation can be applied to multiple variables simultaneously. In this case, the terms  $a_i dx_i$  of all transformed variables are replaced in the differential by  $-x_i da_i$  while those of the untransformed variables remain unchanged.

**Variational properties of Legendre transformation** We define the function

$$\Phi(x, a) = f(x) - ax. \quad (2.7)$$

Its second derivative with respect to  $x$  is

$$\frac{\partial^2 \Phi}{\partial^2 x} = \frac{\partial^2 f}{\partial^2 x}.$$

$f$  and  $\Phi$  are both convex functions of  $x$  or both concave, because they have the same second derivative. Let us assume they are concave. Then we have the following property:

*The Legendre transformation of  $f(x)$  is the maximum of  $\Phi$  with respect to  $x$  at fixed  $a$ :*

$$h(a) = \max_x \{f(x) - ax\}. \quad (2.8)$$

The proof is straightforward. The condition that maximizes  $\Phi$  is

$$\frac{\partial \Phi}{\partial x} = \frac{df}{dx} - a = 0,$$

Then  $a = f'$  and the maximum value of  $\Phi$  is the Legendre transformation of  $f$ . That the extremum is a maximum follows from the fact that  $\Phi$  is concave in  $x$ . Equation (2.8) is the definition that usually appears in mathematical textbooks.<sup>2</sup>

### 2.1.1 The Helmholtz function

Let us apply the Legendre transformation to the microcanonical differential by transforming with respect to  $x_1 = T$ . The transformed function is the Helmholtz free energy:

$$U - \left( \frac{\partial U}{\partial S} \right)_{V n_i} = S = U - TS \equiv A. \quad (2.9)$$

The differential of  $A$  is obtained by application of the Legendre formula:

$$dA = -SdT - PdV + \sum_i \mu_i dn_i. \quad (2.10)$$

We make the following identifications:

$$S = - \left( \frac{\partial A}{\partial T} \right)_{V n_i} \quad (2.11)$$

$$P = - \left( \frac{\partial A}{\partial V} \right)_{T n_i} \quad (2.12)$$

$$\mu_i = \left( \frac{\partial A}{\partial n_i} \right)_{TV n_{j \neq i}}. \quad (2.13)$$

We also obtain the corresponding Maxwell equation by application of the criterion of exactness on  $S$  and  $P$ :

$$\left( \frac{\partial S}{\partial V} \right)_{T n_i} = \left( \frac{\partial P}{\partial T} \right)_{V n_i}. \quad (2.14)$$

As one last result we give the result of the Euler integration of the Helmholtz differential:

$$A = -PV + \sum_i \mu_i n_i. \quad (2.15)$$

The proof is left as an exercise.

### 2.1.2 The enthalpy function

We now transform  $U$  with respect to  $V$ . The new function is *enthalpy*:

$$U - V \left( \frac{\partial U}{\partial V} \right)_{S n_i} = U + PV \equiv H, \quad (2.16)$$

whose differential is

$$dH = TdS + VdP + \sum_i \mu_i dn_i. \quad (2.17)$$

The following results are now easily obtained:

$$T = \left( \frac{\partial H}{\partial S} \right)_{P n_i} \quad (2.18)$$

$$V = \left( \frac{\partial H}{\partial P} \right)_{S n_i} \quad (2.19)$$

$$\mu_i = \left( \frac{\partial H}{\partial n_i} \right)_{SP n_{j \neq i}} \quad (2.20)$$

and

$$\left(\frac{\partial T}{\partial P}\right)_{S n_i} = \left(\frac{\partial V}{\partial S}\right)_{P n_i}. \quad (2.21)$$

The same result may be obtained by Euler integration of  $dH$ . This derivation is left as an exercise.

### 2.1.3 The Gibbs function

We perform the Legendre transformation of the internal energy with respect to both  $S$  and  $V$ . The result of this transformation is the *Gibbs function*,  $G$ :

$$U - S\left(\frac{\partial U}{\partial S}\right)_{V n_i} - V\left(\frac{\partial U}{\partial V}\right)_{S n_i} = U - TS + PV \equiv G. \quad (2.22)$$

The differential of the Gibbs function is

$$dG = -SdT + VdP + \sum_i \mu_i dn_i, \quad (2.23)$$

and from this we obtain the following relationships:

$$S = -\left(\frac{\partial G}{\partial T}\right)_{P n_i} \quad (2.24)$$

$$V = \left(\frac{\partial G}{\partial P}\right)_{T n_i} \quad (2.25)$$

$$\mu_i = \left(\frac{\partial G}{\partial n_i}\right)_{TP n_{j \neq i}} \quad (2.26)$$

$$\left(\frac{\partial S}{\partial P}\right)_{T n_i} = -\left(\frac{\partial V}{\partial T}\right)_{P n_i}. \quad (2.27)$$

The Gibbs free energy is particularly useful because its proper variables are temperature, pressure, and the number of moles, all of which can be measured and controlled experimentally very easily, unlike  $S$  and  $V$ , for example, which appear in the microcanonical differential.

## 2.2 Partial molar properties

Partial derivatives with respect to  $n_i$  at fixed  $T$ ,  $P$ ,  $n_{j \neq i}$  are called *partial molar* and play an important role in thermodynamics. Given an extensive property  $F$ , its partial molar derivative  $\bar{F}_i$  is defined as

$$\bar{F}_i = \left(\frac{\partial F}{\partial n_i}\right)_{TP n_{j \neq i}}.$$

Partial molar derivatives are *intensive*.<sup>3</sup> Expressing  $F$  as a function of temperature, pressure and the number of moles, its differential is

<sup>3</sup>See Problem 1.6.

$$dF = \left(\frac{\partial F}{\partial T}\right)_{P n_i} dT + \left(\frac{\partial F}{\partial P}\right)_{T n_i} dP + \sum_i \bar{F}_i dn_i.$$

## 2.2. Partial molar properties

Legendre function	variable set	differential form
$U$	$(S, V, n_i)$	$dU = TdS - PdV + \sum \mu_i dn_i$
$A = U - TS$	$(T, V, n_i)$	$dA = -SdT - PdV + \sum \mu_i dn_i$
$H = U + PV$	$(S, P, n_i)$	$dH = TdS + VdP + \sum \mu_i dn_i$
$G = U - TS + PV$	$(T, P, n_i)$	$dG = -SdT + VdP + \sum \mu_i dn_i$

**Table 2.1:** Summary of Legendre transformations of internal energy.

To integrate this differential along an Euler path, we note that all intensive properties are constant. This implies  $dT = 0$ ,  $dP = 0$ , constant mole fractions, and  $\bar{F}_i = \text{const.}$  along such path. Integration then gives,

$$F = \sum n_i \bar{F}_i. \quad (2.28)$$

This simple relationship also suggests a simple interpretation of partial molar properties: they represent the molar contribution of a chemical species to the extensive property of a mixture. This interpretation applies only to partial molar properties, namely to derivatives taken at fixed *temperature* and *pressure*. For example, between Eqs. (1.23), (2.13), (2.20) and (2.26), only Eq. (2.26) expresses the chemical potential as a *partial molar* property. We may then write

$$\mu_i = \bar{G}_i \quad (2.29)$$

and

$$G = \sum_i n_i \mu_i. \quad (2.30)$$

The last result could have been obtained directly from Eq. (2.23) by Euler integration.

The results of the Legendre approach are summarized in Table 2.1. This is not an exhaustive list of all Legendre transforms that can be derived from internal energy – there are more–, but the three functions defined here,  $A$ ,  $H$  and  $G$ , along with  $U$ , from which they are derived, play a central role and appear frequently in all thermodynamics.

The Legendre transformation has connected four important thermodynamic functions,  $U$ ,  $H$ ,  $A$  and  $G$  in a formal way. Each function comes with a corresponding set of formal variables, which we most easily identify in the differential form of the function (see summary Table 2.1). From these differentials we obtain a number of relationships, first by identifying the partial derivatives in the differential, and then by applying the criterion of exactness between pairs of them. We could summarize them in a table, but they can be just as easily derived on the fly as needed. In fact, an important pedagogical advantage of the Legendre transformation is that it codifies a set of simple rules that can reproduce a large number of key relationships, *without memorization*. The only equation to memorize is the differential of  $dU$ , in Eq. (1.15). To write this differential we had to argue physics – first and second law. Once we have the mathematical expression for  $dU$ , all else is the result of straightforward calculus.

## 2.2. Partial molar properties

---

As a shortcut we will use notation such as  $U_V$ ,  $A_V$  etc. to indicate the partial derivatives,

$$U_V = \left( \frac{\partial U}{\partial V} \right)_{S, n_i}, \quad A_V = \left( \frac{\partial A}{\partial V} \right)_{T, n_i}.$$

In such cases we will understand the derivatives to be within the *proper* set of independent variables. For example, the derivative  $U_V$  is understood to be at fixed  $S$  and  $n_i$  whereas  $A_V$ , though also a derivative with respect to  $V$ , it is taken at fixed  $T$  and  $n_i$ . Similarly, the second order derivative  $U_{SV}$  is understood to refer to

$$U_{SV} = \left( \frac{\partial^2 U}{\partial S \partial V} \right)$$

with  $U = U(S, V, n_i)$ .

### Example 2.2.

---

Identify  $A_V$  with its more common name.

*Solution* From Eq. (2.10) we make the identification,  $A_V = -P$ .

### Example 2.3.

---

Identify  $A_{VT}$  with a first-order derivative.

*Solution* Since  $A_V = -P$ , then

$$A_{VT} = - \left( \frac{\partial P}{\partial T} \right)_{V, n_i}.$$

However, since  $A_{VT} = A_{TV}$  and  $A_T = -S$ , we may also write

$$A_{VT} = A_{TV} = -S_V = - \left( \frac{\partial S}{\partial V} \right)_{T, n_i}.$$

The two results express the Maxwell relationship in Eq. (2.14).

### Example 2.4.

---

Obtain the partial derivatives

$$\left( \frac{\partial H}{\partial T} \right)_{P, n_i}, \quad \left( \frac{\partial H}{\partial P} \right)_{T, n_i}$$

in terms of  $G(T, P, n_i)$ .

*Solution* We write  $H$  in terms of  $G$ :

$$H = G - TG_T$$

and express derivative with respect to  $T$  as

$$\left( \frac{\partial H}{\partial T} \right)_{P, n_i} = G_T - G_T - TG_{TT} = -TG_{TT}.$$

This derivative is the constant-pressure heat capacity, so the final result is

$$C_P = -TG_{TT}. \tag{2.31}$$

For the derivative with respect to  $P$  we have

$$\left(\frac{\partial H}{\partial P}\right)_{T,n_i} = G_{PP} - TG_{PT}.$$

To express these results in a more useful form, we interpret the derivatives on the right-hand side. Noting that the derivative with respect to temperature is the definition of  $C_P$ , we have

$$\left(\frac{\partial H}{\partial T}\right)_{P,n_i} = C_P = -TG_{TT} = T\left(\frac{\partial S}{\partial T}\right)_{P,n_i}. \quad (2.32)$$

Using  $G_{PT} = G_{TP} = (\partial V/\partial T)_{P,n_i}$ , the pressure derivative becomes

$$\left(\frac{\partial H}{\partial P}\right)_{T,n_i} = V - T\left(\frac{\partial V}{\partial T}\right)_{P,n_i}. \quad (2.33)$$

This derivative can be calculated if the equation of state is known.

**Exercise 2.1.** Show that the corresponding derivatives for entropy are

$$\left(\frac{\partial S}{\partial T}\right)_{P,n_i} = -G_{TT} = \frac{C_P}{T}. \quad (2.34)$$

and

$$\left(\frac{\partial S}{\partial P}\right)_{T,n_i} = -\left(\frac{\partial V}{\partial T}\right)_{P,n_i} \quad (2.35)$$

### 2.3 Jacobians

In the calculation of thermodynamic properties we often face the situation where the variables in a partial derivative do not match the base set of the Legendre transform. This requires a transformation of variables, a process that can be facilitated using the Jacobian determinant. Given two functions  $f$  and  $g$  with independent variables  $x$  and  $y$ , the Jacobian determinant (or simply Jacobian) is

$$\frac{\partial(f, g)}{\partial(x, y)} = \begin{vmatrix} \left(\frac{\partial f}{\partial x}\right)_y & \left(\frac{\partial f}{\partial y}\right)_x \\ \left(\frac{\partial g}{\partial x}\right)_y & \left(\frac{\partial g}{\partial y}\right)_x \end{vmatrix} = f_x g_y - f_y g_x. \quad (2.36)$$

In the shorthand notation  $f_x, f_y$ , etc., the partial derivatives are understood to be with respect to the set  $(x, y)$  that appears in the denominator  $\partial(x, y)$  in the Jacobian notation  $\partial(f, g)/\partial(x, y)$ . The Jacobian satisfies the inversion property

$$\frac{\partial(f, g)}{\partial(x, y)} = 1 / \frac{\partial(x, y)}{\partial(f, g)}.$$

and the chain rule:

$$\frac{\partial(f, g)}{\partial(x, y)} = \frac{\partial(f, g)}{\partial(x', y')} \cdot \frac{\partial(x', y')}{\partial(x, y)}. \quad (2.37)$$

If we set  $g = y$  we obtain

$$\frac{\partial(f, y)}{\partial(x, y)} = \left( \frac{\partial f}{\partial x} \right)_y = f_x, \quad (2.38)$$

which gives the partial derivative  $f_x$  in Jacobian notation. Applying the chain rule to the above result we have

$$\frac{\partial(f, y)}{\partial(x, y)} = \frac{\partial(f, y)}{\partial(x', y')} \frac{\partial(x', y')}{\partial(x, y)}, \quad (2.39)$$

which we write as

$$\frac{\partial(f, y)}{\partial(x, y)} = \frac{\partial(f, y)}{\partial(x', y')} \bigg/ \frac{\partial(x, y)}{\partial(x', y')}. \quad (2.40)$$

On the left-hand side we have the the derivative of  $f$  expressed in terms  $(x, y)$ ; on the right-hand side we have the ratio of derivatives expressed in terms of the new variable set  $(x', y')$ . Equation (2.40) is the basis for changing variables in the derivative of multivariate functions.

**Example 2.5.**

Obtain  $C_P$  in terms of  $U(S, V, n_i)$ .

*Solution* We express enthalpy in terms of  $U(S, V, n_i)$  and write the  $C_P$  in the form

$$C_P = \left( \frac{\partial H}{\partial T} \right)_{P, n_i} = \left( \frac{\partial(U - VU_V)}{\partial T} \right)_{P, n_i}.$$

We apply the chain rule with  $(x, y) = (T, P)$  and  $(x', y') = (S, V)$  noting that the  $n_i$  remain constant throughout these differentiations:

$$\begin{aligned} C_P &= \frac{\partial(H, P)}{\partial(T, P)} = \frac{\partial(H, P)}{\partial(S, V)} \bigg/ \frac{\partial(T, P)}{\partial(S, V)} \\ &= \left| \begin{array}{cc} \left( \frac{\partial H}{\partial S} \right)_{V, n_i} & \left( \frac{\partial H}{\partial V} \right)_{S, n_i} \\ \left( \frac{\partial P}{\partial S} \right)_{V, n_i} & \left( \frac{\partial P}{\partial V} \right)_{S, n_i} \end{array} \right| \bigg/ \left| \begin{array}{cc} \left( \frac{\partial T}{\partial S} \right)_{V, n_i} & \left( \frac{\partial T}{\partial V} \right)_{S, n_i} \\ \left( \frac{\partial P}{\partial S} \right)_{V, n_i} & \left( \frac{\partial P}{\partial V} \right)_{S, n_i} \end{array} \right|. \end{aligned}$$

All derivatives on the right-hand side are expressed in terms of  $U(S, V, n_i)$ :

$$\begin{aligned} \left( \frac{\partial H}{\partial S} \right)_{V, n_i} &= U_S - VU_{SV} & \left( \frac{\partial H}{\partial V} \right)_{S, n_i} &= -VU_{VV} \\ \left( \frac{\partial P}{\partial S} \right)_{V, n_i} &= -U_{SV} & \left( \frac{\partial P}{\partial V} \right)_{S, n_i} &= -U_{VV} \\ \left( \frac{\partial T}{\partial S} \right)_{V, n_i} &= U_{SS} & \left( \frac{\partial T}{\partial V} \right)_{S, n_i} &= U_{SV} \end{aligned}$$

The final result is

$$C_P = \frac{U_S}{U_{SS} - (U_{SV})^2 / U_{VV}}. \quad (2.41)$$

**Exercise 2.2.** Show that

$$\frac{1}{C_P} = \frac{1}{C_V} - \frac{(U_{SV})^2}{U_S U_{VV}}.$$

## 2.4 Applied thermodynamics: calculation of properties

A goal of applied thermodynamics is the calculation of thermodynamic properties as a function of temperature, pressure and composition using the minimum possible amount of substance-specific information. The set of equations developed up to this point is not sufficient. These equations provide general relationships between properties that are not substance specific. What is the minimum number of substance specific properties that are required in order to obtain any thermodynamic property as a function of  $(T, P, n_i)$ , and which properties are these? To answer this question we return to the Gibbs energy, whose independent variables are  $T$ ,  $P$  and  $n_i$ . If  $G = G(T, P, n_i)$  is known as a function of its variables, all other properties can be calculated. The differential of the Gibbs energy is

$$dG = G_T dT + G_P dP + \sum_i G_{n_i} dn_i$$

with

$$G_T = -S, \quad G_P = V, \quad G_{n_i} = \mu_i.$$

To integrate it we must know its partial derivatives as a function of  $T$ ,  $P$  and  $n_i$ . This is  $K + 2$  substance specific functions,  $S$ ,  $V$  and  $K$  chemical potentials. The  $K$  chemical potentials are not all independent because they are connected through the Gibbs-Duhem equation. The required number of substance specific properties then is  $K + 1$ : entropy, volume, and  $K - 1$  chemical potentials. These must be known functions of  $T$ ,  $P$  and  $n_i$ , the required information therefore is

$$S(T, P, n_i), \quad V(T, P, n_i), \quad \mu_i(T, P, n_i); \quad i = 1, \dots, K - 1.$$

These substance-specific partial derivatives expressed in terms of temperature, pressure and composition are called *equations of state*, although the term is often reserved for  $V(T, P, n_i)$ , specifically. If we have mathematical expressions for these derivatives we may compute any property. For a pure component  $K = 1$  and the required equations of state are just two: entropy and volume.

The calculation of thermodynamic properties then involves two steps: First, we must identify the equations of state, substance-specific functions that appear as the partial derivatives in the fundamental equation. Once these functions are known, all other properties are obtained using the network of thermodynamics relationships we developed up to this point. In this sense,  $G(T, P, n_i)$  serves as a property generating function, a function from which all other properties arise. Any function in the set  $\{U(S, V, n_i), H(S, P, n_i), A(T, V, n_i), G(T, P, n_i)\}$  or any other function that can be obtained by manipulation of these functions<sup>4</sup> can be used. Obviously, the advantage of using the Gibbs energy is that the result is expressed in terms of temperature, pressure, and the number of moles. The substance-specific equation of state cannot be obtained from the theory we have developed. These can be obtained experimentally or based on physical models of matter.

### 2.4.1 The $V(T, P, n_i)$ equation of state

The relationship between  $V(T, P, n)$  is one of the most studied relationships in experimental thermodynamics. From these studies we know that this relationship is indeed substance specific but has the remarkable property that at low

<sup>4</sup>For example,  $S(U, V, n_i)$ , which is obtained by solving  $U(S, V, n_i)$  for  $S$ , may also be used as a generating function.

pressure it converges to the same asymptotic form for any gas,<sup>5</sup> a form that itself is of remarkable simplicity:

$$\lim_{P \rightarrow 0} \left( \frac{PV}{nRT} \right) = 1, \quad (\text{const. } T, n_i). \quad (2.42)$$

In this limit the equation of state is

$$\frac{V}{n} = \frac{RT}{P}, \quad (2.43)$$

which is the familiar ideal-gas law, with  $R = 8.314 \text{ J/mol K}$ . This limiting form is universal for all gases, pure or mixtures, regardless of composition. In macroscopic thermodynamics the universality of the ideal gas state is an experimental finding; in statistical thermodynamics it is a derived result and a consequence of the discrete nature of matter and the finite range of interactions.

The region of the  $(T, P)$  plane where the ideal gas is obeyed defines the ideal gas state. Mathematically this region is defined by the condition

$$P \rightarrow 0, \quad (\text{const. } T, n_i),$$

or equivalently,

$$V \rightarrow \infty, \quad (\text{const. } T, n_i).$$

Outside this region the equation of state becomes substance specific and its dependence on its variables is complex. It is not possible to write rigorous equations of state that apply over the entire  $(T, P)$  plane. In the region close to the ideal gas state it is possible to write the equation of state as a series expansion with substance specific coefficients. The expansions are written for the compressibility factor,

$$Z = \frac{PV}{nRT} = \frac{Pv}{RT}, \quad (2.44)$$

whose value in the ideal gas state is 1. Expanding  $Z$  at fixed  $T$  and  $n_i$  in power series of  $1/v$  we obtain the virial expansion,

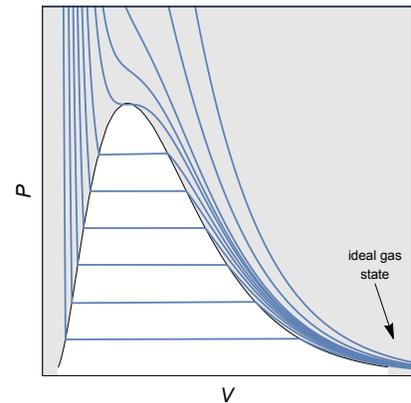
$$Z = 1 + \frac{B_2}{v} + \frac{B_3}{v^2} + \dots \quad (2.45)$$

The virial coefficients  $B_i$  are substance specific and depend on temperature but not on pressure (why?). In the ideal gas limit ( $v \rightarrow \infty$ ) the expansion converges to 1. Outside the ideal gas region its practical applicability is limited by the difficulty of evaluating the virial coefficients higher than the second or third term.

**Semi-empirical equations of state** A difficulty in developing practical equations of state over the entire  $(T, P)$  plane arises from the existence of multiple phases, in particular condensed phases (liquid, solid), whose volume (density) changes very little with pressure and composition. It is possible, however, to develop equations that cover the gas/liquid region of the phase diagram with sufficient accuracy. The prototype for this approach is the van der Waals equation:

$$P = \frac{RT}{v-b} - \frac{a}{v^2}, \quad (2.46)$$

<sup>5</sup>At sufficiently low pressure any substance regardless of composition is in the gas phase.



**Figure 2.2:**  $PVT$  behavior in the vapor-liquid region. Lines are contours of constant temperature.

with  $a$  and  $b$  substance specific constants. Qualitatively, parameter  $b$  accounts for hard-core repulsion between molecules that prevents the volume of the system of ever becoming zero. The term  $-a/v^2$  accounts for attraction between molecules, which effectively reduces the total pressure. The simultaneous presence of attraction/repulsion allow this fairly simple equation to mimic the behavior of a vapor/liquid mixture and to exhibit a critical point and vapor/liquid equilibrium. The van der Waals equation is not accurate enough for technical calculations but it has inspired the development of variations that work quite well for certain molecules, typically small symmetric non polar molecules. Examples of such equations are the Soave-Redlich-Kwong (SRK) and the Peng-Robinson (PR). The reader is referred to undergraduate textbooks for more information.

### 2.4.2 Properties at constant composition

First we will consider systems of constant composition ( $dn_i = 0$  for all components). This encompasses the case of pure component but the results are also valid for mixtures as long as their composition does not change. First we write the differential of enthalpy (entropy works the same way) in the general form

$$dH = \left(\frac{\partial H}{\partial T}\right)_{P,n_i} dT + \left(\frac{\partial H}{\partial P}\right)_{T,n_i} dP \quad (\text{const. } n_i). \quad (2.47)$$

The partial derivatives were obtained in Example 2.4. The temperature derivative is the constant-pressure heat capacity,  $C_P(T, p, n_i)$ ; the pressure derivative is

$$\left(\frac{\partial H}{\partial P}\right)_{T,n_i} = V - T \left(\frac{\partial V}{\partial T}\right)_{P,n_i}, \quad [2.33]$$

whose calculation requires the equation of state  $V(T, P, n_i)$ . Formally, the enthalpy is

$$dH = C_P dT + \left[ V - T \left(\frac{\partial V}{\partial T}\right)_{P,n_i} \right] dP, \quad (\text{const. } n_i). \quad (2.48)$$

The corresponding result for entropy is (see Exercise 2.1)

$$dS = \frac{C_P}{T} dT - \left(\frac{\partial V}{\partial T}\right)_{P,n_i} dP \quad (\text{const. } n_i). \quad (2.49)$$

Since composition is constant we may set  $v = V/n$ ,  $h = H/n$ ,  $s = S/n$ ,  $c_p = C_P/n$  and write these results in the equivalent form

$$dh = c_p dT + \left[ v - T \left(\frac{\partial v}{\partial T}\right)_{P,x_i} \right] dP;$$

$$ds = \frac{c_p}{T} dT - \left(\frac{\partial v}{\partial T}\right)_{P,x_i} dP.$$

In general, all results for the extensive properties under constant composition can be written for the corresponding molar properties.<sup>6</sup>

The calculation of enthalpy and entropy from Eqs. (2.48) and (2.49) requires two substance-specific properties: the constant pressure heat capacity,

<sup>6</sup>Molar properties re functions of temperature, pressure and mole fractions rather than number of moles. This is indicated in the notation of the partial derivatives.

## 2.4. Applied thermodynamics: calculation of properties

$c_p(T, P, x_i)$ ; and the molar volume,  $V(T, P, x_i)$ . For mixtures these functions depend not only on the chemical identity of the components but also on their composition. There is no need to write special equations for any other property because they may be obtained by their relationship to  $c_p$ ,  $v$ ,  $h$  and  $s$ . For example, for the internal energy we have

$$u = h - pv,$$

and similarly for any other property.

### 2.4.3 Properties in the ideal gas state

In the ideal gas state the volumetric equation of state is  $v = RT/P$  and the temperature derivative derivative is

$$\left(\frac{\partial v^{\text{ig}}}{\partial T}\right)_{T, x_i} = \frac{R}{P}. \quad (2.50)$$

From Eqs. (2.33) and (2.35) we then have<sup>7</sup>

$$\left(\frac{\partial h^{\text{ig}}}{\partial P}\right)_{T, x_i} = 0, \quad \left(\frac{\partial s^{\text{ig}}}{\partial P}\right)_{T, x_i} = -\frac{R}{P}. \quad (2.51)$$

<sup>7</sup>The superscript <sup>ig</sup> identifies properties in the ideal-gas state.

The differential forms of enthalpy and entropy become

$$dh^{\text{ig}} = c_p^{\text{ig}}, \quad (2.52)$$

$$ds^{\text{ig}} = \frac{c_p^{\text{ig}}}{T} dT - \frac{R}{P} dP \quad (2.53)$$

and are easily integrated from reference state  $(T_0, P_0)$  to  $(T, P)$ :

$$h^{\text{ig}} = h_0 + \int_{T_0}^T c_p^{\text{ig}}, \quad (2.54)$$

$$s^{\text{ig}} = s_0 + \int_{T_0}^T \frac{c_p^{\text{ig}}}{T} dT - R \ln \frac{P}{P_0}. \quad (2.55)$$

#### Example 2.6. (Ideal gas heat capacity)

Show that the constant-pressure and constant-volume heat capacities in the ideal gas state are independent of pressure and depend only on temperature.

**Solution** It suffices to show that pressure derivatives of  $c_p^{\text{ig}}(T, P, x_i)$  and  $c_v^{\text{ig}}(T, P, x_i)$  are zero. The pressure derivative of  $c_p^{\text{ig}}$  is

$$\left(\frac{\partial c_p^{\text{ig}}}{\partial P}\right)_{T, x_i} = \frac{\partial}{\partial P} \left(\frac{\partial h^{\text{ig}}}{\partial P}\right)_{T, x_i} = 0,$$

which follows from the fact that the pressure derivative of  $h^{\text{ig}}$  is zero. For the constant-volume heat capacity we have

$$\left(\frac{\partial c_v^{\text{ig}}}{\partial P}\right)_{T, x_i} = \frac{\partial}{\partial P} \left(\frac{\partial (h^{\text{ig}} - pv^{\text{ig}})}{\partial P}\right)_{T, x_i} = \frac{\partial}{\partial P} \left(\frac{\partial (h^{\text{ig}} - RT)}{\partial P}\right)_{T, x_i} = 0.$$

**Exercise 2.3.** Show that

$$\left( \frac{\partial c_V^{\text{ig}}}{\partial V} \right)_{T, x_i} = 0.$$

**Ideal gas heat capacity** The constant-pressure heat capacity of pure components is tabulated as a function of temperature for a large number of pure components and is represented by a fitted equation with substance-specific coefficients (see for example the [NIST web book](#)).

#### 2.4.4 Residual properties

Among the many important aspects of the ideal gas state is the fact that the calculation of properties requires only one substance-specific equation of state: the heat capacity—the volumetric equation of state is universal. The fact that this heat capacity is a function of temperature only further simplifies calculations. Indeed all properties in the ideal gas state are straightforward to compute. It is natural to view the ideal gas state as a reference from which to calculate properties at other states by developing corrections that must be added to the ideal-state results. This approach makes use of *residual properties*. A residual property  $F^R$  is defined as the difference between the actual property  $F$  and the hypothetical ideal-gas property  $F^{\text{ig}}$  at the same temperature, pressure and composition:

$$F^R(T, P, n_i) = F(T, P, n_i) - F^{\text{ig}}(T, P, n_i). \quad (2.56)$$

By hypothetical ideal-gas property we refer to the property that is obtained at given  $T, P, n_i$ , using the equations of the ideal-gas state. The term “hypothetical” emphasizes the fact that we use the ideal-gas equation regardless of whether state  $(T, P, n_i)$  is in the ideal gas region or not. The goal is to develop expression for the residual properties. Once  $F^R$  is known the actual property is obtained by adding  $F^{\text{ig}}$ , whose calculation requires only the ideal-gas heat capacity. We will develop equations for the residual enthalpy and entropy. It is not necessary to write expressions for other residual properties because these can be easily obtained by their relationship to  $h^R$ ,  $s^R$  and  $v^R$ .<sup>8</sup> The pressure derivative of the residual enthalpy is

<sup>8</sup>The residual volume is simply

$$v^R = v - \frac{RT}{P} = \frac{RT}{P}(Z - 1).$$

$$\left( \frac{\partial h^R}{\partial P} \right)_{T, x_i} = \left( \frac{\partial h}{\partial P} \right)_{T, x_i} - \left( \frac{\partial h^{\text{ig}}}{\partial P} \right)_{T, x_i} = v - T \left( \frac{\partial v}{\partial T} \right)_{P, x_i}, \quad (2.57)$$

where we have used

$$\left( \frac{\partial h}{\partial P} \right)_{T, x_i} = v - T \left( \frac{\partial v}{\partial T} \right)_{P, x_i}; \quad \left( \frac{\partial h^{\text{ig}}}{\partial P} \right)_{T, x_i} = 0.$$

Equation (2.57) is integrated from the ideal-gas state ( $P = 0, h^R = 0$ ) to arbitrary state  $(P, h^R)$  at fixed temperature and composition:

$$h^R = \int_0^P \left( v - T \left( \frac{\partial v}{\partial T} \right)_{P, x_i} \right) dP, \quad (\text{const. } T, x_i). \quad (2.58)$$

## 2.4. Applied thermodynamics: calculation of properties

**Exercise 2.4.** Show that the residual entropy is

$$s^R = \int_0^P \left( \frac{R}{P} - \left( \frac{\partial v}{\partial T} \right)_{P, x_i} \right) dP, \quad (\text{const. } T, x_i). \quad (2.59)$$

The calculation of the residual enthalpy and entropy in Eqs. (2.58) and (2.59) requires the volumetric equation of state; this equation alone is sufficient to completely specify these residual properties. The calculation of  $h$  and  $s$  is completed by combining the ideal-gas and residual terms:

$$h = h^{\text{ig}} + h^R = h_0 + \int_{T_0}^T c_P^{\text{ig}} + \int_0^P \left( v - T \left( \frac{\partial v}{\partial T} \right)_{P, x_i} \right) dP, \quad (2.60)$$

$$s = s^{\text{ig}} + s^R = s_0 + \int_{T_0}^T \frac{c_P^{\text{ig}}}{T} dT - R \ln \frac{P}{P_0} + \int_0^P \left( \frac{R}{P} - \left( \frac{\partial v}{\partial T} \right)_{P, x_i} \right) dP, \quad (2.61)$$

with the integrals in pressure taken under fixed temperature and composition.

**Reference states** The terms  $h_0$  and  $s_0$  in Eqs. (2.60) and (2.61) are integration constants and drop out when we calculate differences. For tabulation purposes their value must be specified but this can be done arbitrarily by convention since their choice has no effect on the calculation of differences. Two main conventions are in use (but others are possible): *the actual state* and the *hypothetical ideal gas state* at specified temperature, pressure and composition. We first examine the simpler case of pure component.

First we write Eqs. (2.60) and (2.61) in the equivalent form

$$h = h_0 + \Delta h^{\text{ig}} \Big|_{\text{ref}} + h^R - h_0^R \quad (2.62)$$

$$s = s_0 + \Delta s^{\text{ig}} \Big|_{\text{ref}} + s^R - s_0^R \quad (2.63)$$

where  $h$  and  $s$  refer to state  $(T, P)$ ,  $h_0$ ,  $h_0^R$ ,  $s_0$  and  $s_0^R$  refer to reference state  $(T_0, P_0)$  and the differences  $\Delta h^{\text{ig}}$ ,  $\Delta s^{\text{ig}}$  are calculated from the reference state to the state of interest. Under the “actual state” convention we set both  $h_0$  and  $s_0$  to zero and the equations for enthalpy and entropy simplify to

$$h = \Delta h^{\text{ig}} \Big|_{\text{ref}} + h^R - h_0^R$$

$$s = \Delta s^{\text{ig}} \Big|_{\text{ref}} + s^R - s_0^R.$$

Under the “hypothetical ideal gas state” convention we set  $h_0 = h_0^R$  and  $s_0 = s_0^R$ . In this case the equations for enthalpy and entropy take even simpler forms:

$$h = \Delta h^{\text{ig}} \Big|_{\text{ref}} + h^R$$

$$s = \Delta s^{\text{ig}} \Big|_{\text{ref}} + s^R.$$

It is trivial to confirm that under both conventions the expressions for  $\Delta h$  and  $\Delta s$  between any two states are the same.

For mixtures we must specify a reference state for each component. We will return to this question when we discuss the effect of composition.

**Exercise 2.5.** Show that under the “hypothetical ideal-gas” convention the ideal-gas enthalpy and entropy at the reference state is zero, i.e.,  $h_0^{\text{ig}} = 0$  and  $s_0^{\text{ig}} = 0$ . This gives the convention its name. The term “hypothetical” refers to the fact that by “ideal gas” we refer to the ideal-gas *equations*, not to the physical state of the system.

### 2.4.5 Effect of composition

The residual properties in Eqs. (2.58) and (2.59) and the corresponding expressions for enthalpy and entropy in Eqs. (2.60) and (2.61) apply to both pure components and mixtures, but to use them with mixtures we must provide expressions for the the enthalpy and entropy of mixture in the ideal-gas state. This requires physical models of matter that go beyond the scope of macroscopic thermodynamics. We write these equations here without justification and defer their derivation until statistical treatment:

$$h^{\text{ig}} = \sum_i x_i h_i^{\text{ig}} \quad (2.64)$$

$$s^{\text{ig}} = \sum_i x_i s_i^{\text{ig}} - R \sum_i x_i \log x_i. \quad (2.65)$$

Here  $h^{\text{ig}}$  ( $s^{\text{ig}}$ ) is the enthalpy (entropy) of the mixture at  $(T, P, x_i)$  and  $h_i^{\text{ig}}$  ( $s_i^{\text{ig}}$ ) is the enthalpy (entropy) of the pure components at  $(T, P)$ . Accordingly, the enthalpy of the mixture is the same as the sum of the enthalpies of the pure components at the same temperature and pressure while the entropy of the mixture is larger than the sum of the pure components. The difference between the entropy of the mixture and the entropy of the pure components is the ideal entropy of mixing:

$$\Delta S_{\text{mix}} = -R \sum_i x_i \ln x_i. \quad (2.66)$$

The properties of mixture are calculated by combining the ideal-gas properties with the corresponding residuals. The residual properties of mixture require an equation of state that takes into consideration the effect of composition. The same equations of state that are used for pure components can be applied to mixtures by treating the various parameters that appear in these equations as functions of composition.<sup>9</sup>

**Reference states for mixtures** The simplest reference state in the case of mixtures is the hypothetical ideal gas state. First we write the enthalpy and entropy of mixture in the form

$$h = \sum_i h_i^{\text{ig}} + h^R \quad (2.67)$$

$$s = \sum_i s_i^{\text{ig}} - R \sum_i x_i \ln x_i + s^R \quad (2.68)$$

with

$$h_i^{\text{ig}} = h_{i0}^{\text{ig}} + \Delta h_i^{\text{ig}} \Big|_{\text{ref}}$$

$$s_i^{\text{ig}} = s_{i0}^{\text{ig}} + \Delta s_i^{\text{ig}} \Big|_{\text{ref}}$$

<sup>9</sup>A mixing rule is recipe for the calculation of the parameters of an equation of state from those of the pure components. A common mixing rule for the parameters of the van der Waals equation is

$$a = \sum_i \sum_j x_i x_j \sqrt{a_i a_j}$$

$$b = \sum_i b_i$$

where  $a$  and  $b$  refer to the mixture and  $a_i$  and  $b_i$  to the pure components. Several other mixing rules are in use.

with the terms  $\Delta h_i^{\text{ig}}$  and  $\Delta s_i^{\text{ig}}$  of the pure components calculated from the reference state at  $(T_0, P_0)$ . The simplest reference state in this case in the hypothetical ideal gas, which sets  $h_{i0}^{\text{ig}} = 0$  and  $s_{i0}^{\text{ig}} = 0$  for all components at the reference state.

**Exercise 2.6.** Under the “actual state” convention we set the enthalpy and entropy of all pure components at the reference state to zero ( $h_{i0} = 0, s_{i0}$ ). Show that this is equivalent to setting  $h_{i0}^{\text{ig}} = -h_{i0}^R$  and  $s_{i0}^{\text{ig}} = -s_{i0}^R$  in Eqs. (2.67) and (2.68) where  $h_{i0}^{\text{ig}}$  ( $s_{i0}^{\text{ig}}$ ) is the residual enthalpy (residual entropy) of the pure component at the reference state.

## 2.5 Additional reading

A concise outline of the Legendre transformation is given in [McQuarrie \(2001\)](#).

A good source for the mathematical treatment of the fundamental differential, the application of Legendre transformations to thermodynamic properties and the mathematical properties of the Jacobian transformation is given in [Tester and Modell \(1996\)](#).

Residual properties and their use in the calculation of the properties of pure components and their mixtures via equations of state are discussed in most undergraduate textbooks. See for example [Matsoukas \(2013\)](#), chapters 5 and 9.

## 2.6 Problems

**Problem 2.1.** Express the following properties as a function of  $G(P, T, n_i)$  and its derivatives:

- Equation of state.
- Internal energy  $U$ .
- Enthalpy.

**Problem 2.2.** a) Express  $U$ ,  $A$  and  $G$  as Legendre transforms of  $H(S, P, n_i)$ .

b) Express the  $C_V$  in terms of  $H(S, P, n_i)$ .

**Problem 2.3.** a) Simplify the derivative

$$\left( \frac{\partial(A + H_S G_T)}{\partial T} \right)_{V, n_i}$$

b) Evaluate the derivative  $G_{PTT}$  in the ideal-gas state.

c) Evaluate the derivative  $A_{TVT}$  in the ideal-gas state.

d) Evaluate the derivative

$$\left( \frac{\partial H}{\partial V} \right)_{T, n_i}$$

for a gas that obeys the equation of state  $P(V - nb) = nRT$ , where  $b$  is constant.

**Problem 2.4.** Identify the following thermodynamic functions by their more common names:

- $G - U_S G_T$ .
- $G + U_S A_T + U_V G_P$ .
- $U_S G_{TT}$ .

## 2.6. Problems

**Problem 2.5.** a) For a one-component system ( $n_1 = n$ ), show:

$$\left(\frac{\partial \mu}{\partial n}\right)_{T,V} = -\frac{V}{n} \left(\frac{\partial \mu}{\partial V}\right)_{T,n}; \quad \left(\frac{\partial P}{\partial n}\right)_{T,V} = -\frac{V}{n} \left(\frac{\partial P}{\partial V}\right)_{T,n}.$$

b) Obtain  $(\partial \mu / \partial N)_{T,V}$  for a substance that obeys the van der Waals equation.

**Problem 2.6.** a) Obtain the Legendre transform of  $U(S, V, N)$  with independent variables  $T, V, \mu_i$ . Identify the corresponding thermodynamic function and obtain its partial derivatives with respect to these variables.

b) Devise an experimental system that communicates with its surroundings while maintaining  $T, V$  and  $\mu_i$  constant.

**Problem 2.7.** Express the function  $F = TS - PV$  as a Legendre transform of  $U(S, V, n_i)$ . Identify its proper set of variables and write its differential in terms of these variables.

**Problem 2.8.** a) Obtain the differential of  $S(U, V, n_i)$ , identify the partial derivatives in the differential and write the corresponding Maxwell relationships.

b) Show that  $A/RT$  can be expressed as a Legendre transform of  $S(U, V, n_i)$  and obtain its differential in the form

$$d \frac{A}{RT} = U d \left( \frac{1}{RT} \right) - \frac{P dV}{RT} + \sum \frac{\mu_i}{RT} d n_i.$$

**Problem 2.9.** Express the following properties as a function of  $G(P, T, n_i)$  and its derivatives:

- a) Equation of state.
- b) Internal energy  $U$ .
- c) Enthalpy.

**Problem 2.10.** a) Express  $U, A$  and  $G$  as Legendre transforms of  $H(S, P, n_i)$ .

b) Express the  $C_V$  in terms of  $H(S, P, n_i)$ .

**Problem 2.11.** a) Simplify the derivative

$$\left( \frac{\partial(A + H_S G_T)}{\partial T} \right)_{V, n_i}$$

b) Evaluate the derivative  $G_{PTT}$  in the ideal-gas state.

c) Evaluate the derivative  $A_{TVT}$  in the ideal-gas state.

d) Evaluate the derivative

$$\left( \frac{\partial H}{\partial V} \right)_{T, n_i}$$

for a gas that obeys the equation of state  $P(V - nb) = nRT$ , where  $b$  is constant.

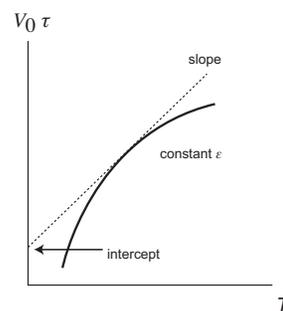
**Problem 2.12.** Identify the following thermodynamic functions by their more common names:

- a)  $G - U_S G_T$ .
- b)  $G + U_S A_T + U_V G_P$ .
- c)  $U_S G_{TT}$ .

**Problem 2.13.** a) For a one-component system ( $n_1 = n$ ), show:

$$\left(\frac{\partial \mu}{\partial n}\right)_{T,V} = -\frac{V}{n} \left(\frac{\partial \mu}{\partial V}\right)_{T,n}; \quad \left(\frac{\partial P}{\partial n}\right)_{T,V} = -\frac{V}{n} \left(\frac{\partial P}{\partial V}\right)_{T,n}.$$

b) Obtain  $(\partial \mu / \partial N)_{T,V}$  for a substance that obeys the van der Waals equation.



**Figure 2.3:** See problem 2.18.

**Problem 2.14.** a) Obtain the Legendre transform of  $U(S, V, N)$  with independent variables  $T, V, \mu_i$ . Identify the corresponding thermodynamic function and obtain its partial derivatives with respect to these variables.

b) Devise an experimental system that communicates with its surroundings while maintaining  $T, V$  and  $\mu_i$  constant.

**Problem 2.15.** Express the function  $F = TS - PV$  as a Legendre transform of  $U(S, V, n_i)$ . Identify its proper set of variables and write its differential in terms of these variables.

**Problem 2.16.** a) Obtain the differential of  $S(U, V, n_i)$ , identify the partial derivatives in the differential and write the corresponding Maxwell relationships.

b) Show that  $A/RT$  can be expressed as a Legendre transform of  $S(U, V, n_i)$  and obtain its differential in the form

$$d \frac{A}{RT} = U d \left( \frac{1}{RT} \right) - \frac{P dV}{RT} + \sum \frac{\mu_i}{RT} dn_i.$$

**Problem 2.17.** a) If  $h(a, y)$  is the Legendre transformation of  $f(x, y)$  with respect to  $x$ , show that its second derivatives are:

$$h_{aa} = -\frac{1}{f_{xx}}$$

$$h_{yy} = f_{yy} - \frac{f_{xy}^2}{f_{xx}}$$

b) Show that if  $f(x, y)$  is a concave surface than  $h(a, y)$  is convex function of  $a$  and a concave function of  $y$ .

**Problem 2.18.** Elasticity is the property of certain materials such as rubber to deform reversibly under the action of a tensile force. The internal energy of a linear elastic body is governed by the equation

$$dU = T dS + V_0 \tau d\epsilon$$

where  $\tau = F/A_0$  is the tensile stress,  $\epsilon = (L - L_0)/L_0$  is the strain (fractional elongation), and  $V_0, A_0, L_0$  are the equilibrium values for the volume, cross sectional area and length of the elastic body ( $V_0 = L_0 A_0$ ).

a) Express the stress  $\tau$  in terms of the following two derivatives:

$$\left( \frac{\partial U}{\partial \epsilon} \right)_T, \quad \left( \frac{\partial S}{\partial \epsilon} \right)_T$$

b) The graph in Fig. 2.3 shows the stress as a function of temperature at constant strain. Relate the slope and the intercept of a tangent line on this graph in terms of  $U, S$ , and their derivatives.

c) It is found experimentally that the relationship between stress and temperature at constant  $\epsilon$  for a certain rubber is described adequately over a range of temperatures, stresses and strains by the equation

$$\tau = 5 \times 10^3 T$$

with  $T$  in Kelvin and  $\tau$  in Pa. Determine the entropy change of the rubber for an isothermal elongation from equilibrium length to  $\epsilon = 2$ .

**Problem 2.19.** The entropy of a pure component is given by

$$s = n \log \left\{ \left( \frac{v}{n} \right)^a \left( \frac{u}{n} \right)^b \right\},$$

## 2.6. Problems

where  $n$  is the number of moles,  $s$ ,  $u = U/U_0$  is dimensionless internal energy,  $v = V/V_0$  is dimensionless volume, and  $U_0$ ,  $V_0$ ,  $a$  and  $b$  are constants.

a) Obtain the following dimensionless properties as functions of  $u$ ,  $v$  and  $n$ : temperature  $t = T/(U_0/k)$ ; pressure  $p = P/(U_0/V_0)$ ; heat capacity  $c_v = C_V/k$ ; enthalpy  $h = H/U_0$ ; Helmholtz energy  $a = A/U_0$ ; Gibbs energy  $g = G/U_0$ .

**Problem 2.20.** The internal energy of a pure gas ( $n = 1$ ) is given by the following equation:

$$U = a + b \left\{ \left( \frac{c}{V-d} \right)^f \exp \left( \frac{S-g}{h} \right) - 1 \right\}$$

where  $V$  is the volume,  $S$  is the entropy, and  $a$ ,  $b$ ,  $c$ ,  $d$ ,  $f$ ,  $g$  and  $h$  are constants.

a) Obtain the equation of state.

b) Obtain  $C_V$ .

c) Write the internal energy of this substance in the form  $U = U(S, V, n)$  for any number of mole  $n$ , where  $S$  and  $V$  are the extensive entropy and volume.

Note: You may find it helpful to do the calculations on Mathematica in order to facilitate otherwise tedious calculus.

**Problem 2.21.** The ideal-gas heat capacity of argon °C is  $2.5R$  and is nearly independent of temperature. Estimate the actual heat capacity at 100 bar,  $-48$  °C. Assume that argon at these conditions is described by the truncated virial equation and that the second virial coefficient is given by Eq. (8-68 in Tester & Modell. Is it appropriate to use the truncated virial equation at these conditions?

**Problem 2.22.** a) Show how the derivative

$$\left( \frac{\partial C_P}{\partial P} \right)_{T, n_i}$$

can be calculated from the equation of state.

b) Obtain the difference  $C_P - C_V$  in terms of quantities that can be calculated from the equation of state.

c) Apply these equations to the ideal-gas state and confirm that  $C_P^{\text{ig}}$  is independent of pressure and that  $C_P^{\text{ig}} - C_V^{\text{ig}} = nR$ .

**Problem 2.23.** a) Show that the  $C_P$  may be expressed in the form

$$C_P = \frac{U_S U_{VV}}{U_{SS} U_{VV} - (U_{SV})^2}.$$

b) Derive the relationship

$$\frac{1}{C_V} - \frac{1}{C_P} = \frac{(U_{SV})^2}{U_S U_{VV}}$$

**Problem 2.24.** Obtain the partial derivatives

$$\left( \frac{\partial S}{\partial T} \right)_{P, n_i} \quad \text{and} \quad \left( \frac{\partial S}{\partial P} \right)_{T, n_i}$$

in terms of  $G(T, P, n_i)$ .

b) Show that the entropy of pure component ( $n_1 = 1$ ,  $n_{i>1} = 0$ ) can be written in the form

$$dS = \frac{C_P}{T} dT - \left( \frac{\partial V}{\partial T} \right)_{P, n_i} dP$$

c) Apply the result to the ideal-gas state to obtain  $\Delta S^{\text{ig}}$  between states  $(T_1, P_1)$  and  $(T_2, P_2)$ .

## 2.6. Problems

**Problem 2.25.** Obtain  $C_V$  in terms of  $A(S, V, n_i)$  and its derivatives. Show that this is the same result as that obtained from  $U(S, V, n_i)$ .

**Problem 2.26.** a) Express the derivative  $(\partial S/\partial T)_{G,n}$  using as independent variables the set  $(T, P, n)$ .

b) Express the derivative  $(\partial A/\partial G)_{T,n}$  using as independent variables the set  $(T, P, n)$ .

**Problem 2.27.** a) Obtain the derivative  $U_{SV}$  in terms of quantities that can be calculated from the  $C_P$  and the equation of state.

b) Evaluate the above derivative in the ideal-gas state.

**Problem 2.28.** a) Obtain the enthalpy and entropy of a mixture that obeys the equation of state

$$P(V - \sum n_i b_i) = nRT,$$

where  $b_i$  are constants for each component and  $n = \sum n_i$ . The heat capacities of the pure components are known and the reference enthalpies and entropies of all pure components are zero at  $T_0$ .

b) As a demonstration, calculate the enthalpy and entropy of a binary mixture with the following properties:

$$T = 420 \text{ K}, P = 20 \text{ bar},$$

$$n_1 = 0.25, n_2 = 0.75,$$

$$b_1 = -0.000075 \text{ m}^3/\text{mol}, b_2 = -0.00021 \text{ m}^3/\text{mol}$$

$$c p_1^{\text{ig}} = 30 \text{ J/mol K}, c p_2^{\text{ig}} = 80 \text{ J/mol K},$$

$$T_0 = 300 \text{ K}, P_0 = 1 \text{ bar}$$

**Problem 2.29.** a) Obtain the residual  $C_P$  of a mixture in terms of the equation of state.

b) Calculate the residual heat capacity of a mixture that obeys the van der Waals equation with the following mixing rules:

$$a = \sqrt{a_1 a_2}, \quad b = b_1 + b_2$$

where  $a_i$  and  $b_i$  are the van der Waals parameters of the pure components.

**Problem 2.30.** a) Obtain the partial derivative

$$\left( \frac{\partial H^R}{\partial V} \right)_{T, n_i}$$

in terms of the equation of state.

b) Show that the residual enthalpy is given by

$$H^R = PV - RT + \int_{\infty}^V \left[ T \left( \frac{\partial P}{\partial T} \right)_{V, n_i} - P \right] dV,$$

with the integration conducted at fixed temperature and composition.

**Problem 2.31.** a) Obtain the partial derivative

$$\left( \frac{\partial S^R}{\partial V} \right)_{T, n_i}$$

in terms of the equation of state.

b) Show that the residual enthalpy is given by

$$S^R = \log Z + \int_{\infty}^V \left[ \left( \frac{\partial P}{\partial T} \right)_{V, n_i} - \frac{R}{V} \right] dV$$

with the integration conducted at fixed temperature and composition.

**Problem 2.32.** An ideal gas mixture at  $(T, V, n_i)$  is formed by mixing the pure components in the following manner. All components are initially at temperature  $T$  and volume  $V$ , the same as the final mixture. Next, the pressure of all components is adjusted reversibly to the pressure of the final mixture. Finally, components are allowed to mix at fixed  $T$  and  $P$ . Show that

$$\begin{aligned} U^{\text{ig}}(T, V, n_i) &= \sum U_i^{\text{ig}} \\ S^{\text{ig}}(T, V, n_i) &= \sum S_i^{\text{ig}} \end{aligned}$$

where  $U_i$  and  $S_i$  are the properties of the pure components at  $T, V$ . The result states that the internal energy and entropy of the mixture is equal to that of the pure components before mixing. How do we reconcile this result with the fact that mixing of ideal-gas components is an irreversible process?

**Problem 2.33.** The residual properties are based on a state defined by  $(T, P, n_i)$ . The set  $(T, V, n_i)$  can also serve as a basis to produce a different set of residual properties that may also be calculated from the equation of state. First, obtain the derivatives

$$\begin{aligned} \left( \frac{\partial U^r}{\partial V} \right)_{T, n_i} &= T \left( \frac{\partial P}{\partial T} \right)_{V, n_i} - P \\ \left( \frac{\partial S^r}{\partial V} \right)_{T, n_i} &= \left( \frac{\partial P}{\partial T} \right)_{V, n_i} - \frac{nR}{V} \end{aligned}$$

Next, define a new residual property  $F^r(T, V, n_i)$  –not to be confused with  $F^R(T, P, n_i)$ – as

$$F^r(T, V, n_i) = F(T, V, n_i) - F^{\text{ig}}(T, V, n_i),$$

and show that

$$F^r(T, V, n_i) = \int_{\infty}^V \left( \frac{\partial F^r}{\partial V} \right)_{T, n_i} dV \quad (\text{constant } T, n_i).$$

Finally, show that the properties of ideal-gas mixture can be written as

$$\begin{aligned} U(T, V, n_i) &= \sum n_i u_i^{\text{ig}}(T, V) + n \int_{\infty}^v \left[ T \left( \frac{\partial P}{\partial T} \right)_{v, x_i} - P \right] dv \\ S(T, V, n_i) &= \sum n_i s_i^{\text{ig}}(T, V) + n \int_{\infty}^v \left[ \left( \frac{\partial P}{\partial T} \right)_{v, x_i} - \frac{R}{v} \right] dv \end{aligned}$$

where  $u, s$  and  $v$  are the molar energy, entropy and volume, respectively,  $x_i = n_i/n$  is the mol fraction, and with both integrals calculated on a path of constant  $T$  and  $x_i$ .

**Problem 2.34.** Obtain the second virial coefficient for the following equations of state:

$$\begin{aligned} P &= \frac{RT}{V-b} - \frac{a}{V^2} && \text{van der Waals} \\ P &= \frac{RT}{V-b} - \frac{a(T)}{V(V-b)} && \text{Soave-Redlich-Kwong} \\ P &= \frac{RT}{V-b} - \frac{a(T)}{V^2 + 2Vb - b^2} && \text{Peng-Robinson} \end{aligned}$$

*Hints* Method 1: Show that  $B = \lim_{V \rightarrow \infty} V(Z-1)$ . Method 2: expand  $Z$  in powers of  $1/V$ .

**Problem 2.35.** The following equation of state is proposed:

$$P = \frac{RT}{V-b} - \frac{a}{V^2+c}$$

where  $a$ ,  $b$  and  $c$  are constants specific to the fluid.

- Show that this equation goes to the ideal-gas law in the ideal-gas limit ( $P \rightarrow 0$  at  $T = \text{const.}$ ).
- Obtain  $a$ ,  $b$  and  $c$  in terms of the critical parameters of the fluid.

**Problem 2.36.** a) The coefficients of isobaric thermal expansion,  $\beta$ , and isothermal compression,  $\kappa$ , are defined as follows:

$$\beta = \frac{1}{V} \left( \frac{\partial V}{\partial T} \right)_P, \quad \kappa = -\frac{1}{V} \left( \frac{\partial V}{\partial P} \right)_T$$

Obtain an expression for the two coefficients for a substance that obeys the van der Waals equation of state:

$$P = \frac{RT}{V-b} - \frac{a}{V^2}$$

where  $a$  and  $b$  are constant.

- A tank is completely filled with liquid octane at 25 °C 1 bar, and is then heated until the temperature increases by 5 °C. Assuming that octane obeys the van der Waals equation, calculate the pressure that develops as a result of the heating. State your assumptions clearly!

The van der Waals parameters of octane can be estimated from the critical temperature and pressure using the following equations:

$$a = \frac{27R^2T_c^2}{64P_c}, \quad b = \frac{RT_c}{8P_c}$$

## Equilibrium and Stability

The equilibrium properties of a thermodynamic system are obtained by applying the second law in its equilibrium form as an exact equality. The general statement of this law is the inequality,

$$[dS]_{UVn_i} \geq 0, \quad (3.1)$$

which we may also write in integrated form as

$$[\Delta S]_{UVn_i} \geq 0. \quad (3.2)$$

The inequality of the second law serves as a compass that points to equilibrium. Between two systems with the same energy, volume and number of moles, the system with the higher entropy is closer to equilibrium and that with the lower entropy is further away. The road to equilibrium is in the direction of increasing entropy. It is understood that the statement refers to a system that is partitioned into a number of non interacting, equilibrium parts. We use this principle to develop criteria for equilibrium by applying it to microcanonical partitioning discussed on p. 10. We begin with a microcanonical system  $(U, V, n_i)$  in equilibrium, such that its entropy is  $S(U, V, n_i)$ . We then partition it into  $\mathcal{K}$  noninteracting microcanonical parts  $(U_k, V_k, n_{ik})$  that satisfy the conservation conditions

$$U = \sum_i U_k, \quad V = \sum_i V_k, \quad n_i = \sum_i n_{ik}. \quad (3.3)$$

Since the parts are non interacting, the total entropy of the partition is the sum of the parts:

$$\sum_k S(U_k, V_k, n_{ik}). \quad (3.4)$$

The entropy change  $\Delta S$  that results from this redistribution is bound by the inequality of Eq. (3.2):

$$[\Delta S]_{UVn_i} = S(U, V, n_i) - \sum_k S(U_k, V_k, n_{ik}) \geq 0, \quad (3.5)$$

or

$$S(U, V, n_i) \geq \sum_k S(U_k, V_k, n_{ik}). \quad (3.6)$$

### 3.1. Conditions for equilibrium

The portioned system is in equilibrium when the inequality reduces to an exact equality. This leads to a variational form of the second law: among *all possible* internal distributions of  $U$ ,  $V$  and  $n_i$  that satisfy the constraints in (3.3), the equilibrium distribution is the one that maximizes the total entropy:

$$S(U, V, n_i) = \max_{U_k, V_k, n_{ik}} \sum_k S(U_k, V_k, n_{ik}). \quad (3.7)$$

In this chapter we study the implications of this variational property of entropy.

#### 3.1 Conditions for equilibrium

The mathematical problem is the maximization of the total entropy in Eq. (3.7) under the constraints

$$\sum_k U_k = U, \quad \sum_k V_k = V, \quad \sum_k n_{ik} = n_i.$$

Using the method of Lagrange multipliers, the equivalent objective function is

$$\mathcal{F} = \sum_k S(U_k, V_k, n_{ik}) + \lambda_U \left( \sum_k U_k - U \right) + \lambda_V \left( \sum_k V_k - V \right) + \sum_i \lambda_i \left( \sum_k n_{ik} - n_i \right),$$

where  $\lambda_U$ ,  $\lambda_V$  and  $\lambda_i$  are the multipliers for each of the constraints. To maximize this function we set the derivatives with respect to  $U_k$ ,  $V_k$  and  $n_{ik}$  to zero:

$$\begin{aligned} \left( \frac{\partial \mathcal{F}}{\partial U_k} \right) = 0 & \Rightarrow \left( \frac{\partial S_k}{\partial U_k} \right)_{V_k, n_{ik}} = -\lambda_U & \Rightarrow T_k = \lambda_U \\ \left( \frac{\partial \mathcal{F}}{\partial V_k} \right) = 0 & \Rightarrow \left( \frac{\partial S_k}{\partial V_k} \right)_{U_k, n_{ik}} = -\lambda_V & \Rightarrow \frac{P_k}{T_k} = \lambda_V \\ \left( \frac{\partial \mathcal{F}}{\partial n_{ik}} \right) = 0 & \Rightarrow \left( \frac{\partial S_k}{\partial n_{ik}} \right)_{U_k, V_k, n_{jk}} = -\lambda_i & \Rightarrow \frac{\mu_i}{T_k} = \lambda_i. \end{aligned}$$

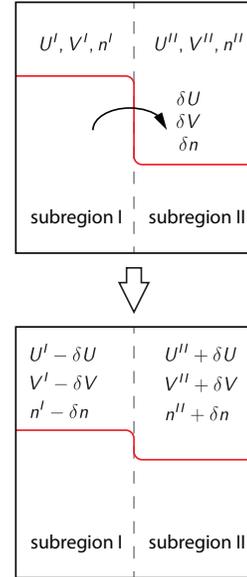
Each of the derivatives  $(\partial S_k / \partial U_k)$ ,  $(\partial S_k / \partial V_k)$  and  $(\partial S_k / \partial n_{ik})$  has a common value in all subsystems. Thus, the internal distribution of energy volume and moles is such that temperature, pressure and the chemical potential of all components are *uniform* throughout the system:

$$T_k = \text{const.} = T \quad (3.8)$$

$$P_k = \text{const.} = P \quad (3.9)$$

$$\mu_{ik} = \text{const.} = \mu_i. \quad (3.10)$$

These equalities express the conditions of thermal, mechanical and chemical equilibrium and are consequences of the variational property of entropy.



**Figure 3.1:** Conceptual system for the determination of equilibrium conditions. The progression towards equilibrium is driven by gradients that set up transfer in the direction of dissipating the gradient.

### 3.2 Geometric properties of entropy

If we partition a system into two parts I and II, the inequality of the second law reads

$$S^{I+II} \geq S^I + S^{II}.$$

We write this in the form

$$\left(\frac{S}{n}\right)^{I+II} \geq \left(\frac{n^I}{n^I + n^{II}}\right)\left(\frac{S}{n}\right)^I + \left(\frac{n^{II}}{n^I + n^{II}}\right)\left(\frac{S}{n}\right)^{II},$$

where  $n^I$  and  $n^{II}$  are the total moles in each compartment. Let  $s = S/n$  be the intensive entropy, a function of the intensive variables  $u = U/n$ ,  $v = V/n$ . The previous inequality can then be written in the more compact form,

$$s^{I+II} \geq \alpha s^I + (1 - \alpha)s^{II}, \quad (3.11)$$

where  $s = S/n$  is the molar entropy,  $\alpha = n^I/(n^I + n^{II})$  is the fraction of moles in subsystem I and  $1 - \alpha$  is the fraction in subsystem II. As an intensive property  $s$  is a function of the intensive variables  $u = U/n$ ,  $v = V/n$ ,  $x_i = n_i/n$ . We explore the above inequality in Fig. 3.2, which shows  $s(u, v, x_i)$  as a function of  $u$  at fixed  $v$  and  $x_i$ . This amounts to distributing energy between the two subsystems while keeping  $v$  and  $x_i$  in each part constant. The linear segment that connects points I and II is described by the parametric equation

$$y = \alpha s^I + (1 - \alpha)s^{II},$$

with parameter  $\alpha$ . According to Eq. (3.11) this equation lies below  $s(u)$  for all  $0 \leq \alpha \leq 1$ . This can only be so if  $s(u)$  is a concave function of  $u$ . The same analysis can be applied to the other variables and leads to the conclusion that  $s(u, v, x_i)$  is a concave function of  $(u, v, x_i)$ . Since  $s$  is concave, so is  $\lambda s$ ; with  $\lambda = n$  we obtain the extensive entropy and we conclude that  $S(U, V, n_i)$  is concave function of  $(U, V, n_i)$ .

The concave shape of  $S$  is expressed mathematically by the condition

$$d^2 S < 0. \quad (3.12)$$

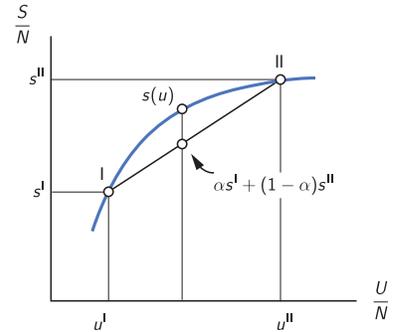
The equilibrium state of an isolated system is given by the condition

$$dS = 0. \quad (3.13)$$

Equation (3.13) states that the equilibrium state is at an extremum of entropy, and Eq. (3.12) states that this extremum is a maximum. We refer to (3.13) as the equilibrium condition and Eq. (3.12) as the stability condition. The geometric interpretation is that stable states must reside on concave portion of  $S$ . If  $S$  contains convex regions, these are physically unattainable.

**Curvature and homogeneity** Entropy is homogeneous and concave function, two properties that are key to understanding its behavior. Concave functions satisfy the concave inequality

$$f(\alpha_1 x_1 + \alpha_2 x_2) \geq \alpha_1 f(x_1) + \alpha_2 f(x_2) \quad (3.14)$$



**Figure 3.2:** Geometric interpretation of the concave inequality of entropy. The line that connects I and II is given by the linear relationship  $\alpha s^I + (1 - \alpha)s^{II}$ . All points of the segment (I)–(II) lie below  $s$ . This means that  $s$  is a concave function of its variable.

### 3.3. Equilibrium conditions under various constraints

for all  $\lambda_i \in [0, 1]$  that satisfy  $\alpha_1 + \alpha_2 = 1$ . It expresses the geometric property that any chord that connects two points on  $f$  lies below  $f$ .

If  $f$  is a function of multiple variables and is concave with respect to several of them, this inequality applies to each of them while holding all other variables constant. If  $f(x, y)$  is positive, concave and homogeneous with degree 1 with respect to  $x$  and  $y$  then we obtain the additional inequality (see Problem 3.4)

$$f(x_1 + x_2, y_1 + y_2) \geq f(x_1, y_1) + f(x_2, y_2), \quad (3.15)$$

which is the inequality of the second law. In purely mathematical terms the second law is a consequence of the curvature and homogeneity of entropy.

### 3.3 Equilibrium conditions under various constraints

The fundamental inequality of entropy serves as the equilibrium criterion under the conditions of fixed total energy, volume and number of moles. Mathematically this requires us to work with  $(U, V, n_i)$  as independent variables. It is possible to develop equilibrium criteria that work with different sets of independent variables. To do this, we consider a system coupled to an equilibrium bath at fixed temperature  $T'$ , pressure  $P'$  and chemical potential  $\mu'$ . The energy, volume and number of particles of the system/bath arrangement are partitioned between system and bath under the condition that the bath remains at equilibrium; their distribution within the system is not necessarily an equilibrium partition. Under a small perturbation the second law requires

$$dS + dS' \geq 0, \quad (3.16)$$

where unprimed variables refer to the system and primed variables to the bath. The bath is in equilibrium and its entropy change is

$$dS' = \frac{dU'}{T'} + \frac{P'dV'}{T'} - \sum_i \frac{\mu' dn'_i}{T'} = -\frac{dU}{T'} - \frac{P'dV}{T'} + \sum_i \frac{\mu' dn_i}{T'},$$

where we have used  $dU' = -dU$ ,  $dV' = -dV$ ,  $dn'_i = dn_i$ . Substituting into Eq. (3.16) we obtain

$$dU - T'dS + P'dV - \sum_i \mu'_i dn_i \leq 0. \quad (3.17)$$

In this inequality the extensive properties  $(U, V, n_i)$  are those of the system and the intensive properties  $(T', P', \mu'_i)$  are those of the bath. We will now consider various special cases.

- *Constant  $T, V, n_i$ :* With  $T = T'$ ,  $dV = 0$ ,  $dn_i = 0$ , Eq. (3.17) becomes

$$(dU - TdS)_{T,V,n_i} \leq 0$$

or

$$\Delta A \Big|_{T,V,n_i} \leq 0. \quad (3.18)$$

This inequality sets a minimum possible value in the Helmholtz energy of a system at fixed  $(T, V, n_i)$ . The minimum is reached when Eq. becomes an exact equality (3.16), i.e., when the system is in equilibrium. The

### 3.3. Equilibrium conditions under various constraints

condition “fixed  $T, V, n_i$ ” is understood to mean that volume and the number of moles are partitioned within the system while temperature is uniform throughout the system (as an intensive property temperature is not partitioned).

- *Constant  $T, P, n_i$* : In this case  $T = T', P' = P, dn_i = 0$  and Eq. (3.17) becomes

$$(dU - TdS + PdV)_{T,P,n_i} \leq 0,$$

or

$$\Delta G \Big|_{T,P,n_i} \leq 0. \quad (3.19)$$

The equilibrium state under fixed  $(S, P, n_i)$  minimizes the Gibbs energy.

- *Constant  $S, V, n_i$* : With  $dS = 0, dV = 0, dn_i = 0$  we obtain

$$\Delta U \Big|_{S,V,n_i} \leq 0. \quad (3.20)$$

Under fixed  $(S, V, n_i)$  equilibrium state minimizes the internal energy.

- *Constant  $S, P, n_i$* : We set  $dS = 0, P = P'$  and  $dn_i = 0$  in Eq. (3.17) to obtain

$$dU + PdV = dU + d(PV) \leq 0,$$

or

$$\Delta H \Big|_{S,P,n_i} \leq 0. \quad (3.21)$$

Under fixed  $(S, P, n_i)$  equilibrium state minimizes the enthalpy.

#### 3.3.1 Equilibration and the potential to produce work

If we fix the temperature, pressure and all chemical potentials in the system to be same as in the bath, i.e.,  $T' = T, P' = P, \mu'_i = \mu_i$ , Eq. (3.17) becomes

$$dU \leq TdS - PdV + \sum_i \mu_i dn_i. \quad (3.22)$$

This is a general inequality that takes simpler forms when we fix the state of the system by constraining one variable in each of the pairs below:

$$(T \text{ or } S) \quad (P \text{ or } V) \quad (\mu_i \text{ or } n_i).$$

If the variable we fix is intensive ( $T, P$  or  $\mu_i$ ) it is understood that it is uniform throughout the system; if the variable we fix is extensive ( $S, V$  or  $n_i$ ) it is understood to be partitioned arbitrarily. Thus by fixing  $T, V$  and  $n_i$  we obtain Eq. (3.18), by fixing  $T, P$  and  $n_i$  we obtain Eq. (3.19) and so on.

We may turn Eq. (3.22) into an exact equality by adding a negative term on the right-hand side with units of energy:

$$dU = TdS - PdV + \sum_i \mu_i dn_i + dW^*. \quad (3.23)$$

### 3.3. Equilibrium conditions under various constraints

This is a statement of energy conservation: on the left-hand side we have the total change in the energy of the system. On the right-hand side each term represents an energy exchange:  $TdS$  is the reversible heat exchanged with the bath;  $-PdV$  is the reversible work associated with the expansion/contraction of the system;  $\mu_i dn_i$  is energy associated with the transfer of chemical species between the system and the bath; the last term,  $dW^*$  represents work that is available after all other interactions with the surroundings have been accounted for. Solved for  $dW^*$  Eq. (3.23) becomes

$$dW^* = dU - TdS + PdV - \sum_i \mu_i dn_i. \quad (3.24)$$

From this we obtain a number of special results under various different constraints. If we fix  $T$ ,  $V$  and  $n_i$  we obtain

$$dW^* \Big|_{T,V,n_i} = dU - TdV = dA,$$

and in integral form,

$$W^* \Big|_{T,V,n_i} = \Delta A \Big|_{T,V,n_i} \leq 0. \quad (3.25)$$

That is, the available work for a process that reaches equilibrium under fixed temperature, volume and number of moles is equal to the change of the Helmholtz energy. This change is negative (see Eq. (3.18)), therefore  $W^*$  represents useful work that can be extracted from the system. By similar steps we obtain the companion results below:

$$W^* \Big|_{T,P,n_i} = \Delta G \Big|_{T,P,n_i} \leq 0 \quad (3.26)$$

$$W^* \Big|_{S,V,n_i} = \Delta U \Big|_{S,V,n_i} \leq 0 \quad (3.27)$$

$$W^* \Big|_{S,P,n_i} = \Delta H \Big|_{T,P,n_i} \leq 0. \quad (3.28)$$

All of these results express the same fundamental principle: a non-equilibrium system has the potential to produce useful work as its transition to the equilibrium state. This work can be extracted by harnessing the gradients that drive the process of equilibration. Conversely, to create a non equilibrium state out of a system in equilibrium we must expend work. This principle has its analogues in mechanical systems. A river that flows from higher to lower elevation has the potential to produce work, via a hydroelectric plant, for example. The available work does not materialize automatically when a system is driven to equilibrium. Without a hydroelectric plant in its path the river will reach sea level without producing any useful work. The extraction of the available work requires human intervention as well as engineering ingenuity. An electric motor must be used to extract work from the poles of a battery, a Carnot engine must be placed between two bodies at different temperature to extract work from the tendency of systems to thermally equilibrate, and so on.

### 3.3.2 Thermodynamic potentials

In non-thermal systems the tendency to reach equilibrium is quantified by potential functions. A mechanical potential  $\Phi(X)$  is defined such that the force that acts on a body is

$$F_X = -\frac{\partial\Phi}{\partial X}.$$

The negative sign means that the force points in the direction of decreasing potential. Mechanical equilibrium is reached when the force is zero, i.e., when the potential is at a minimum with respect to  $X$ . The work associated with moving the system from point  $A$  to point  $B$  is the difference between the potentials at the two points,

$$W = \Delta\Phi_{AB} = \Phi(X_B) - \Phi(X_A).$$

Accordingly, a system on its way to the point of mechanical equilibrium produces work; to move the system away from equilibrium work must be expended. These properties of mechanical potentials are present in the thermodynamic functions

$$U(S, V, n_i) \quad H(S, P, n_i) \quad A(T, V, n_i) \quad G(T, P, n_i).$$

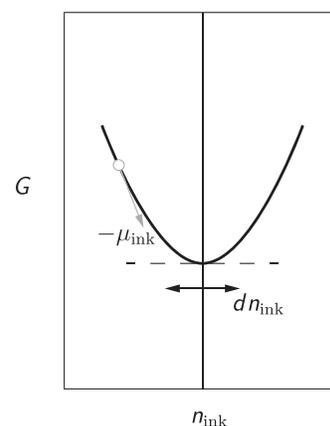
Let us take the Gibbs energy as an example. The conditions of constant  $T$ ,  $P$  and  $n_i$  mean that the only possible process in this system is the distribution of its moles between parts of the system. An experiment that represents this process is simple to construct: place a drop of ink in a glass of water. In this process temperature and pressure are fixed by the ambient conditions and the only observed change is the spread of the ink blob through the liquid. To the extent that the transition to the final equilibrium state may be approximated by a series of partitionings of the ink and the water between internal parts of the system (Fig. 3.3). The Gibbs energy in this case decreases until the equilibrium state is reached. This process can now be interpreted to take place under the action of the thermodynamic potential  $G(T, P, n_i)$  (Fig. 3.4). When  $dn$  moles of ink is transferred from some part  $A$  to some other part  $B$  the change in the Gibbs energy is<sup>1</sup>

$$dG = \mu_A(-dn) + \mu_B(+dn) = -(\mu_A - \mu_B)dn,$$

where  $\mu_A$  and  $\mu_B$  is the chemical potential of ink in each part. The transfer is in the direction of equilibrium if  $dG < 0$  or  $\mu_A > \mu_B$ . This means that ink is transferred from regions of higher chemical potential to regions of lower chemical potential and in this sense  $\delta\mu$  acts as the “thermodynamic force” that moves ink from part  $A$  to part  $B$ . The Gibbs energy acts entirely analogously to a mechanical potential: its negative slope is the force that drives the process of equilibration and the equilibrium state is indicated by its minimization. The same interpretation can be given to  $U(S, V, n_i)$ ,  $H(S, P, n_i)$  and  $A(T, V, n_i)$ . For this reason the four properties,  $U$ ,  $H$ ,  $A$  and  $G$ —all Legendre transformations of each other—are known as *thermodynamic potentials*. In all cases their derivatives with respect to their extensive variables represent the thermodynamic force that drives the process of equilibration; the equilibrium state is a state of zero gradients and is represented mathematically by their minimum with respect to the extensive variables.



**Figure 3.3:** Top: A blob of ink spreads in water at fixed temperature and pressure. Bottom: Representation of the process as a partitioning of the amount of water and ink, shown by different shades of blue.



**Figure 3.4:** Equilibration viewed as an action of a potential function. The negative gradient of the potential is the thermodynamic force that drives the approach to equilibrium.

**Curvature of thermodynamic potentials** The mathematical properties that gives potential functions their special character is curvature and homogeneity. The inequalities in Eqs. (3.18)–(3.21) are all of the form

$$f(x_1 + x_2, y_1 + y_2) \leq f(x_1, y_1) + f(x_2, y_2)$$

where  $f(x)$  is homogeneous function in  $x$  with degree 1 and  $x_i$  is a partition of  $x$ . This is the inverted form of the concave/homogeneous condition in Eq. (3.15) and indicates that  $f$  is convex in  $x$ . The curvature of the potentials can be traced to the curvature of entropy. Using the Helmholtz energy as an example, we write  $A$  in the form

$$\frac{A}{T} = -\left(S - \frac{U}{T}\right) = -(S - S_{US}),$$

which identifies the ratio  $A/T$  as the negative Legendre transformation of entropy with respect to internal energy. The Legendre transformation preserves the curvature with respect to the untransformed variables (here  $V$  and  $n_i$ ) but the negative sign inverts them. Since  $S$  is concave in  $V$  and  $n_i$ ,  $A$  is convex in the same variables. The convexity of the other potentials can be easily established by expressing them as Legendre transformations of  $A$ .

### 3.4 Stability

The equilibrium state of a  $(U, V, n_i)$  system maximizes entropy and satisfies the condition

$$\left[ dS \right]_{U,V,n_i} = 0. \quad (3.29)$$

This condition alone does not guarantee a maximum, only an extremum. To ensure a maximum we further require  $S$  to be concave:

$$\left[ d^2 S \right]_{U,V,n_i} < 0. \quad (3.30)$$

From a physical standpoint this inequality ensures that the equilibrium state is *stable*, i.e., it is restored following small perturbations. The stability condition can be expressed equivalently in terms of other potential functions:

$$\left[ d^2 U \right]_{S,V,n_i} > 0, \quad (3.31)$$

$$\left[ d^2 H \right]_{S,P,n_i} > 0, \quad (3.32)$$

$$\left[ d^2 A \right]_{T,V,n_i} > 0, \quad (3.33)$$

$$\left[ d^2 G \right]_{T,P,n_i} > 0. \quad (3.34)$$

All of these inequalities must be understood in the context of partitioning the *extensive* variables of a potential function while keeping all intensive variables *fixed*.

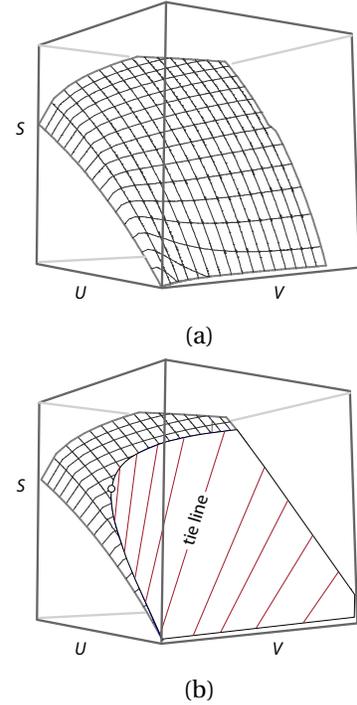
The general form of the the second order differential of a multivariate function  $F(x, y, z \dots)$  is

$$d^2 F = F_{xx} dx^2 + F_{yy} dy^2 + F_{zz} dz^2 \dots + F_{xy} dx dy + F_{xz} dx dz + \dots,$$

<sup>1</sup>Since both parts are internally equilibrated the Gibbs energy is

$$G = \sum_i \mu_i n_i.$$

When  $dn_i$  moles of species  $i$  are transferred to or from the system the change in the Gibbs energy is  $dG = \pm \mu_i |dn_i|$ , with the sign depending on whether the transfer into or out of the system



**Figure 3.5:** (a) The  $S(U, V)$  surface of the van der Waals equation for a pure fluid contains a convex region (most clearly seen in the contours off constant  $U$ ). (b) The same surface with the convex region removed. The tie lines in the VLE region connect the saturated phases (the liquid is the lower left branch, the vapor is on the upper one). Here, we have fixed  $N = 1$ , which makes  $S$ ,  $U$  and  $V$  molar properties.

### 3.5. Thermodynamics of phase equilibrium

where  $F_{xx}$  is the second derivative with respect to  $x$ ,  $F_{xy}$  is the mixed derivative in  $x$  and  $y$ , and so on. Applying the stability condition to  $U$  of pure component ( $n_i = n$ ) we have

$$d^2U = U_{SS}dS^2 + U_{VV}dV^2 + U_{nn}dn^2 + 2(U_{TV}dTdV + U_{Tn}dTdn + U_{Vn}dVdn) > 0.$$

This inequality applies to any independent variation in  $dT$ ,  $dV$  and  $dn$ . For a variation at fixed  $V$  and  $n$  it reduces to

$$U_{SS}dS^2 > 0,$$

which requires  $U_{SS} > 0$ . Recalling the result  $C_V = T/U_{SS}$  this inequality states

$$C_V > 0. \quad (3.35)$$

Thermal stability requires the heat capacity to be positive. In what sense is a negative  $C_V$  “unstable?” Suppose we add heat to a closed system at fixed volume by making thermal contact with a high temperature source. The heat is  $dQ = C_V dT$ , which means that if the  $C_V$  is negative, the temperature of the system *decreases upon heating*. Heat is still flowing from higher to lower temperature, so the second law is not violated. We have, however, a situation in which temperature gradients *increase* as a result of heat transfer induced by the gradient, leading to a runaway process that is incapable of reaching equilibrium. Stability is possible only when gradients dissipate in response to spontaneous transport they produce.

Analogous results to Eq. (3.35) can be obtained using other combinations of variables. In general, if  $F$  is a potential function and  $X$  an extensive variable, the presence of the term  $F_{XX}d^2X$  in  $d^2F$  implies that  $F_{XX}$  must be positive, since variations of  $dX$  keeping all other variables constant must satisfy the stability condition. For pure component, all of the following derivatives must be positive:

$$\left( U_{SS}, U_{VV}, U_{nn}, H_{SS}, H_{nn}, A_{VV}, A_{nn}, G_{nn} \right) > 0. \quad (3.36)$$

We apply these conditions only to extensive variables because we are minimizing the potentials with respect to partitioning the system into subsystems. Only extensive variables can be partitioned.

**Exercise 3.1.** Use stability analysis to obtain the sign of  $F_{xx}$ , where  $F$  is a potential function and  $x$  an intensive variable of  $F$ . Hint: First determine the sign of  $A_{TT}$ , then generalize the result to any potential function with intensive variables.

**Exercise 3.2.** Simplify each of the stability criteria in Eq. (3.36). For each case, explain the stability in physical terms.

### 3.5 Thermodynamics of phase equilibrium

Convex regions on the  $S(U, V, n_i)$  surface are unstable and thus physically unattainable. This produces a gap in the phase diagram that consist of a mixture of two distinct phases. Figure 3.5a shows the surface of a van der Waals

### 3.5. Thermodynamics of phase equilibrium

fluid. This surface contains a convex region that is most clearly seen in the wavy shape of the contours of constant  $U$ . In figure 3.5b the unstable region has been removed and the phase boundary is shown by the bell-shaped curve and the tie lines that connect the equilibrium phases. The numerical calculation of the phase boundary is based on the the equilibrium criteria given in Eqs. (3.8)–(3.10) but some geometric properties of he solution can be determined based on these criteria. Temperature pressure (and chemical potential)<sup>2</sup> are the slopes of the entropy surface with respect to the variables in  $S$ . Since both phases have the same slopes, they both lie on a plane that is tangent at both points where it makes contact with the surface. This geometric construction can be illustrated more clearly using the Helmholtz free energy.

The Helmholtz free energy is convex in its extensive arguments and the unstable regions are concave. Figure 3.6 shows  $A$  of pure component ( $N = 1$ ) as a function of  $V$  at fixed temperature. The graph contains a concave region, which implies there must be two phases. First, we note that the slope of  $A$  on this graph is pressure:

$$\text{slope} = \left( \frac{\partial A}{\partial V} \right)_{T, n_i} = -P.$$

The y intercept of the tangent line drawn at volume  $V$  is

$$\text{intercept} = A - (\text{slope})V = A + PV = G, \quad (3.37)$$

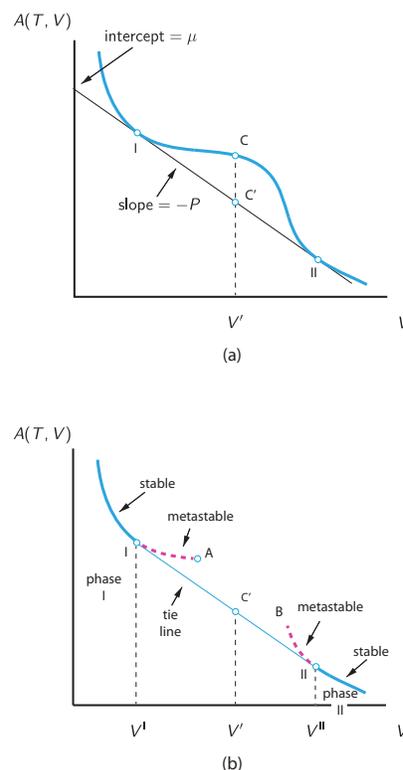
and this, for a pure substance is equal to the chemical potential,  $\mu$ .<sup>3</sup> Since both phases have the same pressure and chemical potential, the tangent lines drawn at each phase have the same slope and same intercept. They must lie, therefore, on a common tangent. This geometric construction defines points I and II, which represent the two equilibrium phases. To the left of point I we have a single phase system (phase I), to the right of point II we have another single phase system (phase II) and between the two points we have a mixture of the two phases. The gap between the two phases is bridged with a straight tie line. A state in this region is a two-phase mixture whose composition is given by the lever rule:

$$\text{mol fraction of phase I} = \frac{V^{\text{II}} - V'}{V^{\text{II}} - V^{\text{I}}}$$

where  $V^{\text{I}}$  and  $V^{\text{II}}$  are the molar volumes of the equilibrium phases and  $V'$  is the overall molar volume of the mixture. The Helmholtz energy of that state corresponds to point  $C'$  on the tie line and is located below point  $C$  on the unstable portion. If we attempt to create other tie lines by connecting a state to the left of  $C$  with a state to the right, the intersection with the vertical line at  $V'$  will always be above state  $C'$ . This is a graphical demonstration of the fact that the equilibrium state at fixed  $T$ ,  $V$  and  $N$  minimizes the Helmholtz energy, and a direct consequence of the fact that  $A$  is convex. The inflection points, marked by  $A$  and  $B$  in Fig. 3.6 lie between the equilibrium phases. The segments I- $A$  and II- $B$  are *metastable*: they are stable, since they are convex, but less so than the equilibrium phases because the result in higher  $A$ . Given the chance, a physical system will choose the most stable state. It is possible, however, to observe metastable states in practice, as when liquids are subcooled below their freezing point or a dissolved solute reaches a state of supersaturation before

<sup>2</sup>To be able to plot the multidimensional surface on paper, we have been working with molar properties. This eliminates the axis for  $n$ .

<sup>3</sup>More precisely,  $A + PV = N\mu$  with  $N = 1$ .



**Figure 3.6:** (a) Molar Helmholtz free energy of pure fluid with an unstable concave region. (b) The corrected graph with the unstable region removed.

it crystallizes. The inflection points represent the limit of intrinsic stability. A physical system cannot exist in the unstable region.

### 3.6 Chemical potential and related properties

The chemical potential is the key property in calculations of phase equilibrium. It is linked to all major thermodynamic functions in the form of a partial derivative with respect to the moles of component,

$$\mu_i = -kT \left( \frac{\partial S}{\partial n_i} \right)_{U,V,n_j} = \left( \frac{\partial U}{\partial n_i} \right)_{S,V,n_j} = \left( \frac{\partial H}{\partial n_i} \right)_{S,P,n_j} = \left( \frac{\partial A}{\partial n_i} \right)_{T,V,n_j} = \left( \frac{\partial G}{\partial n_i} \right)_{P,T,n_j}.$$

Its relationships to the Gibbs free energy is particularly useful because it expresses chemical potential as a partial molar derivative:

$$\mu_i = \bar{G}_i. \quad (3.38)$$

By Euler's theorem we then have

$$G = \sum n_i \mu_i, \quad (3.39)$$

where we understand both sides to be at the same temperature and pressure. The chemical potential is a function of  $T$ ,  $P$  and composition. It is the compositional dependence that complicates the calculation of the chemical potential in the general case. Two special cases that are both important and simple to handle are, pure component, and component in ideal-gas mixture.

**Pure component** For pure component we have  $G(T, P, n) = nG(T, P, 1)$ , where  $G(T, P, 1) \equiv g(T, P)$  is the molar Gibbs free energy. For the chemical potential we then obtain,

$$\mu = \left( \frac{\partial nG(T, P, 1)}{\partial n} \right)_{T,P} = G(T, P, 1) = g(T, P)$$

that is, the chemical potential of pure component is the same as the molar Gibbs free energy.

**Component in ideal-gas mixture** The Gibbs free energy of mixture in the ideal-gas state is

$$\begin{aligned} G^{\text{igm}}(T, P; n_i) &= H^{\text{igm}}(T, P; n_i) - TS^{\text{igm}}(T, P; n_i) \\ &= \sum_i n_i g_i^{\text{ig}}(T, P) + RT \sum_i \log \frac{n_i}{n}, \end{aligned} \quad (3.40)$$

where  $g_i^{\text{ig}}(T, P) = h_i^{\text{ig}}(T, P) - Ts_i^{\text{ig}}(T, P)$  is the molar Gibbs free energy of pure component in the ideal gas state. Applying the partial molar operator, the chemical potential can be easily shown to be<sup>4</sup>

$$\mu_i^{\text{igm}} = \mu_i^{\text{ig,pure}} + RT \log x_i, \quad (3.41)$$

<sup>4</sup>To obtain this result first write the total number of moles as  $n = n_i + C$ , where  $C$  is the sum of the moles of all other components, then differentiate with respect to  $n_i$  keeping  $T$ ,  $P$  and  $C$  constant.

where  $x_i = n_i/n$  is the mol fraction of component  $i$ . Here the superscript <sup>igm</sup> on the left-hand side indicates property of component in ideal gas *mixture*, and the superscript <sup>ig,pure</sup> on the right-hand side indicates the corresponding property of *pure* component in the ideal-gas state.

### 3.6.1 Fugacity

Fugacity is an auxiliary variable, closely related to the chemical potential, that is used in calculations of phase equilibrium in place of the chemical potential. The fugacity of component in a mixture is defined as

$$f_i = x_i P e^{\bar{G}_i^R/RT} \quad (3.42)$$

where  $x_i = n_i/n$  is the mol fraction of component,  $P$  is the total pressure and  $\bar{G}_i^R$  is the residual partial molar Gibbs free energy. This expression can be written in the simpler form

$$f_i = \phi_i x_i P, \quad (3.43)$$

where  $\phi_i$  is the fugacity coefficient and is defined as

$$\phi_i = e^{\bar{G}_i^R/RT}. \quad (3.44)$$

In the ideal-gas state,  $\bar{G}_i^R = 0$ , and Eqs (3.43) and (3.44) give

$$\phi_i^{\text{igm}} = 1, \quad f_i^{\text{igm}} = x_i P. \quad (3.45)$$

As we see, in the ideal-gas state the fugacity coefficient is unity, and fugacity is equal to the partial pressure of the component.

**Relationship to chemical potential** We wrote the definition of fugacity without offering any motivation for it. The nature of fugacity will become more clear once we establish its relationship to chemical potential. We start with the residual partial molar Gibbs free energy and its relationship to the chemical potential:

$$\bar{G}_i^R = \mu_i - \mu_i^{\text{igm}} = \mu_i - \mu_i^{\text{ig,pure}} - RT \log x_i \quad (3.46)$$

where we have used Eq. (3.41) for the chemical potential of component in ideal-gas mixture. Combining with Eq. (3.42) we obtain

$$\mu_i = \mu_i^{\text{ig,pure}} + RT \log \frac{f_i}{P}. \quad (3.47)$$

Consider now component  $i$  in two states  $A$  and  $B$ , both at the same temperature but not necessarily at the same pressure or composition. We write the above

### 3.6. Chemical potential and related properties

equation in each of the two states and take the difference:

$$\begin{aligned}\mu_i^A &= \mu_i^{\text{ig,pure}} \Big|_A + RT \log \frac{f_i^A}{P_A} \\ \mu_i^B &= \mu_i^{\text{ig,pure}} \Big|_B + RT \log \frac{f_i^B}{P_B}\end{aligned}$$

---


$$\mu_i^A - \mu_i^B = RT \log \frac{f_i^A}{f_i^B} + \Delta \mu_i^{\text{ig,pure}} \Big|_B - \log \frac{P_A}{P_B}$$

The ideal-gas difference of the chemical potential of pure component between states  $(T, P_A)$  and  $(T, P_B)$  is equal to  $\log P_A/P_B$ , and with this the final result is<sup>5</sup>

$$\frac{\mu_i^A - \mu_i^B}{RT} = \log \frac{f_i^A}{f_i^B}, \quad (T_A = T_B = T). \quad (3.48)$$

<sup>5</sup>Recall the relationship

$$d\mu^{\text{ig}} = dg^{\text{ig}} = PdV^{\text{ig}} = RTd \log P, \quad (\text{const. } T).$$

This establishes the relationship between fugacity and chemical potential and states that on an isotherm, a difference in chemical potentials is proportional to the difference in the logarithm of fugacities. If states  $A$  and  $B$  refer to two phases in equilibrium, then the equality of chemical potentials implies the equality of fugacities, and vice versa:

$$\mu_i^{\text{phase A}} = \mu_i^{\text{phase B}} \iff f_i^{\text{phase A}} = f_i^{\text{phase B}}. \quad (3.49)$$

Fugacity therefore may be used in place of the chemical potential as the criterion of equilibrium. There are a few practical advantages in working with fugacity: unlike chemical potential, fugacity does not require a reference state; moreover, in the ideal-gas state fugacity is well behaved property whereas chemical potential diverges.<sup>6</sup>

<sup>6</sup>Recall the relationship  $d\mu_i^{\text{ig}} = RTd \log P$  (constant  $T$ ); in the limit  $P \rightarrow 0$  we obtain  $\mu_i^{\text{ig}} \rightarrow -\infty$ .

#### 3.6.2 Fugacity and ideal states

Equation (3.43) for the fugacity of component in mixture may be written in the form,

$$f_i = \phi_i x_i f_i^{\text{ig,pure}},$$

where  $f_i^{\text{ig,pure}}$  is the fugacity of pure component in the ideal-gas state and is equal to pressure. The product  $x_i f_i^{\text{ig,pure}}$  is the fugacity of component in the ideal-gas state and is given as a simple product of the mol fraction of component times a quantity that is independent of composition. The fugacity coefficient is a correction factor that accounts from deviations from the ideal-gas state and the crux of the problem is how to calculate these deviations accurately. For states not too far from the ideal-gas state this approach is satisfactory but for states that are far removed it makes better sense to base the calculation at some reference state other than the ideal-gas state that is hopefully closer to the state of interest so that deviations from that state can be calculated more accurately. For this to work, it should be possible to calculate fugacity exactly at the state we

### 3.6. Chemical potential and related properties

take to be the new reference. We put these ideas into a mathematical form by writing fugacity as a product of three terms by analogy to Eq. (3.43):<sup>7</sup>

$$f_i = \mathcal{C}_i x_i f_i^\circ. \quad (3.50)$$

Here, the product  $x_i f_i^\circ$  represents the fugacity in some “ideal” state,  $f_i^\circ$  is a standard fugacity that is independent of  $x_i$  but is generally a function of temperature and pressure, and  $\mathcal{C}_i$  is a correction factor that accounts from deviations from the chosen state of ideality. Ideality here has the mathematical interpretation

$$f_i^{\text{“ideal”}} = x_i f_i^\circ,$$

and states that the fugacity of component  $i$  in the ideal state is proportional to the mole fraction of the component with a proportionality constant that is independent of  $x_i$ . Three ideal states in common use are, the ideal-gas state, the ideal solution, and the ideal solute. These are discussed below.

**Ideal-gas state** This is the state on which we have based the definition of fugacity and fugacity coefficient:

$$\begin{aligned} \mathcal{C}_i &\rightarrow \phi_i \\ f_i^\circ &\rightarrow P \end{aligned}$$

**Ideal solution** The ideal solution is defined by the following properties:

$$\begin{aligned} V &= \sum n_i V_i \\ H &= \sum n_i H_i \\ S &= \sum S_i - R \sum n_i \log(n_i/n_{\text{tot}}) \end{aligned}$$

where  $V$ ,  $H$ ,  $S$  refer to the solution, and  $V_i$ ,  $H_i$ ,  $S_i$ , refer to the pure components at the same pressure and temperature. The fugacity according to this reference state is given by Eq. (3.50) with

$$\begin{aligned} \mathcal{C}_i &\rightarrow \gamma_i, \\ f_i^\circ &\rightarrow \text{fugacity of pure liquid at } T, P \end{aligned}$$

Here  $\gamma_i$  is the activity coefficient and is related to the excess partial molar Gibbs free energy<sup>8</sup>

$$\log \gamma_i = \frac{\bar{G}_i^E}{RT}.$$

This equation is entirely analogous to Eq. (3.44) that relates the fugacity coefficient to the residual partial molar Gibbs free energy.

**Ideal solute** Here the reference state is the infinite dilution limit, namely a solutions which the mol fraction of component  $i$  is  $x_i \rightarrow 0$ . For this reference state,

$$\begin{aligned} \mathcal{C}_i &= \gamma_i^H, \\ f_i^\circ &= \text{Henry's law constant.} \end{aligned}$$

<sup>8</sup>Excess properties are differences between an actual property and the same property in ideal solution.

The activity coefficient here is based on Henry's law and should not be confused with the activity coefficient that is based on the ideal solution.

**Fugacity and reference states** We mentioned earlier that fugacity does not require a reference state and the above discussion of ideal states does not invalidate the statement. Fugacity itself does not require a reference state, however, the internal separation of fugacity into an “ideal-state” fugacity  $x_i f_i^\circ$  and a correction factor  $\mathcal{C}_i$  does. In each of the above ideal states the correction factor ( $\phi_i$ ,  $\gamma_i$  or  $\gamma_i^H$ ) is different, but if we were to calculate fugacity by all three reference states, we would get the same answer for the fugacity of component. Mathematically, the reference state tells us how to calculate  $\mathcal{C}_i$  for the chosen  $f_i^{\text{ideal}}$  so that the product  $\mathcal{C}_i x_i f_i^\circ$  is independent of the choice of the reference state. Figure ?? demonstrates this behavior for a binary mixture. The three ideal reference states are represented by straight lines with slope  $f_i^\circ$ . The fugacity at A can be calculated from any of these reference states by multiplying by the appropriate value of  $\mathcal{C}_i$ .

### 3.7 Additional reading

1. The discussion of equilibrium and stability of this chapter is similar to that in [Tester and Modell \(1996\)](#).
2. The section on chemical equilibrium and fugacity is a review of material that can be found in most undergraduate texts, see for example ([Matsoukas, 2013](#)).

### 3.8 Problems

**Problem 3.1.** (Homogeneity and curvature) Suppose  $f(x, y)$  is concave in both  $x$  and  $y$ . Let  $(x_1, y_1)$  and  $(x_2, y_2)$  be two points on the  $xy$  plane (these are shown as points A and B in Fig. 3.8) where the function takes the values  $f_1$  and  $f_2$ , respectively (shown as points A' and B'). Draw a straight line between A' and B'. Since  $f$  is concave all points on this line lie below the surface  $f(x, y)$ , except for points A' and B', which lie on the surface.

a) Show that the equation of the line that connects A' and B' on the  $xy$  plane can be expressed in parametric form as

$$x = (1 - a)x_1 + ax_2, \quad y = (1 - a)y_1 + ay_2.$$

where  $a$  is a parameter between 0 and 1.

b) Show that the equation

$$z = (1 - a)f_1 + af_2,$$

with  $x$  and  $y$  given by the parametric equations in part (a) is the equation of the straight line that connects A' and B'.

c) Prove the inequality

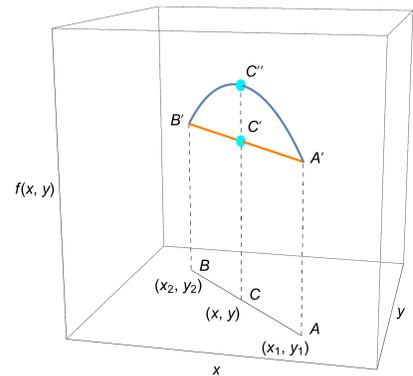
$$f(x, y) \geq (1 - a)f(x_1, y_1) + af_2(x_2, y_2).$$

d) Now show that if  $f(x, y)$  is homogeneous in  $x$  and  $y$  with degree 1, then

$$f(x_1 + x_2, y_1 + y_2) \geq f(x_1, y_1) + f_2(x_2, y_2),$$

for any two points  $(x_1, y_1)$ ,  $(x_2, y_2)$ .

e) How is the previous result different if  $f$  is *convex*?



**Figure 3.8:** Concave inequality (Problem 3.1). The segment A'C''B' is  $f(x, y)$  along the path AB on the  $xy$  plane.

### 3.8. Problems

**Thermodynamic Relevance:** The inequality of part of (d) is the inequality of the second law. We can view the second law in purely mathematical terms to be a consequence of two important properties of entropy: its curvature (it is a concave function of its arguments) and homogeneity.

**Problem 3.2.** (Homogeneity and equilibrium) Here we examine the connection between homogeneity and equilibrium. In the questions below  $f(x, y)$  is homogeneous in  $x$  and  $y$  with degree 1.

a) Show that the partial derivatives  $f_x$  and  $f_y$  are constant for all  $(x, y)$  that satisfy  $y/x = \lambda$ , where  $\lambda$  is constant.

b) Show that  $f(x, y)$  is a straight line along paths defined by the condition  $y/x = \text{const}$ .

c) If  $(x_1, y_1)$  and  $(x_2, y_2)$  are two points on the path defined by the condition  $y_2/x_2 = \lambda$  with  $\lambda$  constant, then

$$f(x_1, y_1) + f(x_2, y_2) = f(x_1 + x_2, y_1 + y_2)$$

Hint: Apply Euler's theorem to  $f$  and use the fact that the partial derivatives are constant along this path.

**Thermodynamic Relevance:** The path described by the condition  $y/x = \text{const}$ . defines the Euler path, a path along which all extensive independent variables (here  $x$  and  $y$ ) are increased by the same factor. This leaves all intensive properties (here the partial derivatives  $f_x$  and  $f_y$ ) constant, and multiplies all dependent extensive variables (here  $f$ ) by the same factor. The Euler path connects a series of states in equilibrium with each other: if we combine two systems, one at  $(x_1, y_1)$  and the other at  $(x_2, y_2)$ , the resulting system  $(x_1 + x_2, y_1 + y_2)$  lies on the same equilibrium line. Then,  $f(x_1, y_1) + f(x_2, y_2)$ , which is the total  $f$  in the two systems, is equal to  $f(x_1 + x_2, y_1 + y_2)$ , which is the amount in the combined system. This equality applies only at equilibrium.

**Problem 3.3.** We partition a system of fixed  $(U, V, n_i)$  into  $K$  volume elements  $V_k$ , not necessarily of the same size. We then partition the energy and the number of moles such that

$$\frac{U_k}{U} = \frac{n_{ik}}{n_i} = \frac{V_k}{V} \equiv \lambda_k$$

Show that this partition is in equilibrium and satisfies Eqs. (3.8), (3.9) and (3.10).

**Problem 3.4.** Function  $f(x_1, x_2)$  has the following properties with respect to  $x_1$  and  $x_2$ : it is monotonically increasing, concave, and homogeneous with degree 1.

a) Show that  $f$  is positive for all  $x_1, x_2 > 0$ .

b) Show that

$$f(x_1 + x'_1, y_1 + y'_1) \geq f(x_1, y_1) + f(x'_1, y'_1).$$

c) Define  $h = x_1 - f/f_1$ , where  $f_1$  is the partial derivative of  $f(x_1, x_2)$  with respect to  $x_1$ . Treating  $h$  as a function of  $f_1$  and  $x_2$ , show that  $h$  is convex with respect to  $f_1$  and concave with respect to  $x_2$ .

d) Show that  $h$  at fixed  $f_1$  is homogeneous with degree 1 with respect to  $x_2$ .

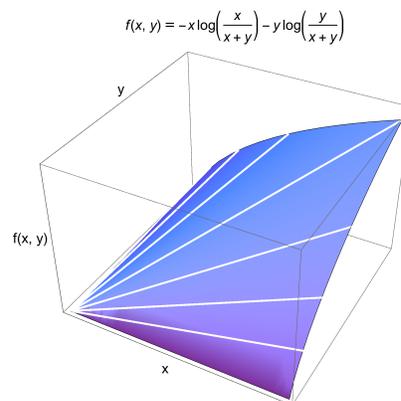
e) Show that  $h$  satisfies the inequality

$$h(f_1, x_2 + x'_2) \leq h(f_1, x_2) + h(f_1, x'_2),$$

for all positive  $x_2, x'_2$ .

**Thermodynamic Relevance:** The problem demonstrates that all thermodynamic relationships can be derived from the mathematical conditions placed on  $f$ . In this example  $h$  is to  $f$  what free energy is to entropy.

**Problem 3.5.** Two non-interacting boxes contain the same ideal gas ( $C_V = 3/2R$ ). One box contains  $n_1$  moles, has volume  $V_{10}$ , and temperature  $T_{10}$ ; the other box contains  $n_2 = 3n_1$  moles, has volume  $V_2 = 0.1V_1$ , and its temperature is  $T_{20} = 4T_{10}$ . We transfer an amount of energy  $\delta U$  and an amount of volume  $\delta V$  from box 1 to 2.



**Figure 3.9:** (Problem 3.2) If  $f(x, y)$  is homogeneous in  $x$  and  $y$  with degree 1, then  $f(x, y)$  is straight line along any path  $y = \lambda x$ . The function plotted here is the ideal entropy of mixing but the same is true for any homogeneous function with degree 1.

### 3.8. Problems

- a) Calculate the entropy of the new arrangement as a function of  $\delta U$  and  $\delta n$ .  
b) Obtain the equilibrium state between the two boxes by maximizing the total entropy with respect to  $\delta U$  and  $\delta V$ .

**Problem 3.6.** You are in charge of a reactor that makes porous particles for a novel separation process. The particles are hollow and have walls that are thin, thermally conducting, and permeable to oxygen but not to any other molecules. The particles come out of the reactor containing in their core pure nitrogen at 500 K, 1 bar. They are then stored in ambient air at 1 bar, 300K.

- (a) Determine the state of the core (pressure, temperature, composition) when the particles come into equilibrium with the air. Assume oxygen and nitrogen to form an ideal-gas mixture and that ambient air is 79% nitrogen, 21% oxygen, by volume.  
(b) To check the quality of the process you periodically look at samples under the microscope and have noticed that a good number of shells are broken. What suggestions do you have that might correct this problem?

**Problem 3.7.** Pure nitrogen and argon are held into two separate, rigid compartments of equal size. Each compartment contains  $N$  moles at pressure  $P_0$  and temperature  $T_0$ .

- a) We take  $M$  moles from the nitrogen compartment and place them in the argon compartment while maintaining the volume and temperature of the two compartments unchanged. Calculate the entropy change of the system and of the universe for this process. Express your result solely in terms of  $N$  and  $x = M/N$ , where  $x$  is the fraction of original amount of nitrogen that was transferred to the argon compartment.  
b) An inventor claims she has invented a non-flow process by which the entire amount of nitrogen can be brought into the argon compartment without the need to expend any work. The final temperature of the mixture is  $T_0$  (same as the temperature at the beginning of the process). Is such process feasible?  
c) Suppose the two compartments that contain the pure gases are brought into contact through a semipermeable material that allows nitrogen to diffuse into the argon side but prevents argon from doing the same. The process is allowed to proceed isothermally until it reaches equilibrium. What is the composition in the argon compartment at equilibrium?

**Problem 3.8.** A box is divided into two parts of equal volume via a partition that is conducting and permeable only to component A. The box is thermally insulated from the surroundings. Consider the following cases:

- a) Compartment 1 contains gas A at  $P_0, T_0$ ; compartment 2 is evacuated.  
b) Compartment 1 contains an equimolar mixture of gases A and B at  $T_0, P_0$ ; compartment 2 is evacuated.  
c) Compartment 1 contains an equimolar mixture of gases A and B at  $T_0, P_0$ ; compartment 2 contains pure gas B at  $T_0, P_0$ .

For each of the above cases, calculate the (i) the equilibrium state in each compartment (pressure, temperature, and if appropriate, composition) and (ii) the entropy generation,  $\Delta S/n_0$ , where  $n_0$  is the total number of moles in the box.

You may assume that both gases are ideal and both have the same heat capacity,  $c_P$ , which does not depend on temperature.

**Problem 3.9.** A gas tank contains  $n$  moles of pure nitrogen at 1 bar, 20°C. There is, however, a small leak through a crack that is permeable only to nitrogen. The tank is stored in ambient air (1 bar, 20°C, 79% by mol nitrogen and the rest may be assumed to be all oxygen). The walls of the tank are rigid and thermally conducting and all components may be assumed to be in the ideal-gas state.

- a) Obtain the entropy change in the tank when  $\delta n$  moles have leaked out.  
b) Obtain the entropy change of the air when  $\delta n$  moles have leak out.

- c) Identify the thermodynamic property that is maximized or minimized under the conditions of this process and apply this condition to obtain the final pressure in the tank.
- d) How would your answer in part (c) change if the tank contained hydrogen instead of nitrogen? In this case the crack is assumed to be permeable to hydrogen only.

**Problem 3.10.** The van der Waals equation of state is

$$P = \frac{RT}{V-b} - \frac{a}{V^2}$$

where  $V$  is the molar volume.

- a) Express the parameters  $a$  and  $b$  in terms of the critical temperature,  $T_c$ , and critical volume,  $V_c$ . *Hint:* At the critical point,

$$\left(\frac{\partial P}{\partial V}\right)_T = \left(\frac{\partial^2 P}{\partial V^2}\right)_T = 0$$

- b) Write the equation in terms of the dimensionless pressure  $P = P/P_c$ , dimensionless temperature,  $t = T/T_c$ , and dimensionless volume,  $v = V/V_c$ .
- b) Calculate the compressibility factor at the critical point.
- c) Obtain the residual enthalpy and residual entropy. Express the results in terms of the dimensionless variables  $p$ ,  $t$  and  $v$ .
- d) Calculate the saturation pressure  $p^{\text{sat}}$  at  $t = 0.9$ .
- e) Make a plot of  $p$  as a function of  $v$  at constant  $t = 0.9$  and indicate the binodal and spinodal points.
- e) Make separate graphs of  $u = U/RT_c$ ,  $h = H/RT_c$ ,  $s = S/R$  and  $g = G/RT_c$  as a function of  $v$  at constant  $t = 0.9$ . On each graph, indicate the points that correspond to the saturated liquid and saturated vapor. For these graphs, use  $c_p = 5R$  and take the reference state for  $h$  and  $s$  to be the ideal-gas state at the critical point.

**Problem 3.11.** a) Determine whether  $G(T, P, n_i)$  is a concave or convex function of  $T$  and  $P$ .

- b) Determine whether  $A(T, V, n_i)$  is a concave or convex function of  $T$ .
- c) Determine whether  $H(S, P, n_i)$  is a concave or convex function of  $P$ .
- d) Show that the  $C_p$  can never be negative.

**Problem 3.12.** a) On a plot of  $A(T, V, n)$  versus  $V$  at fixed  $T$  and  $n$  (pure component), draw the tangent line at some point  $V = V'$ . Identify the thermodynamic quantities represented by the slope and the intercept of the tangent line.

- b) Consider a plot of  $A(T, V, n)$  versus  $n$  at fixed  $T$ ,  $V$  (pure component). Sketch a qualitative graph that exhibits phase splitting, draw the tie line and identify the thermodynamic properties that are represented by the slope and intercept.

**Problem 3.13.** Molecular diffusion is driven by gradients in chemical potential. Suppose we start from a system where component  $A$  is concentrated in one region of the system and  $B$  in another. For example,  $A$  would be a blob of ink in water, and  $B$  is the water. Prove that as diffusion spreads the blob, the chemical potential of the ink in the blob decreases but the chemical potential of the ink in the water away from the blob increases. (Hint: prove that  $\mu_i$  must be a monotonically increasing function of the mole fraction of species  $i$ .)

**Problem 3.14.** The differential of the thermodynamic function  $F$  of pure component is

$$dF = -SdT - PdV - Nd\mu.$$

### 3.8. Problems

- a) Calculate  $F$  for steam at 20 bar, 300 °C. Additional information:  $U = 2750.8$  kJ/kg,  $H = 2994.3$  kJ/kg.
- b) Determine the sign of the derivative  $F_{\mu,\mu}$  for a stable phase.
- c) Draw a qualitative graph that shows  $F$  as a function of  $\mu$  at fixed  $T$  and  $V$ . Relate the slope and the intercept of the tangent line to common thermodynamic properties.

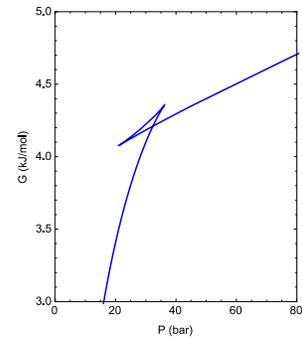
**Problem 3.15.** The differential of a thermodynamic function,  $F$  of a pure component is given by

$$dF = TdS + VdP - Nd\mu$$

- a) Relate  $F$  to common thermodynamic properties.
- b) Determine the value of  $F$  if it is known that  $U = 100$  J,  $A = 80$  J.
- b) Determine the sign of  $F_{\mu\mu}$  for a stable system.

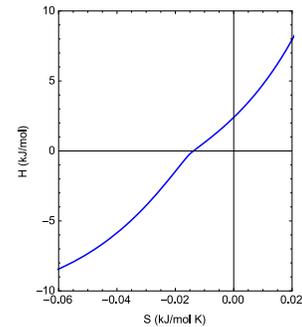
**Problem 3.16.** Figure 3.10 shows the Gibbs free energy of a pure van der Waals fluid at fixed temperature  $T = 270$  K as a function of pressure.

- a) Indicate the stable and unstable parts of the graph.
- b) Determine the saturation pressure at 270 K.
- c) Determine the molar density at 270 K, 40 bar.
- d) Determine the Helmholtz free energy at 270 K, 40 bar.



**Figure 3.10:** Problem 3.16.

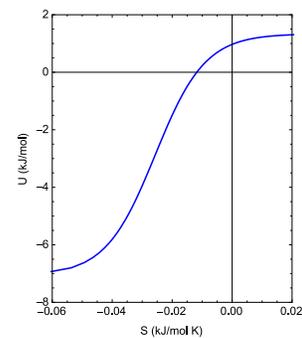
**Problem 3.17.** Figure 3.11 shows the enthalpy of a van der Waals fluid in the vapor-liquid region as a function of entropy at a constant pressure of 10 bar. Determine the saturation temperature, heat of vaporization and chemical potential of the saturated phases at 10 bar.



**Figure 3.11:** Data for Problem 3.17.

**Problem 3.18.** Figure 3.12 shows the internal energy of a pure van der Waals fluid as a function of entropy.

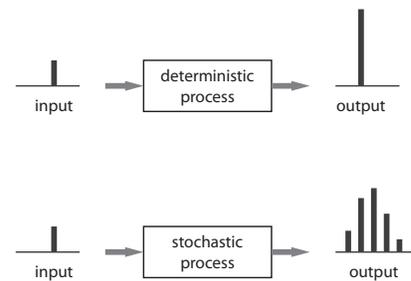
- a) Indicate the stable and unstable parts of the curve.
- b) Obtain the Helmholtz free energy at  $S = 0$  J/mol K.



**Figure 3.12:** Problem 3.18.

## Introduction to Stochastic Calculus

Classical mechanics is deterministic: the initial conditions and the potential field, in which a classical particle moves, fully specify its trajectory in space, both in the forward and backward direction in time. In general, a process is *deterministic* if the same inputs always produce the exact same output. There are many examples of processes that are *not* deterministic. Consider the experimental measurement of a physical quantity: successive measurements under otherwise identical conditions will generally produce a *set* of values that are scattered around the “true” value of the measured quantity. Another example is the double-slit experiment: a beam of electrons that passes through a double slit forms a classical interference pattern. If we send the electrons through the slits one at a time, each lands at some random spot, but when data are collected over a large number of electrons we obtain the diffraction pattern predicted by theory. As a third example, consider the toss of a die. The outcome is a number between 1 and 6 but its precise value cannot be predicted ahead of time. In all of these examples if we were to repeat the process under identical conditions we are not guaranteed the same answer each time. Such process is called *stochastic*. The origin of stochasticity is quite different in each case. In the example of experimental measurement it arises from a combination of instrumental factors and an inherent inability to prepare the system in *exactly* identical ways before measurement. In the double-slit experiment it is a fundamental aspect of quantum mechanics. In the case of a die, it is our inability to control the minute details of the toss, the exact force, angle of the toss, etc. Despite such differences, a common thread that runs through all of these examples is that the final outcome depends on a very large number of variables of which we can directly control only very few. When we say that we repeat an experiment under “identical” conditions we mean it only with respect to those factors that we can control. In thermodynamics the controlled variables are the macroscopic variables of state, for example,  $U$ ,  $V$  and  $n_i$ ; the uncontrolled variables are the precise state of individual molecules that give rise to the observable state. Probability theory is a mathematical language for dealing with stochastic variables. In this chapter we review the basic elements of probability theory. The discussion is brief and incomplete, and its main purpose is to develop tools that we need in the statistical treatment of thermodynamics. The reading section at the end of this chapter offers additional resources for those who need more detail.



**Figure 4.1:** A deterministic process always produces the same output for the same input. We are certain about the value of the output before we feed it to the process. A stochastic process produces a range of outputs with a defined distribution of frequencies. We are uncertain about the output each time but certain about its frequency distribution.

**Probability as frequency** Probabilities in physics are typically manifested as frequencies. Suppose we repeat the same stochastic experiment under identical conditions and tally the outcomes. Let  $n_i$  be the number of times we observe outcome  $i$ , where  $i = 1, 2, \dots$  runs through all possible outcomes. For example,  $i$  could mean “heads” in a coin toss, or it could mean that the measured temperature is found to be in the  $i$ th interval of a predetermined sequence of temperatures,  $T_1, T_2 \dots$ . The nature of physical systems is such that if we repeat the experiment  $N$  times, the ratio  $n_i/N$  converges to a value that is independent of  $N$  when  $N$  is large enough. In this limit  $n_i/N$  becomes an intensive property that characterizes outcome  $i$ . We identify this ratio as the *probability*  $p_i$  of outcome  $i$ :

$$\lim_{N \rightarrow \infty} \frac{n_i}{N} = p_i. \quad (4.1)$$

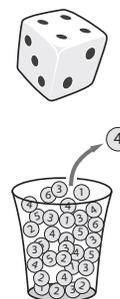
The collected outcomes from successive repetitions of a stochastic experiment represent a *sample* of the stochastic variable, an *ensemble*. This ensemble is homogeneous with respect to  $n_i$  with degree 1: if we increase the size of the population by a factor  $\lambda$ , all  $n_i$  will increase by the same factor, provided that  $N$  is large enough so that the limit in Eq. (4.1) is reached. This homogeneous nature of large samples establishes the relationship between probability, a *mathematical* concept, and relative frequency, a *physical* property.

What physical mechanism gives rise to homogeneity in large samples? When we perform a stochastic experiment we exert control only on those variables of the process that are accessible to us: we use the same die, we hold it with the same hand, we try our best to toss it with the about the same force each time. These factors alone do not uniquely determine the outcome because of the large number of uncontrolled variables that affect the process: the precise position of the die between our fingers, the force of the toss, the angular velocities at the moment of the toss, the texture of the surface at the point where the die lands, etc. The effect of the uncontrolled factors, which are too numerous to list even for a simple problem as that of tossing a die,<sup>1</sup> is to combine somehow in such ways as to materialize each outcome  $i$  a fraction  $p_i$  out of a very large number of repetitions.

We can create a conceptual model that produces the same behavior: We imagine a bucket that contains all possible outcomes, each in proportion to its probability  $p_i$ . With our eyes closed and the outcomes shuffled, we pick one and record it. We place it back in the bucket and repeat. The expectation is that over a large number  $N$  repetitions all outcomes will be sampled  $p_i N$  times on average, i.e., the normalized frequency of  $i$  in the sample will be equal to the probability  $p_i$ . By “eyes closed” and “shuffle” we mean to convey the notion of *randomization* that ensures that any outcome in the bucket is equally likely to be picked. The picture of a stochastic process as an ensemble of outcomes is a useful tool to visualize a stochastic process and the meaning of probability in terms of a thought experiment that relates to common experience.

**Probability as frequency or ignorance** The connection between probability and frequency makes physical sense and indeed most of the time probabilities in physics are manifested as frequencies, as in the physical experiment of tossing a coin, or in the tracking of the position of a Brownian particle. But not all probabilities can be strictly observed as frequencies. The central question in statistical mechanics is the

<sup>1</sup>Simple, that is, relative to the problem of dealing with some  $10^{23}$  molecules in some kind of process.



**Figure 4.2:** In the ensemble view of a stochastic variable we imagine a collection of all possible outcomes present in proportion to their probabilities. The contents of this bucket is the “population.” If we pick randomly elements from the bucket, we will be mimicking the original stochastic process.

*microstate* of matter, which is to say, the exact specification of the positions and momentum vectors of all particles in a given amount of matter at a given time. Statistical mechanics assigns probabilities to these microstates but since microstates are not directly observable, these probabilities cannot be directly compared to frequencies. The conundrum is philosophical and does not affect our ability to apply statistical mechanics or to test its testable predictions against observation. It is just that the probability of microstate itself is not one of those testable statements. Questions like this have led to alternative interpretations of probability. One such interpretation views probability as an extension of formal logic. In logic a statement is either true (truth value 1) or false (truth value 0). Probabilities provide a continuum of truth values between true and false and reflect our degree of confidence in predicting the outcome of a stochastic experiment. If the stochastic experiment is to randomly select a ball from a drawer with an equal number of black and white balls we would be inclined to assign equal probabilities to picking either color. This is a case where the assignment of an unknown probability is guided by known frequencies and the two interpretations give the same result. By interpreting, however, probabilities as measures of certainty or ignorance it is possible to apply them to more general questions that cannot be tested against frequency, such as, what is the probability that a certain individual will live  $x$  number of years. We will not be bogged down by these debates and for the most part we will be served well by treating probability as frequency. This is not to say that we ought to dismiss the deeper questions that arise from the use of probability calculus to physical problems. For the physical scientist the philosophical question is, *why* do the tools of probability work in physics? Why is it that over a long repetition of the same complex experiment the observed frequencies of outcomes *converge*? What are the *physical mechanisms* that lead to this convergence? We relegate these questions to the realm of philosophy because they deal with the deeper meaning of the methods we use. Ultimately, our theories are judged by how well they work, but this does not mean that we should not think about what it means that they work.

#### 4.1 Discrete frequency distributions

Stochastic variables are closely related to distributed populations, so we will start with a discussion of frequency distributions first. We are often dealing with collections of objects (“population”) that are distributed with respect to some property  $x$ . For example,  $x$  could be the radius of particles in a granular material, the number of monomers in a polymer chain, or a card in a deck of playing cards. The distributed variable may be continuous (e.g., particle radius) or discrete (e.g., number of monomers). We will consider discrete variables first, i.e.,  $x$  takes values from a discrete set  $\{x_1, x_2, \dots\}$  that may be finite or infinite. We suppose that we are dealing a sample of  $N$  elements from this population, such that the number of elements with  $x = x_i$  is  $F_i$ . We call  $F_i$  the number distribution, or simply the distribution of  $x$ . The sum of all  $F_i$  is the total number of elements in the population:

$$\sum F_i = N \quad (4.2)$$

with the summation going over all  $x_i$  in the population. The fraction of the population with  $x = x_i$  is

$$f_i = \frac{F_i}{\sum F_i}. \quad (4.3)$$

#### 4.1. Discrete frequency distributions

---

The set of  $f_i$  expresses the distribution in normalized (intensive) form whereas the set of  $F_i$  expresses it in extensive form. Often, the population at hand is a small sample from a large or even infinite population. In this case, we may view  $F_i$  as the extensive form of the distribution and  $f_i$  as the intensive form. Given a large enough sample, the intensive form  $f_i$  that is calculated from this sample is independent of the size of the sample.<sup>2</sup> The difference between  $F_i$  and  $f_i$  is a constant factor ( $1/N$ ), therefore, we may use either representation without loss of information. We generally prefer to work with the intensive form in order to obtain results that are independent of the size of the system. On the other hand,  $F_i$  has tangible physical interpretation (number of elements with the given trait) and if the population undergoes a *process*, polymerization for example, it is  $F_i$  for which we can write continuity equations.

<sup>2</sup>The error between the sampled  $f_i$  and the true  $f_i$  in the infinite population is of the order of  $1/\sqrt{N}$ . We can always make  $N$  large enough to make this sampling error arbitrarily small.

**Means** Given a function  $A(x)$ , its mean value in the population is

$$\bar{A} = \frac{\sum A_i F_i}{\sum F_i} = \sum A_i f_i \quad (4.4)$$

where  $A_i = A(x_i)$ . We may the mean as an operator  $\overline{A(x)}$  that operates on a function  $A(x)$  whose independent variable is distributed with distribution  $f$ . This operator is linear, that is, it has the property

$$\overline{aA(x) + bB(x)} = a\overline{A(x)} + b\overline{B(x)}. \quad (4.5)$$

For notational purposes we write  $\bar{A}$  instead of  $\overline{A(x)}$  because the result of the operator is independent of  $x$  (the summation in  $i$  “averages out” the effect of  $x_i$ .)

**Moments** The moment of order  $k$  is the mean value of  $x^k$  in the population:

$$\mu_k \equiv \overline{x^k} = \sum x_i^k f_i. \quad (4.6)$$

By definition,  $\mu_0 = 1$ ; the first-order moment is the mean value of  $x$ . Extensive moments are defined in analogous manner:

$$M_k = \sum x_i^k F_i. \quad (4.7)$$

The zero-order moment is the total number of elements in the population;  $M_1$  is the total amount of  $x$ . The relationship between extensive and intensive (normalized) moments is

$$\mu_k = \frac{M_k}{M_0}. \quad (4.8)$$

The *variance*  $\text{var}(x)$ , but more often  $\sigma^2$ , is defined as

$$\sigma^2 = \mu_2 - \mu_1^2 = \frac{M_2}{M_1} - \left( \frac{M_1}{M_0} \right), \quad (4.9)$$

The square root of the variance is the standard deviation and is a measure of the spread of the distribution around the mean.

**Example 4.1. Moments**

The quantity  $(x - \bar{x})^k$  is the central moment of order  $k$  and can be viewed as moments of the distribution  $f(x - \bar{x})$ , shifted to zero mean. Show the following:

$$\begin{aligned}\overline{(x - \bar{x})} &= 0 \\ \overline{(x - \bar{x})^2} &= \sigma^2 \\ \overline{(x - \bar{x})^3} &= \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3\end{aligned}$$

**Solution** (Left as an exercise)

**4.2 Continuous distributions**

With continuous variables it does not make sense to ask what is the number of elements whose  $x$  is *exactly equal* to some value. For example, how many people can you find whose height is exactly 1.821... meters, specified to infinite number of decimals? None. This has nothing to do with the fact that we cannot measure height with infinite precision. When  $x$  is continuous the size of the population is infinite; accordingly, the probability to find any given element of this population is zero. It only makes sense to talk about the number  $\delta N(x)$  of elements whose  $x$  falls within a finite interval,  $(x, x + \delta x)$ . If  $\delta x$  is small enough, the ratio  $\delta N/\delta x$  converges to the derivative of  $N$ , which we will call  $F(x)$ :

$$\frac{\delta N(x)}{\delta x} \rightarrow \frac{dN(x)}{dx} \equiv F(x). \quad (4.10)$$

The function  $F(x)$  represents the continuous distribution of  $x$  and its physical interpretation is this:

$F(x)dx$  is equal to the number of elements in  $(x, x + dx)$

The number of elements in a finite interval  $(x_a, x_b)$  is the area of the distribution between these two points:

$$\int_{x_a}^{x_b} F(x)dx = \text{number of elements in } (x_a, x_b). \quad (4.11)$$

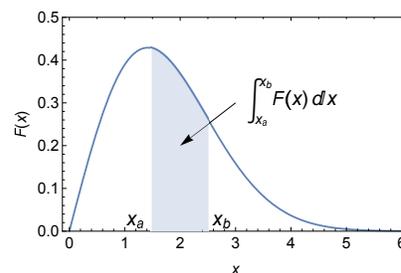
It follows that the total number of elements is the entire area under  $F$ :

$$\int F(x)dx = \text{total number of elements}. \quad (4.12)$$

Here and in the rest, the integration goes over the range of  $x$ . In principle,  $x$  goes from  $-\infty$  to  $\infty$  although usually the range is some smaller interval of the real axis. We will not write the limits in our integrals and will let each problem specify what they should be.

The normalized distribution  $f(x)$  is defined

$$f(x) = \frac{F(x)}{\int F(x)dx}, \quad (4.13)$$



**Figure 4.3:** The area under a continuous distribution represents the size of the population in the region covered by the integral.

and its interpretation is that  $f(x)dx$  represents the fraction of the population in the interval  $(x, x + dx)$ . Normalized and extensive moments are defined by analogy to discrete distributions:

$$m_k = \int x^k f(x) dx,$$

$$M_k = \int x^k F(x) dx.$$

In general, the equations of the discrete case can be converted to their continuous form by substituting summation in  $f_i$  or  $F_i$  with integrals in  $f(x)$  or  $F(x)$ . For example, the mean of  $A(x)$  in the population is

$$\bar{A} = \frac{\int A(x)F(x)dx}{\int F(x)dx} = \int A(x)f(x)dx, \tag{4.14}$$

which follows from Eq. (4.4).

**Transformation of variables** Often times we know the distribution of  $x$  and seek the distribution of some function  $y(x)$  that depends on  $x$ . For example we may know the distribution of velocities in a population of particles and seek the distribution of kinetic energies. In this case  $y = \frac{1}{2}mx^2$ , if  $x$  is velocity. Suppose that the distribution of  $x$  is  $f(x)$  and the distribution of  $y$  is  $h(y)$ , with  $y = y(x)$ . When  $x$  is converted into  $y$  the interval  $(x, x + d)$  is converted into a corresponding interval  $(y, y + dy)$ . Clearly, these two intervals contain the same fraction of the population because they represent the same region of the population, whether this region is expressed in terms of  $x$  or of  $y(x)$ . The fraction of the population in  $(x, x + dx)$  is  $f(x)dx$ , and in  $(y, y + dy)$  is  $h(y)dy$ . Then  $|f(x)dx| = |h(y)dy|$  from which we obtain

$$h(y) = f(x) \left| \frac{dx}{dy} \right|. \tag{4.15}$$

With the substitution  $x(y) = y^{(-1)}(x)$ , the right-hand side is a function of  $y$  and defines the distribution  $h(y)$ . This is a useful trick that allows us to transform the variable of a distribution. Since inverse functions are involved,  $y(x)$  must be strictly monotonic to allow inversion. If  $y$  is an increasing function the absolute value is redundant. If it is decreasing, i.e.  $|dx/dy| = -dx/dy$ , the negative sign ensures that integrations in either  $x$  or  $y$  move from low values to high values even though  $x$  and  $y$  move in opposite directions.

**Example 4.2.** Transformation of variables

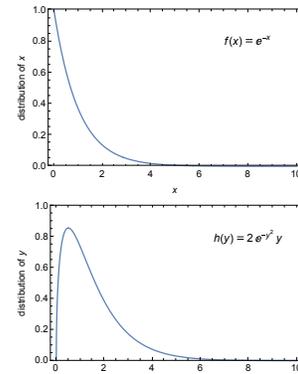
Suppose that the distribution of  $x$  is given by the exponential function

$$f(x) = e^{-x}.$$

Calculate the distribution of  $y = \sqrt{x}$ .

**Solution** The derivative of  $y$  is

$$\frac{dy}{dx} = \frac{1}{2\sqrt{x}}.$$



**Figure 4.4:** The distributions  $f(x)$  and  $h(y)$ .

### 4.3. Some important distributions

Using Eq. (4.15),

$$h(y) = f(x) \frac{dx}{dy} = \frac{e^{-x}}{2\sqrt{x}}.$$

The last step is to express the right-hand side in terms of  $x = y^2$ :

$$h(y) = ye^{-y^2}.$$

If  $f(x)$  is the distribution of energy, then  $h(y)$  is the distribution of magnitude of the velocity (within proportionality constants that are absorbed into  $x$  and  $y$ ). These distributions are shown in Fig. 4.4.

### 4.3 Some important distributions

Of the large and in fact, unlimited number of distributions one may encounter, there are a few that make frequent appearances in statistical thermodynamics and whose basic properties are worth reviewing.

**Exponential distribution** The exponential distribution is of the general form

$$f(x) = \lambda e^{-\lambda x}, \quad (4.16)$$

where  $\lambda$  is a positive constants and  $x$  ranges from 0 to  $\infty$ . It is easy to confirm that this distribution is properly normalized. This is a none-parameter distribution and all of its moments are expressed in terms of  $\lambda$ :

$$\mu_k = \frac{k!}{\lambda^k}.$$

The mean and the variance follow from the above result and we find

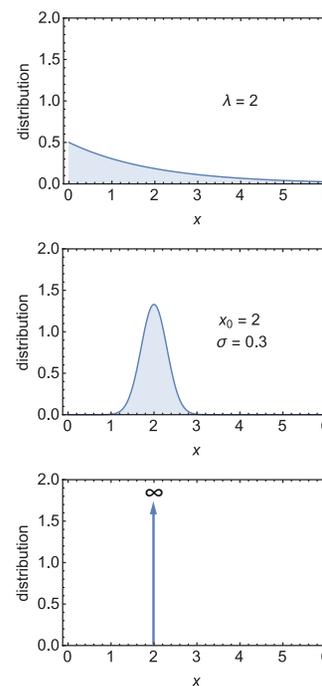
$$\bar{x} = \mu_1 = 1/\lambda, \quad \text{var}(x) = \mu_2 - \mu_1^2 = 1/\lambda^2.$$

The standard deviation, which is given by the square root of the variance, is equal to the mean – this indicates a very broad distribution. Indeed, unlike most other distributions that are more or less bell-shaped and peak near the mean, the exponential distribution peaks at  $x = 0$  and has a tail that persists to values of  $x$  that are several times larger than the mean.

**Gaussian distribution** The Gaussian distribution is a two-parameter distribution and is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-x_0)^2/2\sigma^2}. \quad (4.17)$$

It is normalized to unit area, its mean is  $x_0$  and its variance is  $\sigma^2$ . The Gaussian distribution has a symmetric bell-shaped form whose maximum is at  $x_0$  and whose width at half size is approximately  $\sigma$ . The distribution falls rapidly to zero outside an interval of approximately  $\pm 4\sigma$  around the mean. Even though mathematically the function is defined in the entire range  $(-\infty, +\infty)$ , in practice only the region of a few standard deviations about the mean contributes to any significant degree. This distribution is often used to model variables that are strictly positive on grounds that if its variance is small the distribution is concentrated into a small finite region around the mean. The fact that the distribution is



**Figure 4.5:** Exponential (top), Gaussian (middle) and schematic representation of Dirac delta (bottom) distribution. All three distributions have the same average,  $\bar{x} = 2$ .

normalized to unit area with mean and variance  $x_0$  and  $\sigma^2$ , respectively, means that the following identities can be written:

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-(x-\bar{x})^2/2\sigma^2} = 1 \quad (4.18)$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-(x-\bar{x})^2/2\sigma^2} = x_0, \quad (4.19)$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-x_0)^2 e^{-(x-\bar{x})^2/2\sigma^2} = \sigma^2. \quad (4.20)$$

These integrals are useful to memorize.

**Dirac delta function** The Dirac delta function can be viewed as the limit form of a Gaussian distribution whose variance goes to zero:

$$\delta(x-x_0) = \lim_{\sigma^2 \rightarrow 0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-x_0)^2/2\sigma^2}. \quad (4.21)$$

We visualize the Dirac delta as an infinitely sharp function at  $x = x_0$  that is zero everywhere else, and whose integral from  $-\infty$  to  $\infty$  is 1. The singular nature of this distribution endows it with a simple and very useful mathematical property: The average value of any function  $A(x)$  over the Dirac delta distribution is the value of  $A$  at  $x_0$ :

$$\bar{A} = \int_{-\infty}^{\infty} A(x) \delta(x-x_0) dx = A(x_0). \quad (4.22)$$

That is, integration of a function over a Dirac delta amounts to picking out the value of that function at the singular point of the Dirac delta. We may use this property to calculate the moments of the Dirac delta:

$$\mu_k = \int_{-\infty}^{\infty} x^k \delta(x-x_0) = x_0^k.$$

For the variance in particular we find

$$\text{var}(x) = \mu_2 - \mu_1^2 = 0,$$

which simply states that the distribution is infinitely narrow. Experimentalists working with size distributions call this distribution “monodisperse” to indicate that all of its elements have exactly the same size,  $x_0$ .

#### 4.4 Combinatorics

A frequent problem in probability theory as well as in statistical mechanics is counting discrete finite structures, for example, discrete microstates available to a finite number of particles. This requires use of basic combinatorial tools that are reviewed below.

#### 4.4.1 Permutations

A permutation is an ordered arrangement of objects. We can calculate this number by creating an ordered list one object at a time. Starting with  $N$  objects, we have  $N$  choices for the first object,  $N - 1$  choices for the second object and so on. The number of permutations is equal to their product:

$$(\# \text{ of permutations}) = N \cdot (N - 1) \cdots 1 \equiv N!, \quad (4.23)$$

which is known as the factorial. The factorial is closely related to the Gamma function, defined as

$$\Gamma(x) = \int_0^{\infty} z^{x-1} e^{-z} dz. \quad (4.24)$$

If  $x$  is a positive integer, then

$$\Gamma(x) = (x - 1)!.$$

The Gamma function may be considered an generalization of the factorial to all real numbers.

**Stirling's formula** A very useful asymptotic expression for the factorial is given by Stirling's formula,

$$x! \approx \sqrt{2\pi x} \left(\frac{x}{e}\right)^x. \quad (4.25)$$

As a true asymptote, the ratio of the two sides of this equation approaches 1 as  $x \rightarrow \infty$ . Nonetheless, the accuracy of this equation is remarkable even with small values of  $n$ , as seen in Fig. 4.6. Taking the logarithm of both sides, Eq. (4.25) gives

$$\log x! = x \log x - x + \log \sqrt{2\pi x}.$$

Compared to the other two terms,  $\log \sqrt{2\pi x}$  grows very slowly with  $x$  and can be dropped. This leads to the following approximation for the logarithm of the factorial,

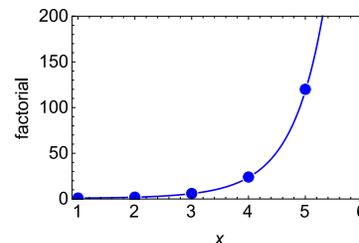
$$\log x! \approx x \log x - x. \quad (4.26)$$

This is an excellent approximation for  $\log x!$  though not for  $x!$  itself.<sup>3</sup> We will mostly be interested in expressions relating to the log of the multinomial coefficient and in such cases Eq. (4.26) suffices. Unless we specify otherwise, when we talk of the Stirling formula we will refer to Eq. (4.26).

#### 4.4.2 Multinomial coefficient

The factorial gives the number of permutations in a list of  $N$  objects that are distinguishable from each other. Suppose that the objects are grouped into several groups such that only objects that belong to different groups are distinguishable from each other while objects in the same group are indistinguishable. For example, the objects might be balls of different colors but otherwise identical in size, weight, texture etc. We can only distinguish two balls if their colors are different. The problem is to calculate the number of permutations of  $N$  objects such that  $n_1$  belong in group (or color) 1,  $n_2$  belong in group 2, and so on. Of course we must have

$$n_1 + n_2 + \cdots = N.$$



**Figure 4.6:** Comparison between the factorial (circles) and Stirling's asymptotic Eq. (4.25), shown by the continuous line.

<sup>3</sup>Using the full asymptotic form of the factorial we obtain

$$\frac{x!}{e^{x \log x - x}} = \sqrt{2\pi x}.$$

This shows clearly that Eq. (4.26) produces an unacceptable approximation for the factorial itself (the error increases without bound) even though it provides an excellent approximation for its logarithm.



There are  $M - 1$  spaces between objects and  $N - 1$  of them must have bars. The number of ways to choose where to draw the bars is equal to the number of ways to choose  $N - 1$  spaces out of  $M - 1$ :

$$V_{M,N} = \binom{M-1}{N-1}. \tag{4.29}$$

**Example 4.3.** *Permutations of colored balls*

We have four balls of which two are red, one is white and one is blue. Enumerate the permutations.

**Solution** Here  $n_R = 2$ ,  $n_W = 1$  and  $n_B = 1$  ( $n_R + n_W + n_B = 4$ ). The number of permutations is

$$\frac{4!}{2!1!1!} = 12.$$

These are enumerated below:

RRWB	RRBW	RWRB	RWBR
RBRW	RBWR	WRRB	WRBR
WBRR	BWRR	BRRW	BRWR

**Example 4.4.** *Paint four balls in three colors*

Count the number of ways to paint four balls, two in red, one in white and one in blue.

**Solution** We assign the four balls into three groups: the “red” group contains two balls, the “white” groups with one ball, and “blue” group with 1 ball. The number of ways to make the assignments is

$$\frac{4!}{2!1!1!} = 12.$$

The ways are enumerated below (balls are numbered 1 through 4, the colors are shown in the header):

R	W	B	R	W	B
1,2	3	4	2,3	1	4
1,2	4	3	2,3	4	1
1,3	2	4	2,4	1	3
1,3	4	2	2,4	3	1
1,4	2	3	3,4	1	2
1,4	3	2	3,4	2	1

Comparing with the previous example we see that the number of ways to choose a ball and its color is equal to the permutations of the painted balls.

**4.4.3 Stirling’s approximation for the multinomial coefficient**

Taking the log of the multinomial and applying Stirling’s approximation  $\log x! = x \log x - x$  to all factorials we obtain

$$\begin{aligned} \log W &= N \log N - N - \sum_i (n_i \log n_i - n_i) = \\ &= \sum_i n_i \log N - \sum_i n_i \log n_i = - \sum_i n_i \log \frac{n_i}{N} \end{aligned}$$

Setting  $p_i = n_i / N$ , this is written in the more compact form as

$$\log W = -N \sum_i p_i \log p_i \quad (4.30)$$

In a typical problem,  $n_i$  is the number of objects of type  $i$  in the collection of  $N$  objects; accordingly,  $p_i$  is the fraction (or probability) of objects of type  $i$  in the collection. The right-hand side of Eq. (4.30) appears prominently in statistical thermodynamics. Since the result was obtained using Stirling approximation for  $\log n!$ , Eq. (4.30) is valid for the logarithm of the multinomial coefficient but not for the coefficient itself.

#### 4.4.4 Multinomial Distribution

Suppose we have a die with  $K$  sides so that each throw may produce any integer between 1 and  $K$  with probability  $p_i$ , not necessarily the same for all outcomes. We throw the dice  $N$  times, what is the probability that we will obtain  $n_1$  one's,  $n_2$  two's and so on? We are interested in predicting the total number of times a number shows up, not in the order in which the successive outcomes occur. The probability to get a specific sequence of outcomes that contains  $n_1$  one's,  $n_2$  two's and so on is given by the product of probabilities

$$p_1^{n_1} p_2^{n_2} \dots$$

However, the same number of outcomes can appear in all possible permutations, a number that is given by the multinomial coefficient,

$$\frac{N!}{n_1! n_2! \dots},$$

where of course  $n_1 + n_2 + \dots = N$ . The answer then is

$$P(n_1, n_2, \dots) = N! \binom{p_1^{n_1}}{n_1!} \binom{p_2^{n_2}}{n_2!} \dots = N! \prod_i \frac{p_i^{n_i}}{n_i!}. \quad (4.31)$$

This is the multinomial distribution and derives its name from its relationship to the multinomial expansion. Recall that this expansion is given by,

$$(p_1 + p_2 + \dots)^N = \sum_{n_1 + n_2 + \dots = N} \frac{N!}{n_1! n_2! \dots} p_1^{n_1} p_2^{n_2} \dots,$$

with the summation on the RHS running over all non-negative integers  $n_i$  whose sum is equal to  $N$ . As we see, the terms of this expansion are the distributions corresponding all possible outcomes of tossing a  $K$ -sided die  $N$  times. Since the sum of  $p_i$  on the LHS is 1, the above result can be written in the form

$$\sum_{n_1 + n_2 + \dots = N} \frac{N!}{n_1! n_2! \dots} p_1^{n_1} p_2^{n_2} \dots = 1,$$

which states that the multinomial distribution is properly normalized. In statistical thermodynamic the multinomial distribution appears when we are dealing with mixtures of particles ( $n_i$  particles of type  $i$ ,  $i = 1, 2, \dots$ ) that do not interact with each other. Then the probability to find a system of  $N$  particles at a particular state is the product of the probabilities to find the individual particles in their corresponding states.

**Binomial Distribution** If there are only two outcomes 1 & 2 with probabilities  $p_1 = p$  and  $p_2 = 1 - p$ , respectively, the multinomial distribution reduces to the binomial distribution

$$P(N, n_1) = \frac{N!}{(N - n_1)!n_1!} p^{n_1} (1 - p)^{N - n_1}$$

which gives the probability of  $n_1$  events of type 1 in  $N$  trials. If we interpret outcome 1 to be “success” and 2 “failure,” the binomial distribution gives the probability of  $n_i$  successes in  $N$  independent trials. For a fair coin we have  $p_1 = p_2 = 1/2$  and the binomial distribution then becomes

$$P(N, n_1) = \frac{1}{2^N} \frac{N!}{(N - n_1)!n_1!}.$$

For large  $N$  this approaches a Gaussian distribution that is sharply peaked at  $n_1 = N/2$ : indeed, for a large number of trials we expect that half of them will lead to success. The mathematical proof was given in section ??, where we determined that the binomial coefficients approach a Gaussian distribution. If  $p_1 \neq p_2$  the distribution again approaches a Gaussian distribution but it is now centered at the expected number of successes,  $p_1 N$ . This proof is entirely analogous to that for  $p_1 = p_2$ .

**Example 4.5.** *Multinomial distribution*

A fair die is tossed 5 times. What is the probability that the sum of the 5 tosses is equal to 10?

**Solution** The probability of  $N$  tosses of a die is a simple example of the binomial distribution. If we define the outcome of  $N$  tosses as the sequential order the outcome of each individual toss in the order it was obtained, then each outcome is equally probable with probability  $p^N = (1/6)^N$ , where  $p = 1/6$  is the probability of each individual outcome. If we group outcomes by the distribution  $\mathbf{n} = (n_1, n_2, \dots)$  where  $n_i$  is the number of times a toss returned the result  $i = 1, \dots, 6$ , then the probability to observe distribution  $\mathbf{n}$  is

$$P(\mathbf{n}) = \mathbf{n}! p^N.$$

For a regular die with 6 sides  $p = 1/6$  but one may imagine a fair die with  $K$  sides ( $p = 1/K$ ) or a biased die with  $K$  sides such that  $p_i$  is not necessarily the same for all sides.

To return to the problem at hand, one possible way to receive a total of 10 is with one 6 and four 1's. These five numbers can appear in any order. To calculate the total number of ways to obtain this particular outcome, we first write the distribution of individual tosses,

$$n_1 = 4, n_2 = n_3 = n_4 = n_5 = 0, n_6 = 1,$$

or in more condensed form,

$$\mathbf{n}(4, 0, 0, 0, 0, 1).$$

The number of ordered sequences is equal to the multinomial factor,

$$\mathbf{n}! = \frac{5!}{4!0!0!0!0!1!} = 5$$

and the probability to observe this distribution is

$$P(\mathbf{n}) = \mathbf{n}! p^N = (5) \left(\frac{1}{6}\right)^5 = 0.000643004$$

## 4.5. Central limit theorem

We need to repeat this calculation for every possible sequence of 5 numbers between 1 and 6 whose sum is 10. The probability to obtain a sum equal to 10 then is

$$P(10) = \left( \sum_{\mathbf{n}} \mathbf{n}! \right) \frac{1}{6^5},$$

with the summation taken over the multinomial coefficients of all distributions of outcomes whose sum is equal to 10. The results are summarized below.

outcomes	$\mathbf{n}$	$\mathbf{n}!$
6 1 1 1 1	4 0 0 0 0 1	5
5 2 1 1 1	3 1 0 0 1 0	20
4 3 1 1 1	3 0 1 1 0 0	20
4 2 2 1 1	2 2 0 1 0 0	30
3 3 2 1 1	2 1 2 0 0 0	30
3 2 2 2 1	1 3 1 0 0 0	20
2 2 2 2 2	0 5 0 0 0 0	1

126

The first 5 columns show the possible ways to obtain a sum of 10 with 5 integers drawn between 1 and 6, the next 6 columns give the distribution of these integers, and the last column is the multinomial coefficient of the distribution. The sum of the multinomial coefficients is 126 and the probability to obtain a sum equal to 10 is finally calculated to be

$$P(10) = (126) \frac{1}{6^5} = 0.0162037,$$

or about 1.6%.

### 4.5 Central limit theorem

Suppose we have a random variable  $x$  with mean  $\bar{x}$  and variance  $\text{var}(x) = \sigma$ . We construct a new variable  $X_N$  as the sum of  $N$  independent values  $x_i$ ,  $i = 1, \dots, N$ ,

$$X_N = x_1 + x_2 + \dots + x_N, \tag{4.32}$$

The central limit theorem states that for large  $N$  the probability distribution of  $X_N$  approaches a Gaussian distribution with mean

$$\bar{X}_N = N\bar{x},$$

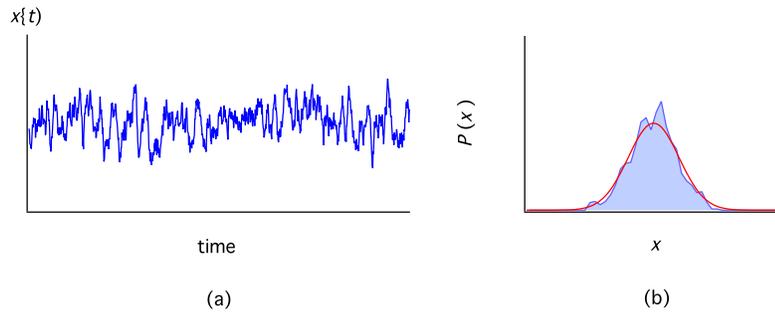
and variance

$$\text{var}(X_N) = N\text{var}(x) = N\sigma^2.$$

This asymptotic behavior is true regardless of the details of the probability density function of  $x_i$ . The only requirements are that all  $x_i$  are drawn from the same distribution, and that  $N$  is large, but fixed. The generality of this theorem<sup>4</sup> is remarkable and has several practical implications. The square root of the variance is the standard deviation and a measure of the variation of  $X$  around its mean value. Its ratio to the mean is

$$\frac{\text{var}(X_N)}{\bar{X}_N} = \frac{\bar{x}/\sigma}{\sqrt{N}}. \tag{4.33}$$

<sup>4</sup>The proof is actually fairly short but requires the use of the characteristic function, a property of probability distributions that we have not introduced.



**Figure 4.7:** (a) Stationary time series and (b) its probability distribution.

Since  $\bar{x}$  and  $\sigma$  are constant, this ratio goes to 0 as  $N$  goes to  $\infty$ . This says that while fluctuations in  $X_N$  increase with  $N$ , when measured relative to the mean  $\bar{X}_N$  they vanish with increasing  $N$ .<sup>5</sup>

Let us define another random variable,  $y$ , as

$$y = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{X_N}{N}.$$

This is the formula for the average of a set of  $N$  measurements.  $y$  is itself a random variable. Let us calculate its mean and variance. For the mean we have

$$\bar{y} = \left( \frac{X_N}{N} \right) = \frac{\bar{X}_N}{N} = \frac{N\bar{x}}{N} = \bar{x}.$$

That is, the mean of  $y$  is indeed equal to the mean of the original variable  $x$ . The variance of  $y$  is<sup>6</sup>

$$\text{var}(y) = \text{var}(X_N/N) = \frac{\text{var}(X_N)}{N^2} = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}.$$

The standard deviation is

$$\sqrt{\text{var}(y)} = \frac{\sigma}{\sqrt{N}},$$

and this goes to zero as  $N$  goes to infinity. We conclude that the arithmetic average is a set of random variables converges to the true mean of the sampled variables, and that the accuracy of this estimation improves as  $1/\sqrt{N}$  with the number of points used in the calculation. This is one consequence of the central limit theorem that finds wide application in statistics. For our purposes, the implication of the central limit theorem that matters is this: a random variable that is formed by accumulation of another random variable becomes progressively less random, in the sense that its variance relative to the mean vanishes as we accumulate more values.<sup>7</sup>

## 4.6 Stochastic process

A stochastic process is a dynamic process in time whose outcome at time  $t$  is a random variable. An example would be the coordinates in space of a Brownian

<sup>6</sup>If  $\sigma^2$  is the variance of  $x$ , then the variance of  $cx$ , where  $c$  is a constant, is  $c^2\sigma^2$ .

<sup>7</sup>For a simple demonstration consider the average fraction of heads in  $K$  tosses of a fair coin. For small  $K$  that fraction fluctuates between independent realizations of the experiment but for large  $K$  it converges to a constant.

particle or the valuation of the stock market as a function. The problems that will be of interest to us are the Brownian particle type in the sense that we will be able to repeat the process under identical preparation as many times as we wish (we cannot say the same for the stock market). If we record the measured quantity  $x(t)$  (e.g., coordinates of the Brownian particle) as a function of time we obtain a time series, namely the trace of  $x$  in time. The value of  $x$  at some point in time  $t$  is a random variable and is characterized by a probability distribution,  $p(x; t)$ .<sup>8</sup> The stochastic process is stationary if the probability does not depend on time. In this case the probability  $p(x)$  may be identified as the fraction of time along the time series that the random variable assumes the value  $x$ . This implies that the statistical weight of each value of  $x$  in the time series is equal to its duration  $dt$ . Imagine a situation where we have an ensemble of time series  $x_i(t)$ ,  $i = 1, \dots$ , recorded under identical conditions. We fix time to  $t = t'$  and collect all the  $x_i(t')$  to obtain a data set of values corresponding to the same time  $t'$ . If the process is stationary, this probability is the same as the one that is calculated from the analysis of a single time series along the time axis. We now have two alternative ways to calculate the mean of a function  $A(x)$ :

$$\bar{A} = \int_x p(x)A(x)dx = \frac{1}{t_2 - t_1} \int_{t=t_1}^{t=t_2} A(x(t))dt, \quad (t_2 - t_1 \rightarrow \infty). \quad (4.34)$$

The first equation is the direct application of the definition of  $\bar{A}$  and involves an integral in  $x$ . The second equation is based on the fact that the statistical weight of  $x$  in the time series is  $dx$  and involves integration in time. The result is independent of the start and end points of the integration,  $t_1$  and  $t_2$ , and only requires the time interval  $t_2 - t_1$  to be long enough for statistical accuracy.

### 4.7 Monte Carlo sampling

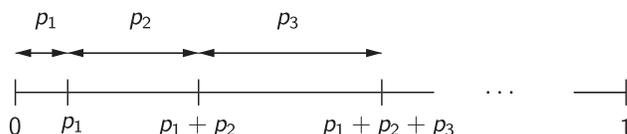
The name Monte Carlo refers to a family of numerical methods for the computer simulation of stochastic processes. At the heart of calculation is the *random number*, a real number that is uniformly distributed between 0 and 1. Several algorithms are available to mimic random numbers<sup>9</sup> but most computational software today comes with built-in random number generators. The name Monte Carlo alludes to games of chance, but it is “chance” rather than “game” that is the important concept here. The idea behind stochastic simulations is to use the computer to replicate the outcome of a real process and obtain various averages by collecting a large number of data points. A simple example is a stochastic process that produces outcomes  $A, B \dots$  with probabilities  $P_A, P_B \dots$ . Here is one way to do this on the computer. Draw a random number  $\text{rnd}$  between 0 and 1, then choose the outcome  $X$  of the toss according to the rules:

$$\begin{aligned} &\text{if } \text{rnd} < P_A, \text{ then choose } A \\ &\text{if } P_A < \text{rnd} < P_A + P_B, \text{ then choose } B \\ &\text{if } P_A + P_B < \text{rnd} < P_A + P_B + P_C, \text{ then choose } C \\ &\text{etc.} \end{aligned} \quad (4.35)$$

The idea is this: divide the interval  $(0, 1)$  into successive intervals whose length is equal to the given probabilities. Since  $\text{rnd}$  is uniformly distributed in  $(0, 1)$ ,

<sup>8</sup>In the case of the Brownian particle, the recorded property is a vector  $(x(t), y(t), z(t))$ , with probability distribution  $p(x, y, z; t)$ .

<sup>9</sup>Technically, these numbers are *pseudo random* in the sense that they are produced by deterministic algorithms but they are highly randomized and suitable for practical applications.



**Figure 4.8:** Graphical representation of Monte Carlo simulation of discrete stochastic variable with arbitrary probability distribution.

the probability to fall in the interval  $(0, P_A)$  is equal to  $P_A$ , and so on, as shown schematically in Fig. 4.8).

#### 4.7.1 The Metropolis method

In Monte Carlo simulations the typical scenario is generally the same: we want to generate a sample of a stochastic variable based on knowledge of its probability distribution. In the previous demonstration the situation was fairly straightforward, there are three possible states and their probabilities are known. Some problems are considerably more complicated. Here is an example: toss a fair die 100 times but only consider results for which the sum of all tosses is 325. There are about  $2 \times 10^{85}$  different ways to form the number 325 by adding 100 tosses of a die.<sup>10</sup> They are all equally probable, which means that the probability of each outcome is  $0.5 \times 10^{-85}$ . To apply the methodology we just described we would have to identify the  $10^{85}$  different outcomes and divide the interval  $(0, 1)$  into an equal number of sub-intervals, an impossible numerical task. A different method is needed to sample states from enormous phase spaces like this one. Obviously we cannot expect to visit *every* possible state when the phase space is so large, but we do want to be able to visit a representative sample of states from which we may calculate reliable averages of various quantities of interest. One simple method that accomplishes this is the Metropolis algorithm. This is a two step algorithm. Starting from the current state, the first step is to identify another state as a possible destination; the second step is to decide if the destination is accepted. If it is, we add it to the list of states that have been visited; if it is not accepted, we add the current state on the list, and repeat. If the current state is  $s_1$  and the target state is  $s_2$ , the Metropolis method accepts the new state with probability

$$P(s_2|s_1) = \min \left\{ \frac{P(s_2)}{P(s_1)}, 1 \right\} \quad (4.36)$$

where  $P(s_1)$  and  $P(s_2)$  are the probabilities of the current and target states, respectively. Algorithmically this is implemented by drawing a random number  $r$  from a uniform distribution in the interval  $(0, 1)$ . If  $r \leq P(s_2|s_1)$  the transition is accepted, otherwise it is rejected and the current state is sampled a second time. If the target state is more probable than the current state then  $P(s_2)/P(s_1)$  and the acceptance probability is 1. This means that the transition to a more probable state is always accepted. If the target state is less probable, the transition is accepted with probability  $P(s_2)/P(s_1)$ . Accordingly, states of low probability are visited less frequently. To complete the algorithm we need an unbiased way to

<sup>10</sup>The exact number is  
19475627808556227046408895717571478  
12300758906664242365493057405646064  
8550849233122720 and is given by  $V_{325,100}$  in  
Eq. (4.29). Even if we could write down a  
trillion of these per nanosecond, to collect all  
of them would require an impossible  $10^{60}$   
years (the universe is only about 15 trillion  
years old).

pick target states. One way to do this is to choose among all possible states with uniform probability. We demonstrate with two examples, one fairly simple, the other one a bit more complicated.

**Example 4.6.** *Metropolis sampling-1*

A random variable receives three possible values,  $A$ ,  $B$  and  $C$ , with probabilities  $P_A = 0.1$ ,  $P_B = 0.4$ ,  $P_C = 0.5$ . Use the Metropolis method to obtain a sample of size  $N = 100$ . Calculate the frequency of  $A$ ,  $B$  and  $C$  in the sample and compare with the given probabilities.

**Solution** We implement the sampling algorithm as follows:

1. Given the current state (say  $A$ ), pick a new state with equal probability among the other two states ( $B$  or  $C$ )
2. Draw a random number  $r$  in  $(0,1)$ . If

$$r \leq \min \left\{ \frac{P(\text{new state})}{P(\text{current state})}, 1 \right\}$$

accept the new state and add it to the list of visited states; otherwise, add the current state to the list

3. Return to step 1 and repeat

At each step of the method we collect a state; this is either a new state, if the transition was accepted, or the old state, if the transition was rejected. The acceptance probabilities for the six possible transitions of this algorithm are:

$$\begin{array}{ll} P_{AB} = 1 & P_{AC} = 1, \\ P_{BA} = 0.25 & P_{CB} = 1, \\ P_{CA} = 0.2 & P_{CB} = 0.8 \end{array}$$

In this implementation we do not consider explicit transitions that return back to the same state. Such transitions occur implicitly when a transition to a new state is rejected. To start we need an initial state, which can be picked arbitrarily.

In a sample of 100 realizations of this random variable we find the following results ( $n_i$  is the number of times state  $i = A, B$  or  $C$  appears,  $n_i/N$  is the estimated probability of  $i$  based on the sample and  $P_i$  is the known probability):

	$A$	$B$	$C$
$n_i$	17	33	50
$n_i/N$	0.17	0.33	0.5
$P_i$	0.1	0.4	0.5

With  $N = 10000$  the estimated probabilities are much closer to the correct values:

	$A$	$B$	$C$
$n_i$	1024	4047	4929
$n_i/N$	1.024	1.01175	0.9858
$P_i$	0.1	0.4	0.5

*Comments:* In this example the number of states is very small and it is easy to select the next state among all states that are possible. Notice that we have not included the current state as an explicit possible destination. It is possible to modify the algorithm so that the next state to be considered is picked with equal probability among all possible states but it is more efficient to attempt to visit a different state from the current one.

## 4.7. Monte Carlo sampling

**Example 4.7.** *Metropolis-2*

40 pieces are placed on an  $8 \times 8$  chessboard so that a square can contain at most one piece. The probability of arrangements is

$$P = Ce^{-N_{11}}$$

where  $N_{11}$  is the number of adjacent pairs of squares that are both occupied. The proportionality constant  $C$  is not known. Simulate this by the Metropolis method and calculate the average number of occupied pairs,  $\bar{N}_{11}$ .

**Solution** The exact number of configurations in this problem is

$$\binom{64}{40} = 250649105469666120 \approx 2.5 \times 10^{17},$$

which is enormously large. One simple way to navigate through this space is to consider transitions to configurations in the vicinity of the current configuration. We do this by selecting an occupied site at random and moving its piece to a vacant site that is also selected randomly. The new configuration differs only in the placement of one single particle and offers a systematic albeit slow method to travel through phase space. Once we have a new configuration, we must still decide whether to transition to it or not. The acceptance probability for the transition is

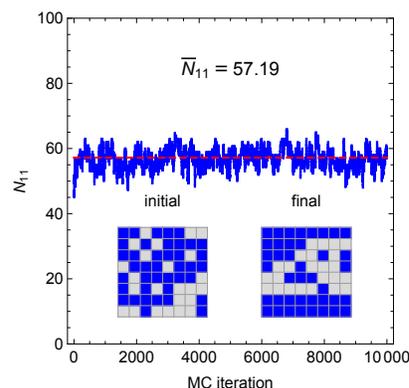
$$P(\text{accept transition}) = \min \left\{ 1, e^{-N'_{11}/N_{11}} \right\}$$

where  $N_{11}$  is the number of nearest-neighbor pairs that are both occupied by a piece in the current configuration and  $N'_{11}$  is the same number in the new configuration. If the transition is accepted, the new configuration is added to the list of visited states; if it is rejected, the current configuration is added to the list. This is repeated until enough statistics have been collected. The method requires an initial configuration, which can be constructed by randomly placing the pieces on the chessboard.

Figure 4.9 shows the number of nearest neighbor pairs over 10,000 Monte Carlo steps. The average number of pairs is 57.19. Clearly, in 10,000 steps the simulation visits a tiny fraction of the phase space (there are more than  $10^{17}$  configurations). Nonetheless, the sample is sufficiently representative to provide a good estimate of the average number of occupied nearest neighbors.

*Comments* This example is typical in two generic respects of those we will encounter in statistical thermodynamics: First, the phase space is very large. The enormity of the phase space motivates us to consider transitions to a much smaller subset of states in the vicinity of the current state. The advantage is that the process is guaranteed to converge to the correct mean number of 11 pairs and more generally to the correct number of almost any mean property. The disadvantage is that each transition represents a small step away from the previous state, which in turn introduces correlations that take a large number of steps to decay.

Second, the probability of state is not only very small, but it is not known exactly because the common proportionality constant is unknown. We could determine this constant by the normalization condition but the enormity of the phase space makes this an impossible (not just difficult!) task. The Metropolis method is based on *ratios* of these probabilities, which renders the proportionality constant irrelevant and also avoids problems with the small magnitude of the absolute probabilities.



**Figure 4.9:** Monte Carlo simulation for Example 4.7 that shows the number of occupied nearest-neighbors over 10,000 Monte Carlo steps. The inset shows the initial and final configurations. The mean number of 11 pairs quickly converges to a stationary value (red line) but to obtain this value with sufficiently accuracy we must accumulate data for sufficiently long time.

### 4.7.2 Markov chains

The Metropolis method produces a sequence of states  $s$ ,

$$\cdots s_n \xrightarrow{p_{n,n+1}} s_{n+1} \cdots$$

in which the probability to obtain a particular state in step  $n+1$  depends entirely on the state in step  $n$ . Here  $s_n$  is the state in step  $n$  chosen from all possible states in phase space. Such sequence is called a Markov chain and is defined by a fixed set of transition probabilities  $p_{ij}$ . These probabilities are independent of  $n$  and depend only on the end states. They satisfy the normalization condition

$$\sum_j p_{ij} = 1 \quad \text{for all } i,$$

which states that all transitions departing from a fixed site must arrive somewhere in the same phase space. The Markov chain describes a stationary stochastic process that produces a sample of random variable  $s$ . The frequency of sampled states converges to the probability  $P_i$  of state  $i$  and is given asymptotically by the recursion

$$[P_i]_n = \sum_j p_{ij} [P_j]_{n-1}. \quad (4.37)$$

The summation on the right-hand side goes over all transitions in the previous step that arrive at state  $i$ . Each term of the summation on the right-hand side is the probability to find the system in state  $j$  times the probability to transition. The summation goes over all states  $j$  that can transition to  $i$ , but since  $p_{ij} = 0$  for those states that do not have a direct transition to  $i$ , we may as well calculate the summation over the entire phase space. The probability  $[P_i]_n$  changes with  $n$  but eventually converges to a constant that represents the probability of state  $i$ .

The Markov chain is an important tool in probability theory that is very useful in modeling stochastic processes. An example would be the arrival and departure of people in a queue (or the addition and removal of monomers from a polymer chain), which can also serve as a model for the attachment and detachment of monomers at the ends of a linear polymer chain. In such problems the transition probabilities are determined by the kinetics of the process and the goal is to obtain the distribution of queue lengths, or molecular weights. But another very useful application is in sampling a large phase space. In this case we know the probabilities of the various states and want to obtain a representative *sample*, which we may then use to calculate various averages. In this case the sequence of states does not represent the outcome of a physical process but rather a sampled sequence of states. To obtain the correct probability of states we need to define the transition probabilities correctly. We do this using the condition of *detailed balance*. This is a special condition that is imposed on the transition probabilities. It is given by

$$P_i p_{ij} = P_j p_{ji} \quad \text{all } i \text{ and } j, \quad (4.38)$$

and states that the rate of arrivals to state  $j$  from state  $i$  is equal to the rate of arrivals to state  $i$  from state  $j$ . This means that all rates to- and from- are matched exactly by their reverse rates and ensures that the probabilities converge to  $P_i$ :

$$\lim_{n \rightarrow \infty} [P_i]_n = P_i. \quad (4.39)$$

If we construct transition probabilities that satisfy detailed balance in Eq. (4.38), then the sampled states will appear with frequencies that match probability  $P_i$ . The transition probabilities of the Metropolis method satisfy detailed balance. To obtain these transition probabilities we first note that Eq. (4.36) gives the *acceptance* probability once the destination state has been chosen. To obtain the transition probability we must multiply the acceptance probability by the probability to choose a particular destination. Since we choose the destination randomly, this probability is  $1/\mathcal{N}$ , where  $\mathcal{N}$  is the number of possible destination from the current state. The Metropolis transition probability then is

$$p_{ij} = \frac{1}{\mathcal{N}} \min \left\{ \frac{P_j}{P_i}, 1 \right\}. \quad (4.40)$$

Suppose that  $P_j > P_i$ , then  $p_{ij} = 1/\mathcal{N}$ ,  $p_{ji} = P_i/(P_j \mathcal{N})$ , and the detailed balance condition is satisfied:

$$P_i \left( \frac{1}{\mathcal{N}} \right) = P_j \left( \frac{1}{\mathcal{N}} \frac{P_i}{P_j} \right).$$

As we see, detailed balance also requires that we have the same number of target states from any departure state. Indeed this is the case in both examples of the Metropolis method given above.

#### 4.8 Additional reading

The discussion of probabilities in this chapter is very brief and cannot do justice to the topic. Although the part of probability theory we need to develop statistical thermodynamics is fairly small, it is important to have a good sense of how probabilities and distributions work. If you did not have an undergraduate class in probabilities and statistics you should consult a textbook in this area. A nice short book is [Rozanov \(1977\)](#). For an even shorter primer, read the first chapter in [van Kampen \(1990\)](#). For those who want a more expanded treatment, a very good resource is *All of Statistics* by [Wasserman \(2004\)](#). This book covers in a series of brief chapters many topics in probabilities and statistics. The treatment takes a more mathematical approach than the other books mentioned here, but emphasizes concepts over derivations and should make for comfortable reading.

#### 4.9 Problems

**Problem 4.1.** a) Suppose  $y(t)$  is a monotonic function of time between  $t_1$  and  $t_2$ . Calculate the fraction of time that  $y$  spends at any particular value  $y$  between  $y_1 = y(t_1)$  and  $y_2 = y(t_2)$ . In other words calculate  $f(y)$  such that  $f(y)dy$  is the probability to find  $y$  in  $(y, y + dy)$  if we pick  $t$  randomly from  $(t_1, t_2)$ .

b) The velocity of the linear harmonic oscillator is given by

$$v(t) = v_0 \cos(at),$$

where  $v_0$  and  $a$  are constants. Determine the probability distribution of  $v$ .

c) Obtain a random sample of  $\cos(x)$  in the interval  $(0, 4\pi)$  of 4,000 points, calculate the distribution of the sampled values and compare with the theoretical result.

## 4.9. Problems

**Problem 4.2.** a) Suppose  $y(t)$  is a monotonic function of time between  $t_1$  and  $t_2$ . Calculate the fraction of time that  $y$  spends at any particular value  $y$  between  $y_1 = y(t_1)$  and  $y_2 = y(t_2)$ . In other words, calculate  $f(y)$  such that  $f(y)dy$  is the probability to find  $y$  in  $(y, y + dy)$  if  $t$  is picked randomly from  $(t_1, t_2)$ .

b) The velocity of the linear harmonic oscillator is given by

$$v(t) = v_0 \cos(at),$$

where  $v_0$  and  $a$  are constants. Determine the probability distribution of  $v$ .

c) Obtain a random sample of  $\cos(x)$  in the interval  $(0, 4\pi)$  of 4,000 points, calculate the distribution of the sampled values and compare with the theoretical result.

**Problem 4.3.** Figure 4.10 shows the evolution of stochastic variable  $y$ . We may take the graph to represent the state of an equilibrium system as a function of time.

a) Construct the probability density function  $P(y)$  and make a graph. Experiment with the width of the bins until you get a “good” representation of the distribution.

b) Determine the probability to find the state of the system between  $y_1 = 14$  and  $y_2 = 16$ . (The data will be available to download in spreadsheet format from the course web site.)

**Problem 4.4.** a) Show that the number of ways to assemble a cluster distribution  $\mathbf{n} = (n_1, n_2, \dots)$  from  $M$  distinguishable monomers is

$$\omega(\mathbf{n}) = M! \prod_i \left( \frac{1}{i!} \right)^{n_i} \frac{1}{n_i!},$$

where  $n_i$  is the number of clusters that contain  $i$  monomers

b) Extend the previous result for the case that a cluster with  $i$  monomers can be produced in  $a_i$  distinct ways and confirm that when  $a_i = 1$  you recover the previous result.

**Problem 4.5.** A system consists of three particles and each particle can be in one of four microstates ( $i = 1, 2, 3$  or  $4$ ). Assuming that no two particles can be in the same state, calculate the number of microstates for the four-particle system if:

a) The particles are distinguishable, (e.g. labeled as  $A, B, C$ ).

b) The particles are indistinguishable, (e.g. labeled as  $A, A, A$ ).

c) Two particles are of one kind, one particle is of another, (e.g. labeled as  $A, A, B$ ).

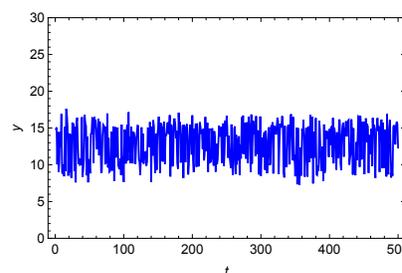
d) Generalize the result for  $N_A$  particles of type  $A$ ,  $N_B$  of type  $B$ , etc, and  $M$  states ( $M > N_A + N_B + \dots$ ).

**Problem 4.6.** a) A system consists  $N$  identical particles. Each particle can be in one of  $M$  states but no two particles can be in the same state, i.e., the occupancy of a state is at most 1. We represent the state of this  $N$ -particle system by the distribution of occupancy numbers,  $\mathbf{n} = \{n_1, n_2, \dots, n_M\}$  where  $n_i$  is the number of particles in state  $i$ . Clearly,  $n_i$  can only take two values, 1, if the state is occupied, or 0 if it is not. Calculate the probability of distribution  $\mathbf{n}$  assuming that all microstates are equally probable. Hint: consider the equivalent problem of  $N$  identical marbles placed into  $M$  distinguishable buckets so that no container has more than one marble; for example, if we may place the containers in a straight line, we can say that container #5, for example, is full (state 1) or empty (state 0).

b) Suppose now that states can have any occupancy number of particles, i.e.,  $n_i$  can have any value from 0 to  $M$ . Calculate the probability of distribution  $\mathbf{n}$  assuming that the state of each particle is independent of the state of any other particle.

d) Which occupancy rule is appropriate for a system of molecules?

**Problem 4.7.** Suppose we have 2 particles that can be placed on a line at distinct positions  $i = 1, 2, \dots, L$ . How many arrangements are possible, if the particles are indistinguishable? Generalize the result to  $N$  particles that may be placed in  $L$  distinct positions ( $L > N$ ). Discuss how these results are relevant to the arrangement of  $N$  particles in the phase space that consists of  $3N$  position coordinates and  $3N$  momentum coordinates.



**Figure 4.10:** Problem 4.3.

## 4.9. Problems

**Problem 4.8.** [Monte Carlo] A die has  $K$  faces that are labelled by consecutive integers,  $i = 1, 2, \dots$  ( $K$  is very large). We roll the die  $n$  times; if the sum is  $m$ , we record the three outcomes, otherwise we discard the results and try again. We wish to find the probability that a successful toss will consist of  $n_1$  ones,  $n_2$  twos, etc. We will do this by hand, and also by simulation. As an example we use  $n = 3$ ,  $m = 10$  and enumerate the possible outcomes. To do this we need to express number 10 as a sum of three positive integers, which can be done as follows:

$$\begin{array}{cccc} 8+1+1, & 7+2+1, & 6+3+1, & 6+2+2 \\ 5+4+1, & 5+3+2, & 4+4+2, & 4+3+3 \end{array}$$

These sets of numbers in all possible permutations constitute the ensemble of possible outcomes, or *phase space*. We will call an ordered sequence a *configuration* (for example, (8, 1, 1) is one such configuration). Each configuration appears in many permutations whose number is given by the multinomial coefficient,

$$\text{number of permutations of configuration} = \frac{n!}{n_1!n_2!\dots} \equiv \mathbf{n}!,$$

where  $n_i$  is the distribution of  $i$ 's in the toss, namely,  $n_1$  is the number of 1's,  $n_2$  is the number of 2's and so on. We call this the *multiplicity* of distribution and notate it as  $\mathbf{n}!$  (we use the vector notation  $\mathbf{n} = (n_1, n_2, \dots)$  as a compact way to refer to the distribution). We now construct the following table:

	tosses	$n_1$	$n_2$	$n_3$	$n_4$	$n_5$	$n_6$	$n_7$	$n_8$	$\mathbf{n}!$
8	1	1	2	0	0	0	0	0	1	3
7	2	1	1	0	0	0	0	1	0	6
6	3	1	0	1	0	0	1	0	0	6
6	2	2	0	2	0	0	1	0	0	3
5	4	1	1	0	0	1	1	0	0	6
5	3	2	0	1	1	0	1	0	0	6
4	4	2	0	1	0	2	0	0	0	3
4	3	3	0	0	2	1	0	0	0	3

Each row contains the outcome of the three tosses (ordered in descending order, not in the order observed during the toss), the corresponding distribution, and the multiplicity of that distribution, namely, the number of permutations with the same number of 1's, 2's, etc. The number of configurations is the sum of the multiplicity factors:

$$\text{number of configurations} \sum \mathbf{n}! = 36$$

and the total number of individual tosses is  $(3)(36) = 108$ . For a fair die, all configurations are equally probable. With this information we can calculate the statistical properties of any quantity in this ensemble. For example, the number of 1's in the ensemble is

$$\sum_{\mathbf{n}} n_1 \mathbf{n}! = (2)(3) + (1)(6) + (1)(6) + (0)(3) + (1)(4) + (0)(4) + (0)(3) + (0)(3) = 24.$$

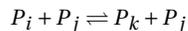
and the probability of 1's in the ensemble is

$$p_1 = \frac{24}{108} = 0.222.$$

- a) Construct the corresponding table for  $n = 4$ ,  $m = 10$ .
- b) What is the total number of ordered configurations?
- c) What is the probability that no number is repeated in four tosses?
- d) Which distribution of numbers is most probable in the ensemble?
- e) Which number is most probable in the ensemble?

A Mathematica notebook will be available for you to download that automates some of these calculations.

**Problem 4.9.** [Monte Carlo] A polymer undergoes the reversible reaction,



with equilibrium constant

$$K = \frac{w(m_k)w(m_l)}{w(m_i)w(m_j)}$$

where  $m_i$ ,  $m_j$ ,  $m_k$  and  $m_l$  are the molecular weights of the molecules in the reaction ( $m_i + m_j = m_k + m_l$ ) and  $w(m)$  is a function of the molecular weight.

a) Which thermodynamic property corresponds to  $w(m)$ ?

b) Write a Monte Carlo simulation using the Metropolis method to obtain samples of the equilibrium distribution if we know that the system contains  $M$  monomers and  $N$  chains. Construct the simulation as follows:

1. Start with a list of  $N$  chains all with the same molecular weight, equal to the average molecular weight  $M/N$ .
2. Pick two chains, and produce two new chains such that the mass of the two chains after the reaction is equal to the mass after the reaction.
3. Accept the results of the reaction with probability proportional to the equilibrium constant. To do this, draw a uniform random number  $\text{rnd}$  in the range  $(0, 1)$  and accept the result if  $\text{rnd} \leq K$ .
4. If the reaction is not accepted, pick two other chains and repeat the process. If the reaction is accepted, replace the two chains on the reactant side of the reaction with the two chains on the product side

By doing this enough times we will reach a point where the distribution of molecular weights fluctuates around the true equilibrium distribution.

c) Run the simulation with  $M = 1000$ ,  $N = 100$  and

$$w(m) = e^{-0.25m^2},$$

to obtain the equilibrium distribution of molecular weights. Monitor the standard deviation of the distribution to determine when equilibrium is reached (corresponds to the point that the standard deviation fluctuates around a stationary value). To obtain a good estimate of the distribution, use an average over several simulated distributions.

d) Repeat the previous part with  $w(m) = 1$ .

e) Compare the entropy of the molecular weight distributions in parts (d) and (e) and discuss the results.

A mathematica Notebook that performs the basic simulation will be available to download from the course web site.

## Statistical Thermodynamics

All macroscopic behavior of thermodynamic systems, be it vapor condensation, solubility or the equation of state, are manifestations of interactions between the microscopic constituents of matter: molecules, atoms, nuclei, electrons, ... . Luckily, the precise details of these interactions are not important at the moment, we will return to them in the next chapter when we consider specific models of matter. What is important is the fact that the internal structure of matter implies an internal state (*microstate*) whose specification requires a large, in fact *enormously* large, number of variables. To use the simplest example of a model for matter, consider  $N$  classical material points in motion within a box of volume  $V$ . The particles may interact with each other via a potential, and may also be subjected to external fields. The internal state of this mechanical system is fully specified by the positions and velocities of all  $N$  particles. As both position and velocity are three-dimensional vectors, the total number of degrees of freedom is  $6N$ : this is the number of independent variables that must be specified in order to fix the internal state. If we apply this model to matter,  $N$  is of the order of Avogadro's number; and if we consider that actual molecules are not point masses but have internal structure with additional degrees of freedom, the total number of degrees of freedom is indeed very large. This may seem to present an insurmountable obstacle but is actually a good thing! When the number of components that make up a system is very large, its overall behavior becomes independent of its details. In this limit *thermodynamics* emerges as the calculus of the *macroscopic observables* of the system. These observables are tied of course to the microscopic interactions through *statistical averages* over the possible internal states of the system. It is this averaging that allows us to handle the enormous microscopic detail in a manner that is both quantitative and computationally manageable. In this chapter we derive the calculus of *statistical thermodynamics*. By this we refer to the mathematical framework that connects microscopic degrees of freedom to macroscopic observables. We will recognize this calculus as identical to that of classical thermodynamics, with one important new element: all non-mechanical variables of the fundamental equation, namely, entropy, temperature and chemical potential, will be firmly linked to statistical properties of the internal state. To derive this calculus we will make no reference to the details of the internal state of matter, this connection will be made in the next chapter where we consider a specific model of matter. Indeed, the broad applicability of statistical thermodynamics and the mere

fact that the theory not only survived the quantum revolution, but seamlessly adapted to it, reflects the fact that thermodynamics is not tied to specific models of matter, or even to material systems. It is a general purpose calculus for stochastic systems whose observable behavior is the result of unobservable interactions between a very large number of degrees of freedom.

## 5.1 Microstates and ensembles

The macroscopic equilibrium state of a thermodynamic system with  $K$  components is fixed by a small number of variables,  $2 + \mathcal{N}$ , to be precise. This is the number of independent variables that appear in the fundamental differential. If these variables are fixed, the state of the system is fully specified and all of its properties have fixed values. The internal microstate of the system, however, is not fixed. Rather, it is in continuous flux as molecules move under the action of the laws of motion. To fix ideas we take the system to consist of material particles, in which case the microstate refers to the specification of all positions and all velocities. It is clear that a fixed macroscopic state corresponds to a very large number of microstates. The collection of all microstates that are possible under fixed macroscopic state constitute a thermodynamic ensemble. For the purposes of this discussion we treat microstates as discrete—we can always do this by discretizing the velocities and coordinates—that way we may distinguish microstates from each other and treat the ensemble as a discrete set. The details of the microstates in this ensemble depend on the macroscopic variables that are used to specify the microstate. Suppose we fix the total energy  $E$ , total volume  $V$  and the number  $N_i$  of particles of component  $i$ . This represents an isolated system, for example, a gas inside a rigid insulated box. All microstates in this ensemble have, by definition, the same energy. Suppose now that instead of energy we use temperature to specify the state, along with  $V$  and  $N_i$ . This represents a closed system in thermal equilibrium with the surroundings, for example, a gas inside a rigid box with conducting walls. The instantaneous energy of this system is not fixed as energy is exchanged between system and surroundings. Accordingly, this ensemble contains microstates with variable energy, unlike the  $(E, V, N_i)$  ensemble in which all microstates have the same energy. Many different ensembles can be constructed depending on the choice of the macroscopic state variables but three that are especially useful are listed below:

*Microcanonical ensemble* This is the ensemble of all microstates of an isolated system whose energy, volume and number of particles are fixed.

*Canonical ensemble* This is the set of all microstates of a system of fixed volume and number of particles in thermal equilibrium with its surroundings.

*Grand canonical ensemble* This is the ensemble of all microstates in an open system of fixed volume in thermal and chemical equilibrium with the surroundings.

The ensemble contains all microstates that can possibly exist within a system with fixed macroscopic state so that if we were to take a snapshot of the microscopic state of the system at some arbitrary time  $t$ , this snapshot would be

## 5.1. Microstates and ensembles

one of the microstates of the ensemble.<sup>1</sup> A physical system transitions between different microstates of the ensemble under the action of the laws of motion as their constituent particles move in space. It is impossible to predict with any certainty which microstate will be observed at any given time. Even if we adopt a fully deterministic view of matter as classical material points moving under deterministic laws, we would need to know the complete microstate –all positions and velocities– in order to integrate the equations of motion and predict the future state. Such level of knowledge is beyond reach. Thus we treat the microstate as a stochastic variable. Suppose that  $X_i$  is the value of mechanical property  $X$  in microstate  $i$ . A mechanical property is any property of the system that can be calculated from a *single microstate*, i.e., from knowledge of the positions and velocities of all particles. Two examples of mechanical properties are the total energy of the system (can be calculated as the sum of kinetic and potential energy), and pressure (can be calculated from the flux of the particles).<sup>2</sup> If we knew the probability of microstate  $i$  then we could calculate the ensemble average value of  $X$  by summing over all microstates,

$$\bar{X} = \sum p_i X_i.$$

The basic premise of statistical thermodynamics is that this ensemble average is the same as the *macroscopic* value of  $X$  that we would measure experimentally. The only piece of information we need in order to put this idea to work is the probability  $p_i$  of microstate  $i$ . The principle that allows us to determine this unknown probability is called the *maximum entropy principle* and is based on a new definition of entropy as a measure of uncertainty.

Suppose we are told that a black box spits out one of six possible numbers (say, 1 through 6) but are given no further information about its workings. We are asked to guess the probability distribution of its six outcomes. Is it even possible to answer this question? Of course it is! We can always take a *guess*. Here is one,

$$p_1 = 0.1, p_2 = 0.33, p_4 = 0.07, p_5 = 0.31, p_6 = 0.18,$$

and here is another one,

$$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6.$$

Most would agree that the first guess is too specific and projects a degree of confidence that is not supported by what we know about the system. On what basis should we assign such high probability to 2 at the expense of 4? The second guess feels more natural. It is far less specific, more difficult to predict its outcome,<sup>3</sup> but more in line with what we know about the system: it treats all outcomes as equivalent because nothing in the available information tells us otherwise. The uniform distribution in this example is the least “specific” among all distributions that could be used. It has more *uncertainty* than our first guess because we do not have any data (information) to remove this uncertainty.<sup>4</sup> *Uncertainty* should be understood in the context of predicting the outcome of a stochastic variable. The harder it is to predict, the more “uncertainty” there is. The maximum uncertainty is when everything that can happen takes place with equal probability.<sup>5</sup> If we can eliminate certain possibilities, or even decrease their probabilities, the overall uncertainty will decrease. But this should be done

<sup>1</sup>To capture this “snapshot” we would require a device that can record the position *and* velocities of all particles.

<sup>2</sup>Temperature, entropy and free energy are examples of *non* mechanical properties. As we will see, such properties require knowledge of all microstates and their probabilities.

<sup>3</sup>If we are had to gamble and could choose the distribution of the die, we would choose the first distribution. In this die, outcomes 2 has twice the probability of the same outcome with a fair die.

<sup>4</sup>In Information Theory, information is *knowledge*; more information is less uncertainty.

<sup>5</sup>Suppose you are debating whether or not to do elective surgery. If the probability of success is 95%, you might go for it. If it is 5% you would probably decide against it. In both cases there is high degree of certainty about the outcome and the decision is easier to make. It is when the probability of success and failure are about the same that it becomes difficult to make a decision.

only based on hard knowledge. If we eliminate possibilities without factual basis, we will be committing an error of judgment.

To put these ideas to work we need a measure of “uncertainty” that we can use to characterize a probability distribution  $p_i$ . Let  $S[p_i]$  be such measure. What properties should such measure have? First, it should be continuous in  $p_i$ . Second, if we are dealing with equally probable outcomes, the more outcomes there are the bigger the uncertainty should be. This is to say that the uncertainty of a fair die (six possible outcomes) ought to be higher than the uncertainty of a fair coin (two outcomes). And third, we require extensive behavior with respect to composite effects such that, if  $p_i$  is expressed via the conditional probability  $p_i = p_k p_{i|k}$ , the uncertainties of  $p_i$ ,  $p_k$  and  $p_{i|k}$  ought to be related as follows

$$S[p_i] = S[p_k] + \sum_k p_k S[p_{i|k}] \quad (5.1)$$

This property requires entropy to be extensive such that it can be decomposed into additive contributions of its constituent elements.<sup>6</sup> The mathematical function that satisfies these requirements turns out to be

$$S[p_i] = -k \sum p_i \log p_i, \quad (5.2)$$

where  $k$  is any positive constant. This expression is a *functional* of  $p_i$ , i.e., a function whose independent variable is not a single number but the entire probability distribution  $p_i$ . We call it *entropy* and we will take it to be the same quantity as the entropy of classical thermodynamic systems. This entropy in Eq. (5.2) has the dimensions of  $k$ , and if  $k$  is chosen to be 1 then entropy is dimensionless. For historical reason  $k$  is the Boltzmann constant,  $k_B = 1.23 \times 10^{-23}$  J/K.

Equation (5.2) gives the entropy of a normalized probability distribution, which in an intensive property. It can be extended to an extensive population, such as a collection of  $N$  outcomes taken from some distribution. Suppose that the population with  $N$  total members (outcomes) contains  $F_i$  members of type  $i$ . The entropy of this population is

$$S[F_i] = - \left( \sum_i F_i \right) \sum_i F_i \log \frac{F_i}{\left( \sum_i F_i \right)} = -N \sum_i F_i \log \frac{n_i}{N}. \quad (5.3)$$

This is the extensive entropy and reduces to Eq. (5.2) when  $\sum F_i = 1$ , i.e., the distribution is normalized.

The entropy defined in Eq. (5.3) has the following properties:

- It is never negative:  $S[F_i] \geq 0$ .
- It is zero only for one distribution:  $F_i = N \delta_{i,i^*}$ , where  $\delta_{i,j}$  is Kronecker's delta. This distribution represents a deterministic variable whose outcome is always  $i^*$ .
- It is homogeneous in  $F_i$  with degree 1:  $S[\lambda F_i] = \lambda S F_i$ .

<sup>6</sup>If outcomes  $i$  and  $i|k$  are independent, then  $S[p_{i|k}]$  is independent of  $k$  and comes out of the summation:

$$S[p_i] = S[p_k] + S[p_{i|k}] \sum_k p_k = S[p_k] + S[p_{i|k}],$$

which states that the joint uncertainty of two independent events is the sum of their uncertainties.

- It is concave, i.e., it satisfies the inequality

$$S[\alpha F_i + \alpha' F'_i] \geq \alpha S[F_i] + \alpha' S[F'_i]$$

for all distributions  $F_i, F'_i$ , and all  $\alpha, \alpha' \geq 0$ .

**Entropy of continuous distribution** Equation (5.3) generalizes to continuous distributions by simply converting the summations into integrals:

$$S[F] = -k \int F(x) \log \frac{F(x)}{M_0} dx, \quad x \geq 0, \quad (5.4)$$

where  $M_0$  is the zeroth order moment,

$$M_0 = \int F(x) dx. \quad (5.5)$$

For nonnegative  $x$  the continuous entropy in Eq. (5.4) has the same properties as its discrete form with the only difference that the distribution for which entropy is zero is the Dirac delta,  $F(x) = a\delta(x - x_0)$ , the continuous equivalent of a deterministic variable. Under a translation of  $x$  it is possible to further extend the definition of entropy to distributions whose argument is not restricted to be positive. Let  $z = ax - b$ . If  $F$  is the probability distribution of  $x$  and  $H$  is the distribution of  $z$ , the two distributions are related as

$$F(x) dx = H(z) dz. \quad (5.6)$$

We apply the definition of entropy in Eq. (5.4) to  $H$ :

$$S[H] = -M_0 \int H(z) \log \frac{H}{M_0} dz = -M_0 \int F(x) \log \left( \frac{F}{M_0} \frac{dx}{dz} \right) dx,$$

where  $M_0[H]$  is the zeroth moment of  $H$ , also equal to the zeroth moment of  $F$  by virtue of Eq. (5.6). The result is equivalent to

$$S[H] = S[F] + \log a.$$

The entropy of  $F$  is nonnegative but depending on the value of  $a$  this is not necessarily true for the entropy of  $H$ .

## 5.2 Microcanonical ensemble

Having identified entropy as a functional of the probability distribution we are ready to formulate the problem of statistical thermodynamics as follows: determine the unknown probability distribution by maximizing entropy subject to the available information about the system. This is the principle of maximum entropy. It is equivalent to that in macroscopic thermodynamics except that now we will be using the entropy functional to solve for the probability distribution of the equilibrium state.

We consider an isolated system of  $N$  particles in volume  $V$  with fixed total energy  $E$ . The total number of microstates available to this system is  $\Omega$  and clearly depends on  $E, V$  and  $N$ . We may further deduce that  $\Omega$  is an increasing function of all three variables because the number of microstates available to the system  $(E + \delta E, V + \delta V, N + \delta N)$  with all delta's positive must be *at least* as large as that for  $(E, V, N)$ . We seek the probability  $p_i$  to find the system in

microstate  $i$ . According to the principle of maximum entropy we must maximize the entropy functional,

$$\frac{S}{k} = - \sum_{i=1}^{\Omega} p_i \log p_i$$

under the sole constraint that  $p_i$  must be properly normalized:

$$\sum_{i=1}^{\Omega} p_i = 1. \quad (5.7)$$

Using the method of Lagrange multipliers, the function to be maximized is

$$- \sum_{i=1}^{\Omega} p_i \log p_i - \alpha \sum_{i=1}^{\Omega} (p_i - 1),$$

where  $\alpha$  is the Lagrange multiplier corresponding to the constraint in Eq. (5.7). We take the derivative of this function with respect to  $p_i$  and set it equal to zero:

$$-\log p_i - 1 - \alpha = 0 \Rightarrow \log p_i = -(1 - \alpha)$$

from which we conclude that  $p_i = 1/C$ , with  $C = \exp(-1 - \alpha)$  is the same for all microstates of the ensemble. From the normalization condition we obtain  $C = \Omega$ , which leads to the final result for the microcanonical probability

$$p_i = \frac{1}{\Omega}. \quad (5.8)$$

We refer to  $\Omega$  as the *microcanonical partition function*.

To establish connection with thermodynamics we insert this probability into the entropy functional:

$$\frac{S}{k} = - \sum_{i=1}^{\Omega} \left( \frac{1}{\Omega} \right) \log \left( \frac{1}{\Omega} \right),$$

which simplifies to

$$\frac{S}{k} = \log \Omega. \quad (5.9)$$

Thus we have identified the entropy of the microcanonical ensemble as the logarithm of the total number of microcanonical microstates.

### 5.2.1 Connection to thermodynamics

The partition function  $\Omega$  depends solely on  $E$ ,  $V$  and  $N$ : once these variables are fixed, the total number of microstates is also fixed. We express  $\log \Omega$  as a function of  $(E, V, N)$ ,

$$\log \Omega = \Omega(E, V, N), \quad (5.10)$$

and write its differential,

$$d \log \Omega = \left( \frac{\partial \log \Omega}{\partial E} \right)_{VN} dE + \left( \frac{\partial \log \Omega}{\partial V} \right)_{EN} dV + \left( \frac{\partial \log \Omega}{\partial N} \right)_{EV} dN. \quad (5.11)$$

Comparing this to the differential of entropy in classical thermodynamics we make the identifications,

$$\frac{dS}{k} = \frac{dE}{kT} + \frac{PdV}{kT} - \frac{\mu dN}{kT},$$

we make the identifications,

$$\frac{1}{kT} = \left( \frac{\partial \log \Omega}{\partial E} \right)_{VN}, \quad (5.12)$$

$$\frac{P}{kT} = \left( \frac{\partial \log \Omega}{\partial V} \right)_{EN}, \quad (5.13)$$

$$\frac{\mu}{kT} = - \left( \frac{\partial \log \Omega}{\partial N} \right)_{EV}. \quad (5.14)$$

With Eqs. (5.9–5.14) we have established something very important: a direct connection between macroscopic and microscopic properties. The macroscopic properties that appear in these equations are  $T$ ,  $P$ ,  $\mu$ ,  $E$ ,  $V$  and  $N$ . The microscopic property that ties them together is  $\log \Omega$ , the logarithm of the total number of available microstates. These equations allow us to obtain temperature, pressure and chemical potential from the derivatives of the microcanonical partition function. These results are very important: they provide a direct connection between macroscopic properties and the microscopic world that gives rise to them.

### 5.2.2 Second law in the microcanonical ensemble

The association of entropy with  $\log \Omega$  rather than  $\Omega$  itself is a reassuring result. While  $\Omega$  is astronomically large, its logarithm is of reasonable magnitude.<sup>7</sup> Let us explore some implications of this relationship. We start with two isolated systems in macroscopic states  $(E_1, V_1, N_1)$  and  $(E_2, V_2, N_2)$ , respectively. The number of microstates in system 1 is  $\Omega_1 = \Omega(E_1, V_1, N_1)$ , and in system 2 is  $\Omega_2 = \Omega(E_2, V_2, N_2)$ . The probability to find system 1 in microstate  $i_1$  is  $p_1 = 1/\Omega_1$ ; similarly, the probability to find system 2 in microstate  $i_2$  is  $p_2 = 1/\Omega_2$ . Consider now the ensemble of microstates that can be constructed by combining every microstate of system 1 with every microstate of system 2. There is a total number of  $\Omega_1\Omega_2$  microstates and all of them have the same probability

$$p_{i_1, i_2} = p_{i_1} p_{i_2} = \frac{1}{\Omega_1 \Omega_2}$$

All of these microstates have total energy  $E_1 + E_2$ , volume  $V_1 + V_2$ , and number of particles  $N_1 + N_2$ . They belong, that is, to the microcanonical ensemble  $(E_1 + E_2, V_1 + V_2, N_1 + N_2)$ . The combined system therefore must have at least  $\Omega_1\Omega_2$  microstates, and possibly more –but certainly no fewer. This implies that

$$\Omega_{1+2} > \Omega_1 \Omega_2. \quad (5.15)$$

The probability of microstate in the large microcanonical ensemble is

$$p^* = \frac{1}{\Omega_{1+2}}$$

<sup>7</sup>To appreciate the vast difference between a huge number and its logarithm, take  $10^{23}$ , which is of the order of Avogadro's number, and express it in seconds. The resulting number is more than 200,000 times the age of the universe! But its log expressed in the same units, 52.96 s, is not even a minute. (The age of the universe is estimated to be about 14 billion years.)

## 5.2. Microcanonical ensemble

and since  $\Omega_{1+2}$  is larger than, or equal to  $\Omega_1\Omega_2$ , we conclude that the equilibrium distribution maximizes  $\log\Omega$  with respect to any partitioning of the total energy, volume or number of particles.

Figure 5.1 demonstrates this inequality with a discrete example. System 1 consists of two buckets containing one ball; system 2 consists of three buckets with two balls. Boxes are distinguishable, balls are not. At most one ball can be placed in a bucket. A microstate refers to an arrangement of the balls into buckets. With reference to system 2, for example, microstate (1, 0, 1) means that the first and third bucket contain a ball but the second bucket is empty. By straightforward combinatorial counting we obtain

$$\Omega_1 = \frac{2!}{1!1!} = 2, \quad \Omega_2 = \frac{3!}{2!1!} = 3.$$

If we consider the single system that is formed with 5 buckets and 3 balls the number of possible microstates is

$$\Omega_{1+2} = \frac{5!}{3!2!} = 10.$$

This is larger than the product of  $\Omega_1\Omega_2 = 6$  and the fundamental inequality in Eq. (5.16) is shown to hold. Inspection of the microstates explains what is going on. The 6 microstates that are formed by combining the microstates of the two parts are constrained by the fact that the first two buckets may share one ball, and the other three buckets may share two balls. In the combined system this restriction is not present. We may place more balls in each part and thus form additional microstates.

Taking the logarithm of Eq. (5.15) we have

$$\log\Omega(E_1 + E_2, V_1, V_2, N_1 + N_2) > \log\Omega(E_1, V_1, N_1) + \log\Omega(E_2, V_2, N_2).$$

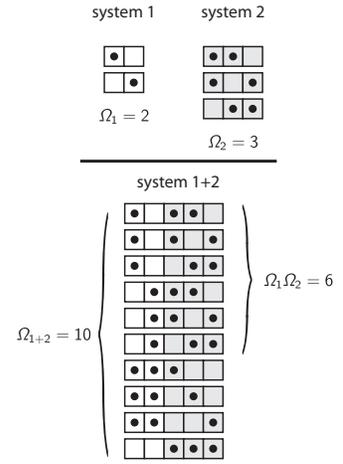
While the  $\Omega$ 's themselves satisfy Eq. (5.15) as a strict *inequality*, the corresponding inequality for the logarithms can, under certain conditions, revert asymptotically to an equality. Thus we have

$$\log\Omega(E_1 + E_2, V_1 + V_2, N_1 + N_2) \geq \log\Omega(E_1, V_1, N_1) + \log\Omega(E_2, V_2, N_2).$$

and in more condensed form,

$$\log\Omega_{1+2} \geq \log\Omega_1 + \log\Omega_2. \quad (5.16)$$

This condition expresses the fundamental inequality of the microcanonical ensemble and is a statement of the second law: the entropy of two non-interacting systems is lower than the entropy of the single system that is formed when the two systems are allowed to interact. Systems 1 and 2 are in *equilibrium* with each other when this condition becomes an equality. This establishes the variational property of  $\log\Omega$ : it is maximized with respect to any equilibrium partitioning of the whole. This provides us with the first strong argument in support of the notion that the entropy functional in Eq. (5.2) and the entropy of classical thermodynamics represent the same fundamental property.



**Figure 5.1:** The system that is formed by combining systems 1 and 2 has a total of 10 microstates. Only 6 of them can be produced by combining the microstates of the two subsystems. The product  $\Omega_1\Omega_2$  is the number of microstates of two *non-interacting* systems taken as one;  $\Omega_{1+2}$  is the number of microstates when the two systems are allowed to interact (in this case, exchange particles).

### 5.2.3 Quasistatic process

In the language of statistical thermodynamics a process is a continuous transformation of the probability distribution  $p_i$ . A *quasistatic* process is one in which  $p_i$  is always at a maximum at every step of the way. Under these conditions thermodynamic entropy ( $S/k$ ), and the probability functional in Eq. (5.2) are the same:

$$S = -\sum_i p_i \log p_i. \quad (5.17)$$

We now subject a system with fixed number of particles to a quasistatic process. The mean energy is

$$\bar{E} = \sum_i p_i E_i \quad (5.18)$$

and its change along the process is

$$d\bar{E} = \sum_i p_i dE_i + \sum_i E_i dp_i \quad (5.19)$$

and expresses the first law in terms of the probability of microstate. The first summation on the right-hand side is the change in energy at fixed  $p_i$  but since fixed  $p_i$  means “fixed  $S$ ,” according to Eq. (5.17), this sum is the energy change for isentropic process and is equal to the reversible work:<sup>8</sup>

$$\sum_i p_i dE_i = dW_{\text{rev}}. \quad (5.20)$$

Then the second sum in Eq. (5.19) must be heat:

$$\sum_i E_i dp_i = dQ_{\text{rev}}. \quad (5.21)$$

Reversible work is associated with energy stored in all microstates without altering their probabilities; reversible heat is associated with altered probabilities of microstates. Since the “microstate” is an entity with fixed energy, we must be careful in interpreting the meaning of  $dE_i$  in these equations. A process transforms the microstates of the initial state to those of the final state. In a reversible process every state  $i$  is transformed to a state  $i'$  such that there is a one-to-one correspondence between every microstate in the initial and final state, schematically

$$i \xleftrightarrow{\text{reversible process}} i'.$$

(see Fig. 5.2). This mapping is implied in Eq. (5.19) where we have tacitly assumed that each term of the summation that defines  $\bar{E}$  contributes a corresponding term to the differential  $d\bar{E}$ . We should understand then  $i$  to mean “microstate  $i$  initial state, or microstate  $i'$  in final state” and  $dE_i$  or  $dp_i$  to refer to the difference in energy or probability between these two states. In this picture isentropic work represents a change in all  $E_i$  at fixed probabilities, and constant-volume heating or cooling represents a change in the probabilities at fixed energy.

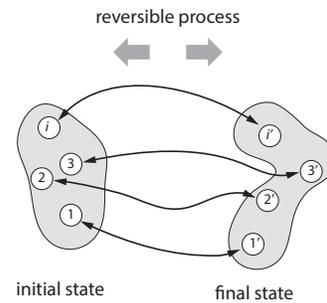
The one-to-one mapping between the initial and final states means that if we reverse the direction of the external inputs that drive the process the entire process will be reversed as each microstate individually reverts to its initial state.

<sup>8</sup>Recall that at fixed  $N$ ,

$$dU = TdS - PdV,$$

then with  $S = \text{const.}$

$$dU = -PdV = dW_{\text{rev}}.$$



**Figure 5.2:** In reversible process there is a one-to-one mapping between all microstates of the initial and final state: the volume of phase space is preserved.

**Example 5.1.** *A discrete microcanonical ensemble*

We are not yet equipped to write realistic models for molecular systems but we may still perform *exact* calculations for some very simple models. Here we consider a *lattice* model. We have a three dimensional cubic lattice of length  $l$  that contains  $l \times l \times l = V$  identical sites. On this lattice we place  $N$  identical particles in such way that a lattice site may contain no more that one particle. Each particle receives energy in discrete integer units  $\{1, 2, 3, \dots\}$  and the total energy of the lattice is given by the sum of all particle energies,

$$E = \sum_{i=1}^N E_i,$$

We wish to obtain the microcanonical partition function  $\Omega(E, V, N)$ , where  $V$  is dimensionless volume,  $E$  is dimensionless energy and  $N$  particles.

**Solution** All particles belong to a single component, i.e., they should be treated as indistinguishable from each other. The lattice sites on the other hand should be treated as distinguishable because they can be identified by their  $(x_i, y_i, z_i)$  coordinates. The problem then is to calculate the total number of microstates with fixed energy  $E$ , volume  $V$  and number of particles  $N$ .

A microstate in this system is a distinct arrangement of particles on the lattice with energies such that their sum is  $E$ . In this problem the assignment of positions on the lattice and the assignment of energies are independent of each other; accordingly, the total number of arrangements is the product of the number of arrangements in space (number of configurations),  $\Omega_c$ , and the number of ways to distribute the energy,  $\Omega_E$ :

$$\Omega(E, V, N) = \Omega_c(V, N) \cdot \Omega_E(E, N)$$

where  $\Omega_c(V, N)$  is the number of arrangements (configurations) in space and  $\Omega_E(E, N)$  is the number of arrangements with respect to energy. The number of arrangements in space is equal to the total number we can choose  $N$  lattice sites to occupy out  $V$  sites total:

$$\Omega_c(V, N) = \binom{V}{N}.$$

With respect to energy, we seek the number of ways to assign integer energies to  $N$  particles in such way that the total energy is  $E$ . This is equal to the number of ways to split integer  $E$  into  $N$  non negative integers. In number theory this is called a partition of  $E$  into  $N$  parts and the number of ways to accomplish this partitioning is<sup>9</sup>

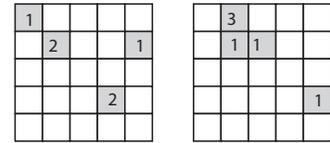
$$\Omega_E(E, N) = \binom{E-1}{N-1}.$$

Combining these results we finally obtain the microcanonical partition function of this system in the form,

$$\Omega(E, V, N) = \binom{V}{N} \binom{E-1}{N-1}. \quad (5.22)$$

The quantity that matters in thermodynamics is the logarithm of the microcanonical partition function. Using the Stirling approximation  $\log x! = x \log x - x$  and the additional approximations,  $(N-1) \log(N-1) \approx N \log N$ ,  $(E-1) \log(E-1) \approx E \log E$ , both of which are appropriate when  $N$  and  $E$  are large, the log of the microcanonical partition function becomes

$$\log \Omega = N \left\{ \log \frac{VE}{N^2} - \left( \frac{V}{N} - 1 \right) \log \left( \frac{V}{N} - 1 \right) - \left( \frac{E}{N} - 1 \right) \log \left( \frac{E}{N} - 1 \right) \right\}. \quad (5.23)$$



**Figure 5.3:** Two microcanonical microstates on a two-dimensional lattice with  $V = 25$ . Both microstates have  $N = 6$  particles and total energy  $E = 6$ .

<sup>9</sup>This result was derived in chapter 4; see Eq. (4.29).

### 5.3. Canonical ensemble

Notice that  $E/N$  is the energy per particle and  $V/N$  the volume per particle, both intensive properties. The above result then states that  $\log\Omega$  is an extensive property. Indeed this is what we expect of entropy. All properties of the system may now be calculated.

**Equation of state** The pressure of the system is given by the derivative of  $\log\Omega$  with respect to  $V$ :

$$\frac{P}{kT} = \left( \frac{\partial \log\Omega}{\partial V} \right)_{E,N} = \left( \frac{\partial \log\Omega}{\partial V} \right)_{E,N} = \log \frac{V}{V-N}.$$

Let us obtain the asymptotic form of this equation in the limit of a very dilute system,  $V/N \gg 1$ . To do this we set  $\nu = V/N$  and write the equation of state in the form<sup>10</sup>

$$\frac{P}{kT} = -\log(1 - \nu) \rightarrow \nu = \frac{V}{N},$$

<sup>10</sup>Recall that  $\log(1+x) \rightarrow x$  when  $x \rightarrow 0$ .

or

$$PV = NkT. \quad (5.24)$$

This is the ideal-gas law! The result is quite remarkable if we consider that we have recovered the correct equation of state of a dilute gas using a rather crude physical model.

**Lessons from a simple lattice model** The lattice model is a simple system on which we can perform exact calculations. It lacks in physical realism, so we cannot take it as a reliable model of matter, but it serves as an example on which to apply the theory of the microcanonical ensemble. Despite its remarkable simplicity, it is more general than it looks. The geometry of the lattice does not matter – we specified it as cubic, but we did not make use of that piece of information. Our results are general for any uniform lattice, namely one whose sites are all equivalent. Similarly, we did not use the fact that the lattice was in the shape of a three dimensional cube. It may have any shape (square, rectangular, other) and any number of dimensions (it could be a line, a two dimensional surface, or a solid of higher dimensionality). It could serve, for example, as a (crude) model of adsorption, if we take the lattice to be two dimensional and the energy to refer to the interaction energy with the substrate. The main limitation of this model is that it does not include interactions between particles and for this reason it cannot model phase transitions. We will add interactions to study phase transitions in chapter 8. Our focus here is in the development of the theory of statistical mechanics and our examples will be meant to provide demonstrations of the theory, not necessarily physical realism.

### 5.3 Canonical ensemble

The microcanonical ensemble represents the microstates of an isolated system. Suppose now that the system is in thermal equilibrium with the surroundings. The system is closed and has fixed volume, i.e.,  $V$  and  $N$  remain fixed, but because the system exchanges energy with the surroundings, its energy is not fixed but varies around its mean energy  $\bar{E}$ . For a system in equilibrium,  $\bar{E}$  is constant and a property of the equilibrium state. The microstates of this system are all microstates that can be constructed with  $N$  particles in volume  $V$  and with *any* energy. There is usually a limit on the *smallest* energy the system can have (ground state); there is no limit on the maximum energy.<sup>11</sup> If we operate at energies much higher than that of the ground state the details of the ground state are not very important. Then we can take the range of energies to be from  $-\infty$  to  $+\infty$  without loss of generality. The canonical ensemble consists of all microstates that can be constructed with fixed  $V$  and  $N$  and any possible

<sup>11</sup>In a collection of molecules interacting with an intermolecular potential, the minimum energy corresponds to a microstates with zero kinetic energy (kinetic energy cannot be negative) with the particles arranged in such way as to produce the minimum potential energy. This minimum energy may be positive or negative.

energy. We now want to determine the probability of microstates over this infinite phase space under the condition that the mean energy is fixed. Of the infinite number of probability distributions with mean  $\bar{E}$ , we choose the distribution that maximizes entropy. This amounts to choosing the distribution with the maximum uncertainty given that the mean of the distribution is known. Mathematically we see to maximize

$$\max_{p_i} \left\{ -\sum p_i \log p_i \right\}$$

with respect to the unknown probabilities  $p_i$  subject to the constraints,

$$\sum_i p_i = 1, \quad (5.25)$$

$$\sum_i E_i p_i = \bar{E}. \quad (5.26)$$

The first constraint is the standard normalization condition; the second constraint expresses the fact that the mean energy is fixed. Both summations are over all microstates  $i$ , and  $E_i$  refers to the energy of microstate  $i$ . Using Lagrange multipliers the function to maximize is

$$\max_{p_i, \alpha, \beta} \left\{ -\sum p_i \log p_i - \alpha (\sum p_i - 1) - \beta (\sum E_i p_i - \bar{E}) \right\}$$

where  $\alpha$  and  $\beta$  are the Lagrange multipliers for the two constraints in Eqs. (5.25) and (5.26). Setting the derivative with respect to  $p_i$  equal to zero we obtain

$$-\log p_i - 1 - \alpha - \beta E_i = 0.$$

Solving for  $p_i$  we obtain the probability of canonical microstate,

$$p_i = \frac{e^{-\beta E_i}}{Q}, \quad (5.27)$$

where  $Q = \exp(-1 - \alpha)$ . The canonical probability depends on a single property of the microstate, its energy. More specifically, the probability of microstate  $i$  is weighted by the factor  $\exp(-\beta E_i)$ , known as the Boltzmann factor. This means that between two microstates, the one with the lowest energy is more probable. It also means that between two microstates with the same energy, both are equally probable. The normalization constant  $Q$  in Eq. (5.27) is called the *canonical partition function* and is the major property of the canonical ensemble. Inserting the canonical probability into the first constraint we obtain the following expression for the partition function,

$$Q = \sum e^{-\beta E_i}, \quad (5.28)$$

with the summation going over all canonical microstates. The canonical partition function is a function of  $\beta$ , which appears in the exponent of the summand, but not of  $E_i$ , since we sum over all microstates. It is also a function of volume and number of particles, since both variables determine the actual microstates over which the above summation is conducted.<sup>12</sup> The canonical partition func-

tion is therefor a mathematical function of  $\beta$ ,  $V$  and  $N$ :

$$Q = Q(\beta, V, N).$$

We now insert the canonical probability into the second constraint and write the result in the form,

$$\frac{1}{Q} \sum_i E_i e^{-\beta E_i} = \bar{E}. \quad (5.29)$$

This equation can be expressed in the alternative form

$$\bar{E} = \frac{\sum_i E_i e^{-\beta E_i}}{\sum_i e^{-\beta E_i}} = \frac{-1}{\sum_i e^{-\beta E_i}} \left( \frac{\partial}{\partial \beta} \sum_i e^{-\beta E_i} \right),$$

and finally as

$$\bar{E} = - \left( \frac{\partial \log Q}{\partial \beta} \right)_{VN}. \quad (5.30)$$

Thus we have the mean energy of the ensemble expressed as a derivative of the canonical partition function. As we will see, all properties of the canonical ensemble can be expressed in terms of the canonical partition function and its derivatives.

### 5.3.1 Entropy of the canonical ensemble

We obtain the entropy of the ensemble by inserting the canonical probability into the entropy functional:

$$\frac{S}{k} = - \sum_i p_i \log p_i = - \sum_i p_i (-\beta E_i - \log Q)$$

or

$$\frac{S}{k} = \beta \bar{E} + \log Q. \quad (5.31)$$

Using Eq. (5.30), this result can written as

$$\frac{S}{k} = \log Q - \beta \left( \frac{\partial \log Q}{\partial \beta} \right)_{VN}. \quad (5.32)$$

This result identifies entropy as the Legendre transform of the canonical partition function with respect to  $\beta$ . Conversely,  $\log Q$  is a Legendre transformation of entropy – but with respect to which variable? To answer this question we rewrite Eq. (5.31) in the form

$$\log Q = \frac{S}{k} - \beta \bar{E}.$$

For the right-hand side to be a Legendre of  $S$ ,  $\beta$  must be the derivative of entropy with respect to  $E$ <sup>13</sup>

$$\beta = \frac{1}{kT}. \quad (5.33)$$

<sup>13</sup>Recall that the differential of entropy is

$$\frac{dS}{k} = \frac{dE}{kT} + \frac{PdV}{kT} - \frac{\mu dN}{kT}.$$

Then the canonical partition function is

$$\log Q = \frac{S}{k} - \frac{E}{kT} = \frac{TS - E}{kT},$$

or

$$\log Q = -\frac{A}{kT}, \quad (5.34)$$

where  $A$  is the Helmholtz free energy. To complete the thermodynamic connection we write the differential of  $\log Q$  by Legendre-transforming the differential of  $S$  with respect to energy:

$$d \log Q = -\bar{E} d\left(\frac{1}{kT}\right) + \frac{PdV}{kT} - \frac{\mu dN}{kT} = -\bar{E} d\beta + \beta PdV - \beta \mu dN.$$

where we have used  $\beta$  as a shortcut for  $1/kT$ . From this differential we obtain the following equations:

$$\bar{E} = -\left(\frac{\partial \log Q}{\partial \beta}\right)_{VN} \quad (5.30)$$

$$\frac{P}{kT} = \left(\frac{\partial \log Q}{\partial V}\right)_{\beta N} \quad (5.35)$$

$$\frac{\mu}{kT} = -\left(\frac{\partial \log Q}{\partial N}\right)_{\beta N} \quad (5.36)$$

the first of which was already obtained earlier as Eq. (5.30). These equations allow us to calculate all properties from knowledge of the canonical partition function expressed as a function of  $\beta$ ,  $V$  and  $N$ .

### 5.3.2 Relationship to the microcanonical ensemble

The canonical and the microcanonical ensembles appear quite different from each other. They consist of different sets of microstates and these microstates are characterized by different probability distributions. Despite these differences they are thermodynamically equivalent and produce identical results for all thermodynamic properties. Here we will establish a connection between the partition functions in the two ensembles. The canonical ensemble consists of all microstates that can be constructed by  $N$  particles in volume  $V$  having *any* possible energy. By contrast the microcanonical ensemble strictly contains microstates with a single value of energy. If we collect all canonical microstates with fixed energy, say  $E'$ , we will obtain all microstates of the microcanonical ensemble  $(E', V, N)$ . We may view then the canonical ensemble as a collection of an infinite number of microcanonical ensembles that cover all possible energies. The name “microcanonical” was coined by Gibbs to indicate that it represents an infinitesimal piece of the canonical.<sup>14</sup> Recognizing that the canonical microstates can be grouped by energy into a series of microcanonical ensembles, we apply this grouping to construct the summation over microstates in Eq. (5.28). There are  $\Omega(E, V, N)$  microstates with energy  $E$  and they all contribute  $\Omega(E, V, N) \exp(-\beta E)$  to the summation. The canonical partition function can be

<sup>14</sup>Canonical carries the connotation of *proper, normal, or follows the rule*, and conveys a preferred or default representation. Gibbs developed the canonical ensemble first. He then obtained the *micro* canonical ensemble as an infinitesimal slice of the canonical ensemble at fixed energy, and the *grand* canonical by combining an infinite number of canonical ensembles, each with a different number of particles. Viewed that way it makes sense to consider the canonical ensemble as the “default” from which to construct the other ensemble. But we can also take the opposite view and consider the microcanonical ensemble to consist of a collection canonical microstates by partitioning the system in subsystems of fixed  $V$  and  $N$  but each with each own energy. The canonical ensemble, however, is more convenient to work with and forms the basis for most theoretical analyses.

expressed then in the form

$$Q(\beta, V, N) = \sum_E \Omega(E, V, N) e^{-\beta E}. \quad (5.37)$$

Here the summation goes over energy, not microstates, and its limits are, in principle, from  $-\infty$  to  $\infty$ . Real systems usually have a limit on the lowest energy they can possess (ground state) and in this case the lower limit of the summation must be set to that energy. For a system of classical particles, this would be a state of zero kinetic energy with the particles arranged in fixed volume so as to have the smallest possible potential energy.<sup>15</sup> There is no limit as to the highest energy that is possible (particles can have unlimited kinetic energy). In practice the summand in Eq. (5.37) typically peaks within some finite range of energies and is zero outside that range. This allows us to take the limits to be from  $-\infty$  to  $\infty$  without causing mathematical or physical error.

Suppose now that we have  $K$  systems with identical volume  $V$  and number of particles  $N$  in thermal contact with each other such that the total energy of the combined system is fixed. The combined system has fixed total energy  $E'$ , fixed volume  $V' = KV$  and fixed number of particles  $N' = KN$ . Its microstates then come from the microcanonical ensemble  $(E'V'N')$ . A microstate of this combined system consists of a sequence of canonical microstates, such that system 1 is in microstate  $i_1$ , system 2 in microstate  $i_2$  and so on. The probability to find system  $i$  in microstate  $i_k$  is given by the canonical probability

$$p_{i_k} = \frac{e^{-\beta E_{i_k}}}{Q},$$

where  $Q = Q(\beta, V, N)$  is the canonical partition function of the canonical systems. The probability that system 1 is at microstate  $i_1$  and system 2 is at microstate  $i_2$  and system 3  $\dots$  and so on is the product of these probabilities,

$$p(i_1, i_2, \dots) = \frac{e^{-\beta E_{i_1}}}{Q} \frac{e^{-\beta E_{i_2}}}{Q} \dots = \frac{e^{-\beta(E_{i_1} + E_{i_2} + \dots)}}{Q^K} = \frac{e^{-\beta E'}}{Q^K}.$$

This is the probability to find each canonical system at the specified microstates. Since the sequence of these microstates forms a microcanonical microstate for the combined system, this probability is also equal to the microcanonical probability<sup>16</sup>

$$p(i_1, i_2, \dots) = \frac{1}{\Omega(E', V', N')}.$$

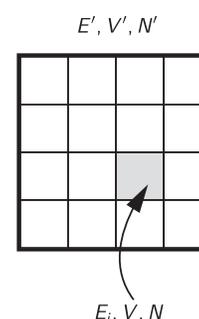
Equating the two results and taking the log we obtain

$$\frac{\log \Omega(E', V', N')}{K} = \log Q(\beta, V, N) + \frac{\beta E'}{K}.$$

By the homogeneous property of  $\log \Omega$ , the left-hand side is the log of the microcanonical partition function at  $E'/K = \bar{E}$ ,  $V'/K = V$ ,  $N'/K = N$ .<sup>17</sup> Therefore,

$$\log \Omega(\bar{E}, V, N) = \log Q(\beta, V, N) + \beta \bar{E}. \quad (5.38)$$

<sup>15</sup>This presumes infinite repulsion at zero distance between particles. A system of particles with infinite attraction at zero distance is not stable at any temperature and does not represent a physical system.



**Figure 5.4:**  $K$  canonical systems with the same  $V$  and  $N$  in thermal contact with each other.

<sup>16</sup>Strictly speaking, the microcanonical probability ought to be corrected for the fact that the total number of microstates that can be constructed by combining systems via conducting walls is *smaller* than the total number of microstates that are possible when these walls are not present. This was demonstrated in Fig. 5.1. To correct the microcanonical probability we should write

$$p(i_1, i_2, \dots) = \frac{1}{\Omega(E'V'N') - \delta\Omega}$$

where  $\delta\Omega$  is the number of the additional microstates that become available to the microcanonical system when the internal walls are removed. It turns out when we take the log of the denominator, we find

$$\log(\Omega - \delta\Omega) \rightarrow \log \Omega - O(\log N).$$

Since  $\log \Omega \sim N$ , this correction vanishes when we work in log space and this is why we do not include it in our analysis.

<sup>17</sup>Since  $\log \Omega$  is extensive (homogeneous with degree 1) in  $E, V, N$ , we have

$$\frac{\log \Omega(E', V', N')}{K} = \log \Omega(E'/K, V'/K, N'/K).$$

In Eq. (5.31) we determined that  $\log Q + \beta \bar{E}$  is the entropy of the canonical ensemble. Equation (5.38) then says that the entropy of the canonical ensemble is equal to the entropy of the *microcanonical* ensemble whose energy is equal to the average energy in the canonical. This means that every canonical ensemble is associated with a microcanonical ensemble that has the same entropy. If we match the energy of the microcanonical ensemble to the mean energy of the canonical ensemble, both ensembles will produce *identical* thermodynamic properties.

**On the notation  $E$  versus  $\bar{E}$**  The energy of canonical microstates is not fixed, therefore when we speak of the “energy of the canonical ensemble” we refer to the mean energy  $\bar{E}$ . This is the property that we identify as the *internal energy* of a system in thermal equilibrium with its surroundings. In the microcanonical ensemble energy is strictly constant and we notate it as  $E$  (with no bar). This is the property we identify as *internal energy* in an isolated system. Both  $E$  and  $\bar{E}$  refer to the *same* physical property, the macroscopic energy of the system. The different notation is meant to convey information about the nature of the corresponding microstates. In the canonical ensemble we use  $E$  for the energy of microstates (which can be any number from  $-\infty$  to  $\infty$ ) so we use special notation to refer to the the mean energy of the ensemble. In the microcanonical ensemble this distinction is not necessary.

**Example 5.2.** *A lattice calculation of the canonical partition function* \_\_\_\_\_  
As a demonstration of the theory of the canonical ensemble we will calculate the canonical partition of the lattice system we discussed in Example 5.1. As we recall, in this model particles are placed on discrete points of a lattice with size  $V$ , particles receive energies that are positive integer numbers, and the total energy is the sum of the particle energies,

$$E = \sum_{j=1}^N E_j$$

with  $E_j$  takes values from the integer set  $\{1, 2, 3, \dots\}$ . (To avoid confusion between summations over particles and summations over microstates, we use  $j$  to indicate summations over the  $N$  particles of the system and  $i$  for summations over microstates.) We will obtain the canonical partition function in the limit  $V/N \gg 1$ , and  $E/N \gg 1$ .

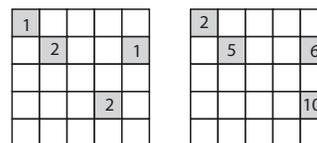
**Solution** We will work out the solution completely in terms of the dimensionless energy  $E$  and the dimensionless volume  $V$ .

The canonical partition function is given by the summation<sup>18</sup>

$$Q(\beta, V, N) = \sum_{i=1}^{\infty} e^{-\beta E_i} \quad (5.39)$$

which goes over all canonical microstates with fixed  $V$  (or  $V$ ) and  $N$ . We will organize this summation as follows. First we fix energy to some value  $E$  and assign energies to the  $N$  particles so that their sum is equal to the chosen value  $E$ . The number of ways to do this was determined in Example (5.1) and was found to be

$$\Omega_e = \binom{E-1}{N-1}$$



**Figure 5.5:** Two canonical microstates on a lattice with  $V = 25$  and  $N = 4$ . One microstate has energy  $E = 6$  (left) and the other  $E = 23$ .

<sup>18</sup>The upper limit in the summation is the total number of canonical microstates. This number is infinite because the system can have any possible energy between  $N$  (all particles at ground state) and infinity.

### 5.3. Canonical ensemble

Each of these energy arrangements can be placed in  $\Omega_c$  number of ways on the lattice; this number is

$$\Omega_c = \binom{V}{N}.$$

And since they have the same energy, they all contribute to the summation in Eq. (5.39) the same factor,

$$\binom{V}{N} \binom{E-1}{N-1} e^{-\beta E}.$$

The canonical partition function is the sum of these terms over all possible energies we can assign to the lattice:

$$Q(\beta, V, N) = \sum_{E=N}^{\infty} \binom{V}{N} \binom{E-1}{N-1} e^{-\beta E}. \quad (5.40)$$

To obtain thermodynamic relationships from this partition function we need to express it into a simpler mathematical function of  $\beta$ ,  $V$  and  $N$  and this requires the evaluation of the summation in the limit that  $E$ ,  $V$  and  $N$  are large numbers. In addition, we also assume take  $V/N \gg 1$  and  $E/N \gg 1$ . This additional condition is not necessary, we apply it here for mathematical convenience.

Our general approach to accomplish this is as follows. First, we define the auxiliary function  $F(E)$  to be the summand in Eq. (5.40),

$$F(E) = \Omega(E, V, N) e^{-\beta E} = \binom{V}{N} \binom{E-1}{N-1} e^{-\beta E}. \quad (5.41)$$

We will obtain the asymptotic representation of  $F(E)$  in the form of a continuous function of energy and will approximate the summation by the integral

$$\sum_{E=N}^{\infty} \Omega(E, V, N) e^{-\beta E} \rightarrow \int_N^{\infty} F(E) dE.$$

The combinatorial term in  $F$  increases with energy while the exponential term decreases. As a result,  $F(E)$  has a maximum in  $E$ . This maximum is actually quite sharp, as shown in Fig. 5.6. For large  $E$  and  $N$ ,  $F$  approaches a symmetric Gaussian function. To obtain this asymptotic Gaussian we expand  $\log F$  in powers of  $E$  around the energy  $E_*$  that maximizes  $\log F$  and truncate the expansion past the second order term:

$$\log F = \log F_* + \frac{1}{2} \frac{d^2 \log F_*}{dE^2} (E - E_*)^2 + \dots$$

Inverting the logarithm we obtain  $F$  in the form

$$F(E) = F_* \exp \left[ -\frac{(E - E_*)^2}{2\sigma_E^2} \right] \quad (5.42)$$

with

$$\sigma_E^2 = -1 \left/ \frac{d^2 \log F_*}{dE^2} \right.$$

Since  $\log F$  has a maximum, its second derivative at the maximum is negative, and this guarantees that  $\sigma_E^2$  is positive. To obtain  $E_*$  and  $\sigma_E$  we evaluate the first and second derivatives of  $\log F$ .

### 5.3. Canonical ensemble

To evaluate  $E_*$  and  $\sigma_E$  we need the derivatives of  $\log F$ . First we write  $\log F = \log \Omega - \beta E$  and use Eq. (5.23) to obtain  $\log F$  in the form

$$\log F = -\beta E + N \left\{ \log \frac{VE}{N^2} - \left( \frac{V}{N} - 1 \right) \log \left( \frac{V}{N} - 1 \right) - \left( \frac{E}{N} - 1 \right) \log \left( \frac{E}{N} - 1 \right) \right\}. \quad (5.43)$$

To obtain  $E_*$  we set the first derivative equal to zero and solve for  $E_*$ :

$$\frac{d \log F_*}{dE} = -\beta + \frac{E_*}{E_* - N} = 0,$$

from which we obtain

$$E_* = \frac{N}{1 - e^{-\beta}}. \quad (5.44)$$

The second derivative is

$$\frac{d^2 \log F_*}{dE_*^2} = -\frac{N}{E_*(E_* - N)}$$

from which we obtain the variance,

$$\sigma_E^2 = \frac{E_*(E_* - N)}{N}. \quad (5.45)$$

Notice that  $\sigma_E^2$  is extensive and scales as  $\sim N$  when the size of the system increases under all intensive properties constant. The ratio  $\sigma_E/E_*$  scales as  $1/\sqrt{N}$  and this means that the width of the distribution becomes narrower when measured relative to the mean as the size of the system increases. This is the reason we do not need terms higher than second order to represent the asymptotic distribution.

The last step is to determine the integral of  $F$  over all energies. This integral is to be evaluated from  $E = N$  to  $\infty$ , but since the range of energies that contribute to this integral is limited to a narrow region around  $E_*$  we may as well shift the lower limit to  $-\infty$  without error. Then we have,

$$Q = \sum_E F(E) \rightarrow \int_0^\infty F_* \exp \left[ -\frac{(E - E_*)^2}{2\sigma_E^2} \right] dE_* = \sqrt{2\pi\sigma_E^2} F_*$$

with  $F_* = \Omega(E_*, V, N) e^{-\beta E_*}$  and  $\sigma_E^2$  from Eq. (5.45). The result is

$$Q = \sqrt{2\pi\sigma_E^2} \Omega(E_*, V, N) e^{-\beta E_*} \quad (5.46)$$

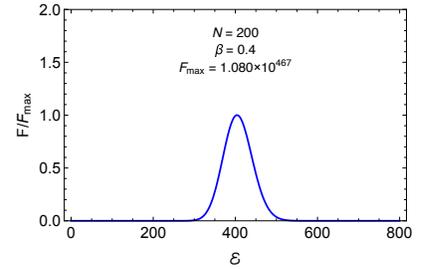
and finally

$$Q = \sqrt{2\pi \frac{E_*(E_* - N)}{N}} \left( \frac{V}{N} \right) \left( \frac{E_* - 1}{N - 1} \right) e^{-\beta E_*}. \quad (5.47)$$

Recall that  $E_*$  is a function of  $\beta$  (Eq. 5.44), therefore the canonical partition function is properly expressed in terms of the variable set  $(\beta, V, N)$ .

The partition function itself is of no particular interest, it its *logarithm* that matters in thermodynamics. The log of  $Q$  in Eq. (5.47) is,

$$\log Q = \log \Omega(E_*, V, N) - \beta E_* + O(\log N).$$



**Figure 5.6:** For large  $E, N$ , function  $F(E)$  approaches the Gaussian form expressed by Eq. (5.42). The value of  $F$  at the maximum is astronomically enormous even though  $N$  is minuscule compared to Avogadro's number!

The last term is the log of the square-root factor. This term scales as  $\sim \log N$  whereas  $\log \Omega(\bar{E}_*)$  is extensive and scales as  $\sim N$ . We drop this term and notice that the result is the same as that in Eq. (5.43) with  $E$  replaced by  $E_*$ :

$$\log Q = -\beta E_* + N \left\{ \log \frac{VE_*}{N^2} - \left( \frac{V}{N} - 1 \right) \log \left( \frac{V}{N} - 1 \right) - \left( \frac{E_*}{N} - 1 \right) \log \left( \frac{E_*}{N} - 1 \right) \right\}. \quad (5.48)$$

Using Eq. (5.44) for  $E_*$  the result may be expressed in terms of  $\beta$  but this is just a matter of tedious algebra. All thermodynamic properties may now be calculated from this equation.

**Equation of state** The pressure is given by the volume derivative of the partition function,

$$\frac{P}{kT} = \left( \frac{\partial \log Q}{\partial V} \right)_{\beta, N} = \left( \frac{\partial \log Q}{\partial V} \right)_{\beta, N}.$$

Since  $E_*$  depends on  $\beta$  and  $N$ , differentiation at fixed  $(\beta, N)$  is equivalent to differentiation at fixed  $(E_*, N)$ . The result is

$$\frac{P}{kT} = \log \frac{V}{V - N}.$$

This is exactly the same result we obtained in the microcanonical ensemble. This agreement demonstrates that even though the two ensembles consist of very different microstates, the lead to identical thermodynamics.

### 5.3.3 The maximum term method

The calculation of the canonical partition function in the lattice system of the previous example highlights an important element of statistical thermodynamics that is not limited to the simple system considered here but applies with great generality. The summation in energy that produces the canonical partition function consists of terms that peak very sharply around a single energy value (in the lattice example,  $E_*$ ) and this is the *only* energy that matters out of the infinitely many possible energies that are available to a canonical system. From a physical perspective this result is reassuring: even though we permit a canonical system to have any conceivable energy, the system chooses microstates only in the narrow vicinity of a single energy. Indeed, this is the behavior that we expect from a system in stable equilibrium. From a mathematical perspective this observation simplifies enormously the calculation of summations of the form of those in Eq. (5.37). Here we provide a more general form of the argument we used in Example (5.2) that allows us to calculate the summation of the canonical partition function as an integral over a Gaussian function.

The summand in Eq. (5.37),

$$F(E) = \Omega(E, V, N) e^{-\beta E},$$

is a peaked function around a value that we will call  $E_*$ . This is a general result:  $\Omega(E, V, N)$  increases with increasing  $E$  (the higher the energy the more ways to distribute it among a fixed number of particles) while  $\exp(-\beta E)$  decreases. The competing effects produce a maximum. We follow the general procedure as in Example 5.2 and expand  $\log F$  in power series of  $E$  around  $(E - E_*)$ . Why  $\log F$

instead of  $F$ ? Because it is the log of  $F$  that is associated with *extensive* behavior, a mathematical fact we will use. The expansion is

$$\log F = \log F_* + \frac{\log F_*^{(2)}}{2!} (E - E_*)^2 + \dots$$

where  $F_* = F(E_*)$  is the value at the maximum, and  $\log F_*^{(2)}$  is the second derivative evaluated at  $E = E_*$ . We drop terms of order three and higher and invert the logarithm to obtain the Gaussian representation of  $F$ :

$$F(E) = F(E_*) \exp \left\{ -\frac{(E - E_*)^2}{2\sigma_E^2} \right\} \quad (5.49)$$

where

$$\sigma_E^2 = -\frac{1}{F_*^{(2)}}. \quad (5.50)$$

Since  $F$  has a maximum at  $E_*$ , the second derivative is negative and  $\sigma_E^2$  is positive. How does  $\sigma_E^2$  scale with the size of system? The log of  $F_*$  is extensive in  $E_*$ , i.e., homogeneous with degree 1. This makes  $\log F_*^{(2)}$  homogeneous with degree  $-1$ , and  $\sigma_E^2$  homogeneous with degree 1; that is

$$\sigma_E^2 \sim E_*. \quad (5.51)$$

For the relative magnitude of deviations from  $E_*$  we find

$$\frac{\sigma_E}{E_*} \sim \frac{1}{\sqrt{E_*}}. \quad (5.52)$$

As an extensive property,  $E_* \rightarrow \infty$  when the size of the system is increased while holding all intensive properties constant. This means that  $\sigma_E/E_*$  goes to zero, making the Gaussian in Eq. (5.49) increasingly narrower. For a macroscopic system, Eq. (5.49) represents the asymptotic form of  $F(E)$ .<sup>19</sup> The summation of the canonical partition function is now evaluated as an integral,<sup>20</sup>

$$Q = \sum_E \Omega(E, V, N) e^{-\beta E} = \sum_E F(E) \rightarrow \int_{-\infty}^{\infty} F_* \exp \left\{ -\frac{(E - E_*)^2}{2\sigma_E^2} \right\} dE = \sqrt{2\pi\sigma_E^2} F_*. \quad (5.53)$$

For the log of the partition function we obtain

$$\log Q = \log F_* + \log \sqrt{2\pi\sigma_E^2} = \log F_* + O(\log E_*)$$

Recall that  $\sigma_E^2$  scales as  $E_*$ , which means the last term in the equation above scales as  $\log E_*$ . We may drop this term relative to  $\log F_*$ , which is extensive in  $E_*$ . Thus we have  $\log Q = \log F_*$ , and finally, using the original definition of  $F(E)$ ,

$$\log Q(\beta, V, N) = \log \Omega(E_*, V, N) - \beta E_*. \quad (5.54)$$

This is the same as Eq. (5.38) except that the energy that appears in the above equation is the location of the maximum in  $\Omega(\beta, V, N)$ , not the mean energy. But

<sup>19</sup>If  $a(x)$  is the asymptotic form of  $b(x)$  in the limit  $x \rightarrow x_*$ , then

$$\lim_{x \rightarrow x_*} \frac{a(x)}{b(x)} = 1.$$

This is to say that an asymptotic form is not an approximation but the true limiting expression for the function of interest.

<sup>20</sup>Since the Gaussian function becomes very sharply peaked about  $E_*$ , we may replace the lower integration limit by  $-\infty$  regardless of the actual value of  $E$  at the ground state.

the two are the same, as we can confirm (see Example ??). Therefore, the final result of this analysis is

$$\log Q(\beta, V, N) = \log \Omega(\bar{E}, V, N) - \beta \bar{E}, \quad (5.55)$$

in full agreement with Eq. (5.38).

Equation (5.55) restates a result we obtained previously, here however we have followed an approach here is more rigorous as well as more general. What have we accomplished? We have shown that the term

$$\Omega(E, V, N) \exp(-\beta E)$$

is sharply peaked about the mean energy of the canonical ensemble; that the maximum term is

$$\Omega(\bar{E}, V, N) \exp(-\beta \bar{E});$$

and that the log of the canonical partition function is equal to the log of the maximum term in the summation. This situation appears often and the same shortcut works: a summation of terms over an extensive property, here energy, peaks very sharply. We may then obtain the logarithm of the summation by simply taking the logarithm of the maximum term. Aptly, this shortcut is known as the maximum-term method.

### 5.3.4 Energy distribution

The microstates the canonical ensemble can have any energy between 0 and  $\infty$ . What is the probability to find a microstate with energy  $E$ ? It is *not*  $\exp -\beta E / Q$ , because this gives the probability to obtain a *specific* microstate whose energy is  $E$ . Here instead we are interested in the probability to find *a* microstate with energy  $E$ . There are  $\Omega(E, V, N)$  microstates with energy  $E$  and all of them have the same probability  $\exp -\beta E / Q$ . Therefore, the probability to find a microstate with energy  $E$  is

$$f(E) = \frac{\Omega(E, V, N) e^{-\beta E}}{Q(\beta, V, N)}. \quad (5.56)$$

This probability is properly normalized, as we can confirm easily:

$$\sum_E f(E) = \frac{\sum_E \Omega(E, V, N) e^{-\beta E}}{Q(\beta, V, N)} = 1,$$

since the numerator is equal to the canonical partition function (see Eq. (5.37)). Equation (5.56) the distribution of energy in the canonical ensemble.

In section 5.3.3 we obtained the asymptotic form of the numerator and denominator that appear in Eq. (5.56). There we found

$$\Omega(E, V, N) e^{-\beta E} \rightarrow F_* \exp \left\{ -\frac{(E - E_*)^2}{2\sigma_E^2} \right\}$$

and

$$Q(\beta, V, N) = \sqrt{2\pi\sigma_E^2} F_*$$

with

$$\frac{1}{\sigma_E^2} = -\frac{d^2 \log F_*}{dE^2}.$$

Putting these results together, the energy distribution is

$$f(E) = \frac{1}{\sqrt{2\pi\sigma_E^2}} \exp\left\{-\frac{(E-E_*)^2}{2\sigma_E^2}\right\}. \quad (5.57)$$

In this form  $F(E)$  is a continuous function of energy and represents a probability density function such that  $F(E)dE$  is the probability to find the system at a microstate with energy  $E$ .

This is a Gaussian distribution normalized to unit area with mean  $\bar{E} = E_*$  and variance  $\sigma_E^2$ . We see clearly that the most probable energy (the energy at the peak of  $f(E)$ ) and the mean energy of the canonical ensemble are the same. Both  $\bar{E}$  and  $\sigma_E^2$  can be obtained from the canonical partition function. We already have the mean energy in terms of the canonical partition function in Eq. (5.30),

$$\bar{E} = -\left(\frac{\partial \log Q}{\partial \beta}\right)_{VN}. \quad (5.30)$$

A similar result will be obtained for the variance. The easiest approach is to differentiate the above equation with respect to  $\beta$ :

$$\begin{aligned} \frac{\partial^2 \log Q}{\partial \beta^2} &= \frac{\partial(-\bar{E})}{\partial \beta} = -\frac{\partial}{\partial \beta} \left( \sum_i \frac{E_i e^{-\beta E_i}}{Q} \right) = \\ &= \underbrace{\sum_i \frac{E_i^2 e^{-\beta E_i}}{Q}}_{\bar{E}^2} + \underbrace{\sum_i \frac{E_i e^{-\beta E_i}}{Q}}_{\bar{E}} \underbrace{\frac{1}{Q} \left( \frac{\partial Q}{\partial \beta} \right)}_{-\bar{E}}. \end{aligned}$$

The right-hand side is simply the variance of the energy distribution, thus we have:

$$\sigma_E^2 = \left( \frac{\partial^2 \log Q}{\partial^2 \beta} \right)_{VN}. \quad (5.58)$$

We can use a simple scaling argument to obtain the asymptotic behavior of the variance:  $\log Q$  in an extensive property and its derivative with respect to  $\beta$ , which is intensive, leaves the result extensive. The standard deviation  $\sigma_E$  scales as  $\sim N$  whereas the mean energy scales as  $\sim N$  with the size of the system.<sup>21</sup> As a result the magnitude of fluctuations relative to the mean energy decreases as  $1/\sqrt{N}$ . In fact we may express the variance of energy fluctuations in a more explicit form by evaluating the derivative in Eq. (5.58). Noting the the first derivative of  $\log Q$  with respect to  $\beta$  is  $-\bar{E}$  we have

$$\sigma_E^2 = -\left(\frac{\partial \bar{E}}{\partial \beta}\right)_{VN} = -\left(\frac{\partial \bar{E}}{\partial T}\right)_{VN} \left(\frac{dT}{d\beta}\right) = -C_V (-kT^2)$$

and finally

$$\sigma_E^2 = kT^2 C_V. \quad (5.59)$$

<sup>21</sup>To obtain this scaling we increase the size of the system keeping all intensive properties constant. All extensive properties then grow in proportion to  $N$ .

Here  $C_V$  is the extensive heat capacity ( $C_V = Nc_V$ ) and scales as  $\sim N$ , in agreement with the scaling analysis. For the magnitude of fluctuations we now have

$$\frac{\sigma_E}{\bar{E}} = \frac{\sqrt{kT^2 C_V}}{\bar{E}}.$$

Since both  $C_V$  and  $\bar{E}$  are proportional to  $N$ , the relative magnitude of fluctuations  $\sigma_E/\bar{E}$  vanishes as  $1/\sqrt{N}$ .<sup>22</sup> For systems with  $N$  of the order of  $10^{23}$ , energy fluctuations are imperceptible. For all practical purposes the energy of a macroscopic system in thermal equilibrium with the surroundings is constant.

The fact that energy fluctuations are predicted to vanish in macroscopic systems is reassuring. No such fluctuations are observed experimentally; if theory said otherwise, that would mean trouble. More importantly, the vanishing magnitude of energy fluctuations implies that even though the system has access to microstates of *any* energy from zero to infinity, it prefers with overwhelmingly high probability only microstates whose energy is within an extremely narrow band of energies around the mean energy of the system. This is a hallmark of stable equilibrium. Consider a closed system inside insulating walls that is placed inside a bath with the same temperature as the system. Because of the adiabatic walls the microstates of the system are microcanonical and restricted to a single energy. Suppose now we turn the walls from insulating to conducting. This allows the system to explore microstates with energies other than that of the original system. And yet, in this thought experiment we find that the state of the system remains unchanged before and after the change to conducting walls. That's because the system was *already in equilibrium* with the bath even before the walls were turned conducting. The adiabatic constraint is redundant, we do not need it to maintain the equilibrium state of the system. Turning the wall conducting changes nothing at the macroscopic level. Microscopically the picture is different but the net result is the same. The removal of the adiabatic constraint opens up an infinite number of new microstates but have extremely low probabilities, all of them except for those whose energy is extremely close to the energy of the system initially. Effectively, the canonical ensemble is identical to a microcanonical ensemble with energy  $\bar{E}$ . The difference between the populated microstates of the canonical ensemble and the microstates of the microcanonical ensemble is in the fluctuations, but these have very small magnitude and involve an insignificant fraction of microstates. It is the narrow magnitude of energy fluctuations that establishes the full equivalency between the canonical and the microcanonical ensembles.

**Example 5.3.** *Energy fluctuations in lattice model*

Obtain the energy distribution in the lattice model of Example 5.2.

**Solution** We start with Eq. (5.56) which gives the energy distribution in the form,

$$f(E) = \frac{\Omega(E, V, N)e^{-\beta E}}{Q(\beta, V, N)}. \quad 5.56$$

We determined that the numerator on the right-hand side is given asymptotically by

$$\Omega(E, V, N)e^{\beta E} = \Omega(E_*, V, N)e^{\beta E_*} \exp\left[-\frac{(E - E_*)^2}{2\sigma_E^2}\right] \quad 5.42$$

<sup>22</sup>We found the same scaling in Eq. (5.52), except that we used  $E_*$  rather than  $N$  to fix the size of the system. Since we increase the size keeping all intensive properties constant, all extensive properties increase by the same factor. We can use any extensive variable to indicate the size of the system, and all of them give the same one-over-square-root result.

## 5.4. Grand canonical ensemble

and the denominator by

$$Q = \sqrt{2\pi \frac{E_*(E_* - N)}{N}} \Omega(E_*, V, N) e^{\beta E_*}. \quad (5.47)$$

Combining these results the probability distribution of energy becomes

$$f(E) = \frac{1}{\sqrt{2\pi\sigma_E^2}} \exp\left[-\frac{(E - E_*)^2}{2\sigma_E^2}\right] \quad (5.60)$$

with variance

$$\sigma_E^2 = \frac{E_*(E_* - N)}{N}$$

This is a symmetric Gaussian distribution with mean  $\bar{E} = E_*$ . We recognize now that  $E_*$  in all of our previous results may be replaced with the mean energy.

Figure 5.7 shows the energy distribution at fixed  $\bar{E}/N = 10$ , which corresponds to  $\beta = 0.1054$ . As  $N$  is increased at fixed  $\beta$  the distribution moves to larger energies and becomes broader. However, the mean of the distribution grows faster than the standard deviation, which makes the distribution narrower when energy is measured relative to the mean. To see this more clearly we calculate the distribution of the intensive energy  $\epsilon = E/N$ . If we call this distribution  $h(\epsilon)$  we have

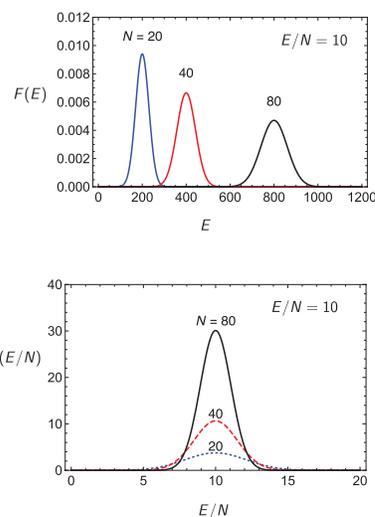
$$h(\epsilon) = f(E) \frac{dE}{d\epsilon} = N f(E/N).$$

This is plotted at the bottom of Fig. 5.7. In this graph as the size of the system increases, the distribution of intensive energy becomes increasingly sharper around its mean value. With  $N$  of the order of Avogadro's number the distribution is practically a delta function at  $E = \bar{E}$ . For all practical reasons, fluctuations of energy in an isothermal system are negligible.

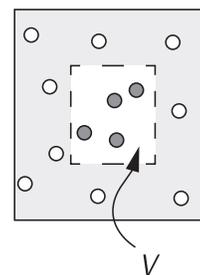
### 5.4 Grand canonical ensemble

So far we have constructed the microstates of an isolated system (microcanonical ensemble) and of a closed system in thermal equilibrium with the surroundings (canonical ensemble). We will now consider an open system of fixed volume that exchanges energy and particles with the surroundings. The physical picture for is a rigid box with holes, is placed inside a bigger system (Fig. 5.8). The system exchanges both energy and particles with the surroundings. On an instantaneous basis the energy and number of particles in the box both fluctuate but at equilibrium their average value is constant over time. In principle, the number of particles in the system may be any number between 0 and  $\infty$ ; and its energy can be anywhere between  $-\infty$  and  $\infty$ . The ensemble consist of all microstates that can be constructed in volume  $V$  with *any* number of particles and *any* total energy. This ensemble is called grand canonical. The microstates of the canonical and of the microcanonical ensemble can be obtained as subsets of the grand canonical.

We now want to determine the probability of grand canonical microstates under the condition that the mean energy is  $\bar{E}$  and the mean number of particles is  $\bar{N}$ . The two specifications state that the system is in equilibrium. Following



**Figure 5.7:** Top: Energy distribution in the lattice model at fixed  $\bar{E}/N = 10$  ( $\beta = 0.1054$ ). Bottom: Distribution of the intensive energy  $E/N$ .



**Figure 5.8:** Schematic of grand canonical system that exchanges energy and mass with the surroundings.

#### 5.4. Grand canonical ensemble

our standard procedure, we determine the unknown probability distribution by maximizing entropy under the constraints

$$\sum_i p_i = 1, \quad (5.61)$$

$$\sum_i E_i p_i = \bar{E} \quad (5.62)$$

$$\sum_i N_i p_i = \bar{N}, \quad (5.63)$$

where  $E_i$  is the energy of microstate  $i$  and  $N_i$  is the number of particles in that microstate. WE use Lagrange multipliers to construct the objective function objective function

$$\max_{p_i, \alpha, \beta, \gamma} \left\{ -\sum_i p_i \log p_i - \alpha \left( \sum_i p_i - 1 \right) - \beta \left( \sum_i E_i p_i - \bar{E} \right) - \gamma \left( \sum_i N_i p_i - \bar{N} \right) \right\}.$$

Maximization with respect to  $p_i$  gives

$$-\log p_i - 1 - \alpha - \beta E_i - \gamma N_i = 0.$$

Setting  $\Xi = -(1 + \alpha)$ , the probability of the grand canonical microstate is

$$p_i = \frac{e^{-\beta E_i - \gamma N_i}}{\Xi}. \quad (5.64)$$

The normalization factor  $\Xi$  is the grand canonical partition function.<sup>23</sup> Inserting this into the first constraint we obtain the partition function as a summation over all microstates of the ensemble:

<sup>23</sup>Pronounced “ksi” or “ksee”.

$$\Xi(\beta, V, \gamma) = \sum_i e^{-\beta E_i - \gamma N_i}, \quad (5.65)$$

The grand canonical partition function is a mathematical function of  $\beta$  and  $\gamma$ , both of which appear in the exponent of the summand, as well as of volume, which defines the microstates of the ensemble. Two results follow immediately. The mean energy is given by

$$\bar{E} = - \left( \frac{\partial \log \Xi}{\partial \beta} \right)_{V, \gamma}, \quad (5.66)$$

and the mean number of particles is

$$\bar{N} = - \left( \frac{\partial \log \Xi}{\partial \gamma} \right)_{\beta, V}, \quad (5.67)$$

Both results follow in the manner of Eq. (5.30) and their derivation is left as an exercise.

**Connection to thermodynamics** The common thread in all ensembles is entropy, and is the property that unlocks the connection between an ensemble of microstates and thermodynamics. The entropy of the grand canonical ensemble is obtained by substituting the probability from Eq. (5.64) into the entropy functional:

$$\frac{S}{k} = - \sum_i p_i (-\beta E_i - \gamma N_i - \log \Xi),$$

or

$$\frac{S}{k} = \log \Xi + \beta \bar{E} + \gamma \bar{N}. \quad (5.68)$$

We solve the above equation for  $\log \Xi$  and use Eqs. (5.66) and (5.67) to express the result in the form,

$$\frac{S}{k} = \log \Xi - \beta \left( \frac{\partial \log \Xi}{\partial \beta} \right)_{V, \gamma} - \gamma \left( \frac{\partial \log \Xi}{\partial \gamma} \right)_{\beta, V}.$$

Written in this form the above equation states that entropy is a Legendre of  $S/k$  with respect to  $\beta$  and  $\gamma$ . Conversely,  $\log \Xi$  is a Legendre transform of entropy. To identify the transformed variables we write 5.68 as

$$\log \Xi = S/k - \beta \bar{E} - \gamma \bar{N}. \quad (5.69)$$

Since  $E$  is the derivative of  $S/k$  with respect to  $1/\beta$ , and  $-N$  is the partial derivative of the same with respect to  $\mu/kT$  (see Eq. (5.11)), we conclude the the right-hand side of the above equation is the Legendre with respect to  $E$  and  $N$ , leading to the immediate identifications,

$$\beta = \frac{1}{kT} \quad (5.70)$$

$$\gamma = -\frac{\mu}{kT} \quad (5.71)$$

By Legendre transforming  $S/k$  with respect to  $E$  and  $N$  we obtain

$$\log \Xi = \frac{S}{k} - \frac{E}{kT} + \frac{\mu N}{kT},$$

or<sup>24</sup>

$$\log \Xi(T, V, \mu) = \frac{PV}{kT}. \quad (5.72)$$

Thus we have identified the thermodynamic significance of the grand canonical partition function. All properties can now be expressed in terms of the grand canonical partition function and its derivatives. The starting point is the differential of  $\log \Xi$ , which we obtain by Legendre transform of the differential of  $S/k$ :

$$d \log \Xi = -\bar{E} d \left( \frac{1}{kT} \right) + \frac{PdV}{kT} + \bar{N} d \left( \frac{\mu}{kT} \right) \quad (5.73)$$

and in a more condensed form,

$$d \log \Xi = -\bar{E} d\beta + \frac{PdV}{kT} + \bar{N} d\gamma. \quad (5.74)$$

Here we use  $\bar{E}$  for the energy and  $\bar{N}$  for the number of particles to indicate that in the grand canonical ensemble these quantities are not fixed but are allowed to fluctuate.

<sup>24</sup>We will also use the alternative notation  $\log \Xi(\beta, V, \gamma)$ , where  $\beta = 1/kT$ ,  $\gamma = -\mu/kT$ .

## 5.4.1 Relationship to other ensembles

The grand canonical partition function in Eq. (5.65) is expressed as a summation over all microstates  $i$  of the ensemble. This summation can be organized by first grouping all microstates with fixed number of particles, then summing over all  $N$ :

$$\Xi = \sum_N e^{\gamma N} \underbrace{\left( \sum_i e^{-\beta E_i} \right)}_{\text{microstates with fixed } N}$$

The inner summation is the canonical partition function,  $Q(\beta, V, N)$ , therefore we have

$$\Xi(\beta, V, \gamma) = \sum_N e^{-\gamma N} Q(\beta, V, N). \quad (5.75)$$

Expressing the canonical partition function as a summation of microcanonical partition functions using Eq. (5.37) as also have

$$\Xi(\beta, V, \gamma) = \sum_N \sum_E e^{-\beta E - \gamma N} \Omega(E, V, N). \quad (5.76)$$

Equations (5.75) and (5.76) express the grand canonical partition function as a summation of weighted canonical or microcanonical partition functions, respectively. The term  $e^{-\beta E_i - \gamma N_i}$  is the statistical weight of the grand canonical microstate, its probability depends only on the number of particles in the microstate and their total energy. The term  $\Omega(E, V, N) e^{-\beta E - \gamma N}$  is the statistical weight of microstates with energy  $E$  and number of particles  $N$ , regardless of all other details of the microstate, namely, particle positions and momenta. This term, which contributes additively to the total value of the grand canonical partition function, is sharply peaked. We can examine the asymptotic behavior of this term by the same methods we used for the factor  $\Omega(E, V, N) e^{-\beta E}$  in the canonical ensemble. Setting

$$F = \log \Omega - \beta E - \gamma N$$

we expand its log in powers of  $E$  and  $N$  around the point  $(E_*, N_*)$  where  $F$  is maximized. Retaining only terms of order 2 we have

$$\log F = \log F_* - \frac{a_1}{2} (E - E_*)^2 - \frac{a_2}{2} (N - N_*)^2 + a_3 (E - E_*)(N - N_*) + \dots$$

where<sup>25</sup>

$$a_1 = -\frac{d^2 \log F_*}{dE^2}, \quad a_2 = -\frac{d^2 \log F_*}{dN^2}, \quad a_3 = \frac{d^2 \log F_*}{dE dN}. \quad (5.77)$$

Coefficients  $a_1$  and  $a_2$  are both positive because the second derivatives at the maximum are negative. We obtain  $F$  by inverting the logarithm,

$$F = F_* \exp \left( -\frac{a_1 (E - E_*)^2}{2} - \frac{a_2 (N - N_*)^2}{2} + a_3 (E - E_*)(N - N_*) \right). \quad (5.78)$$

<sup>25</sup>Notice that  $F$  is a function of  $E$ ,  $N$ ,  $\beta$  and  $\gamma$ . In taking these derivatives,  $\beta$  and  $\gamma$  are held constant. All derivatives are evaluated at  $E = E_*$ ,  $N = N_*$  such that

$$\frac{d \log F_*}{dE} = \frac{d \log F_*}{dN} = 0.$$

This is a two-dimensional Gaussian distribution with mean energy  $\bar{E} = E_*$  and mean number of particles  $\bar{N} = N_*$ . The coefficients  $a_1$ ,  $a_2$  and  $a_3$  are related to the variance and covariance of the distribution. They all scale as  $\sim 1/N$ , which means that the  $F$  becomes increasing sharper around its maximum as the size of the system increases to macroscopic dimensions. The integral of  $F$  over all  $N$  and  $E$  is the grand canonical partition function:<sup>26</sup>

$$\Xi \rightarrow \int_{-\infty}^{-\infty} \int_{-\infty}^{-\infty} F(E, N) dN dE = \frac{2\pi F_*}{\sqrt{a_1 a_2 - a_3^2}}. \quad (5.79)$$

Its log is

$$\log \Xi = \log F_* + \log \frac{2\pi}{\sqrt{a_1 a_2 - a_3^2}} = \log F_* + O(N).$$

The term with the coefficients  $a_1$ ,  $a_2$  and  $a_3$  scales logarithmically with  $N$  (recall that all  $a_i$  scale as  $N$ ) whereas  $\log F_*$  scales as  $N$ . We may drop the slow logarithmic terms. Setting  $\log F_* = \log \Omega(\bar{E}, V, \bar{N}) - \beta \bar{E} - \gamma \bar{N}$ , the final result for the log of the partition function is

$$\log \Xi(\beta, V, N) = \log \Omega(\bar{E}, V, \bar{N}) - \beta \bar{E} - \gamma \bar{N} \quad (5.80)$$

where we have used the fact that  $E_* = \bar{E}$  and  $N_* = \bar{N}$ . This equation is not a new result, it expresses the fact that  $\log \Xi$  is a Legendre transformation of entropy, which we established in Eq. (5.69). Here, however, we have added a new element this relationship: we have shown that the microcanonical ensemble that is associated to the grand canonical is the one with energy  $\bar{E}$  and number of particles  $\bar{N}$ .

*Exercise* Equation (5.80) demonstrates the maximum term method: it says that the log of  $\Xi$  is equal to the log of the maximum term in Eq. (5.76). Show by similar analysis on Eq. (5.75) that the log of the grand canonical partition function can be expressed as

$$\log \Xi(\beta, V, \gamma) = \log Q(\beta, V, \bar{N}) - \gamma \bar{N}.$$

What Legendre transform is indicated by this result?

#### 5.4.2 Grand canonical fluctuations

In the grand canonical ensemble the energy and number of particles are not fixed but vary among microstates. Let us calculate the probability to find a grand canonical microstate with energy  $E$  and number of particles  $N$ . There are  $\Omega(E, V, N)$  microstates that fit this description and all of them have the same probability  $e^{-\beta E - \gamma N} / \Xi$ . The probability to find any of these is

$$f(E, N) = \frac{\Omega(E, V, N)}{\Xi(\beta, V, \gamma)} e^{-\beta E - \gamma N}. \quad (5.81)$$

The numerator of this expression is the function  $F$  of the previous section, whose asymptotic form is given by Eq. (5.78). Combining with the asymptotic form for

<sup>26</sup>Since  $F$  is significant only in a narrow region around its maximum, we can take the integration limits for both variables to range from  $-\infty$  to  $\infty$  without committing error. The result is a standard integral that we can find in mathematical handbooks. The condition for the existence of this integral is  $a_1 a_2 > a_3^2$ . We will not offer a proof that this is indeed so but we will demonstrate this in Example 5.4) using a lattice model.

From Eq. (5.79), we obtain

$$F(E, N) = \frac{\sqrt{a_1 a_2 - a_3^2}}{2\pi} \exp\left(-\frac{a_1(E - E_*)^2}{2} - \frac{a_2(N - N_*)^2}{2} + a_3(E - E_*)(N - N_*)\right), \quad (5.82)$$

with  $a_1$ ,  $a_2$  and  $a_3$  given in Eq. (5.77). This is the joint probability to find a microstate whose energy is  $E$  and whose number of particles is  $N$ . We may obtain the distribution of energies regardless of the number of particles. This is given by the integral of  $f(E, N)$  over all  $N$ :

$$f(E) = \int_{-\infty}^{\infty} F(E, N) dN = \left(\frac{a_1 a_2 - a_3^2}{2\pi a_2}\right)^{1/2} e^{-\frac{(a_1 a_2 - a_3^2)}{2a_2}(E - \bar{E})^2}. \quad (5.83)$$

The result is a Gaussian distribution with mean  $\bar{E}$  and variance  $\sigma_E^2 = a_2 / (a_1 a_2 - a_3^2)$ . The same procedure can be used to obtain the probability to find a microstate with  $N$  particles regardless of its energy:<sup>27</sup>

$$f(N) = \int_{-\infty}^{\infty} F(E, N) dE = \left(\frac{a_1 a_2 - a_3^2}{2\pi a_1}\right)^{1/2} e^{-\frac{(a_1 a_2 - a_3^2)}{2a_1}(N - \bar{N})^2}. \quad (5.84)$$

The conclusion from the analysis of fluctuations is this: while grand canonical microstates are free to have any energy and any number of particles, their vast majority have energy  $\bar{E}$  and number of particles  $\bar{N}$ . Effectively, the grand canonical ensemble is equivalent to a microcanonical ensemble with energy  $\bar{E}$  and number of particles  $\bar{N}$ .

*Exercise* Show that the variance  $\sigma_E$  of energy fluctuations in the grand canonical ensemble is

$$\sigma_E^2 = kT^2 C_V,$$

that is, it is the same as that in the canonical ensemble.

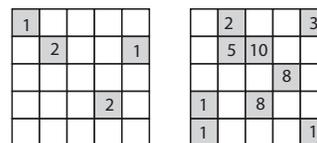
**Example 5.4.** A lattice system in the grand canonical ensemble \_\_\_\_\_  
The best way to demonstrate the theory of the grand canonical ensemble is with an example. We will use the same lattice model of Examples 5.1 and 5.2. In this model we place particles on discrete points of the lattice and give them integer energies. The grand canonical ensemble consists of all microstates that can be constructed with *any* number of particles ( $N = 0, 1, \dots$ ) and with *any* total energy ( $E = N, N + 1, \dots$ ).<sup>28</sup> By selecting microstates with probability  $p_i = e^{-\beta E - \gamma N} / \Xi$  we will obtain the thermodynamic behavior of an equilibrium system with temperature  $1/\beta$  and chemical potential  $\mu$ . We will calculate the grand canonical partition function, the equation of state and the distribution  $F(E, N)$  of energies and particles over the microstates of the ensemble.

**Solution** Since we have determined the microcanonical partition function  $\Omega(E, V, N)$  to be

$$\Omega(E, V, N) = \binom{V}{N} \binom{E-1}{N-1},$$

<sup>27</sup>The result can be obtained simply by switching  $a_1$  and  $a_2$  in Eq. (5.83)

<sup>28</sup>Recall that the minimum energy for a particle is 1; therefore, the minimum total energy of a microstate is  $N$ , corresponding to all particles having the minimum possible energy.



**Figure 5.9:** Two microstates from the grand canonical ensemble of a lattice with  $V = 25$ . One microstate contains  $N = 4$  particles with total energy  $E = 6$  (left) and the other  $N = 9$ ,  $E = 39$  (right).

## 5.4. Grand canonical ensemble

we will construct the grand canonical partition using Eq. (5.76), which expresses  $\Xi$  as a summation of weighted microcanonical partition functions,

$$\Xi = \sum_N \sum_E F(E, N),$$

where

$$F(E, N) = \binom{V}{N} \binom{E-1}{N-1} e^{-\beta E - \gamma N}, \quad (5.85)$$

Before we evaluate the summation it is instructive to inspect the form of  $F(E, N)$ . This is shown in Fig. 5.10 and is calculated for  $V = 950$ ,  $\beta = 1$ ,  $\gamma = 0.9821$ . This function is very sharply peaked at  $E_* = 268.95$ ,  $N_* = 170$  and is of the form of Eq. (5.82) predicted by the asymptotic analysis, i.e., it is Gaussian with respect to both  $E$  and  $N$ . Therefore we have  $\bar{E} = E_*$  and  $\bar{N} = N_*$ .

*Asymptotic form of  $F(E, N)$*  To evaluate the grand canonical partition function we will obtain its asymptotic form by series expansion of its log around its maximum following the procedure discussed in section 5.4.1. Using the Stirling formula  $x! \approx x \log x - x$ , the log of  $F$  is<sup>29</sup>

$$\begin{aligned} \log F = & -\beta E - \gamma N + V \log V - 2N \log N \\ & - E \log E - (V - N) \log(V - N) - (E - N) \log(E - N) \end{aligned} \quad (5.86)$$

To obtain the position of the maximum we set the derivatives with respect to  $E$  and  $N$  equal to zero:

$$\begin{aligned} \frac{d \log F}{dE} = 0 &= -\beta + \log \frac{E_*}{E_* - N_*}, \\ \frac{d \log F}{dN} = 0 &= -\gamma + \log \frac{(V - N_*)(E_* - N_*)}{N_*^2}. \end{aligned}$$

Solving for  $E_*$  and  $N_*$  we obtain

$$E_* = \frac{e^\beta V}{(e^\beta - 1)(e^{\beta+\gamma} - e^\gamma + 1)}, \quad N_* = \frac{V}{e^{\beta+\gamma} - e^\gamma + 1}. \quad (5.87)$$

Next we obtain the parameters  $a_1$ ,  $a_2$  and  $a_3$  from Eq. (5.77),

$$a_1 = -\frac{d^2 \log F_*}{E^2} = \frac{N_*}{N_* E_* (E_* - 1)}, \quad (5.88)$$

$$a_2 = -\frac{d^2 \log F_*}{N^2} = \frac{2VE_* - N_*(V + E_*)}{N_*(V - N_*)(E_* - N_*)}, \quad (5.89)$$

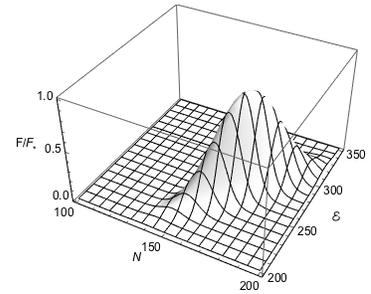
$$a_3 = \frac{d^2 \log F_*}{dE dN} = \frac{1}{E_* - N_*} \quad (5.90)$$

These can be expressed in terms of  $\beta$ ,  $\gamma$  and  $V$  using Eq. (5.87) but this is merely a matter of tedious algebra and we will not do it here. Following Eq. (5.79), the grand canonical partition function is

$$\Xi(\beta, V, \gamma) = \frac{2\pi}{\sqrt{a_1 a_2 - a_3^2}} \binom{V}{\bar{N}} \binom{\bar{E}-1}{\bar{N}-1} e^{-\beta \bar{E} - \gamma \bar{N}}, \quad (5.91)$$

where we have used  $E_* = \bar{E}$  and  $N_* = \bar{N}$ . Along with Eq. (5.87), the right-hand side of this expression is a function of  $\beta$ ,  $V$  and  $\gamma$ . For numerical calculations the factorials should

<sup>29</sup>To obtain this result we assume that  $E, N \gg 1$ . The result is homogeneous with degree 1 in  $E$ ,  $V$  and  $N$ .



**Figure 5.10:** The graph shows function  $F(E, N)$ , defined in Eq. (5.85), evaluated at  $V = 950$ ,  $\beta = 1$ ,  $\gamma = 0.9821$ . The maximum value is  $F_* = 2.25 \times 10^{81}$  and is located at  $E_* = 268.95$ ,  $N_* = 170$ . This two dimensional function is narrowly peaked and occupies a very small area of the  $E \times N$  plane.

## 5.5. Statistical Thermodynamics

be replaced by Stirling's asymptotic formula  $x! \rightarrow \sqrt{2\pi x}(x/e)^x$ . Nonetheless, the partition function itself is of no special interest and we will not invest any more effort on such manipulations.

*Energy fluctuations* We have all the ingredients to calculate the distribution of energies and particle numbers. The energy distribution is given by Eq. (5.83)

$$f(E) = \left( \frac{a_1 a_2 - a_3^2}{2\pi a_2} \right)^{1/2} e^{-\frac{(a_1 a_2 - a_3^2)}{2a_2}(E-\bar{E})^2}, \quad (5.83)$$

with  $a_1$ ,  $a_2$  and  $a_3$  from Eq. (5.88).

*Number fluctuations* The number distribution is

$$f(N) = \left( \frac{a_1 a_2 - a_3^2}{2\pi a_1} \right)^{1/2} e^{-\frac{(a_1 a_2 - a_3^2)}{2a_1}(N-\bar{N})^2}, \quad (5.84)$$

*Logarithm of the grand canonical partition function* The log of the grand canonical partition function from Eq. (5.91) is

$$\log \Xi = \log \Omega(\bar{E}, V, \bar{N}) - \beta \bar{E} - \gamma \bar{N} + O(\log N).$$

The last term is the log of the factor that contains the coefficients  $a_1$ ,  $a_2$  and  $a_3$ . This term increases logarithmically with the size of the system and its contribution can be neglected relative to all the other terms which grow linearly with size. Then,  $\log \Xi$  is equal to  $\log F$  in Eq. (5.86) with  $E$  replaced by  $\bar{E}$  and  $N$  replaced by  $\bar{N}$ :

$$\begin{aligned} \log \Xi = & -\beta \bar{E} - \gamma \bar{N} + V \log V - \bar{N} \log \bar{N} \\ & - E \log \bar{E} - (V - \bar{N}) \log(V - \bar{N}) - (E - \bar{N}) \log(E - \bar{N}). \end{aligned} \quad (5.92)$$

Using Eq. (5.87) and the fact that  $\bar{E} = E_*$ ,  $\bar{N} = N_*$ , we have obtained the grand canonical partition function in terms of its proper variable set,  $(\beta, V, \gamma)$ .

### 5.5 Statistical Thermodynamics

We may now restate the theory and results we have obtained. The log of the microcanonical partition function is a homogeneous function of  $(E, V, n_i)$  with degree 1. By Euler's theorem,

$$\log \Omega = \beta E + \epsilon V + \gamma N, \quad (5.93)$$

where  $\beta$ ,  $\epsilon$  and  $\gamma$  are the derivatives with respect to these variables (Eqs. ??-??). The differential of  $\log \Omega$  is

$$d \log \Omega = \beta dE + \epsilon dV + \gamma dN \quad (5.94)$$

The Legendre transforms of the microcanonical partition function are themselves partition functions of an associated ensemble.<sup>30</sup>

<sup>30</sup>More precisely, the Legendre transform of the *logarithm* of the partition function is the *logarithm* of another partition function. Since we do not have a special name for the log of the partition function, we use the same term, "partition function," to refer to the function itself or its logarithm. The distinction should be clear from the context.

$\log \Omega$  is the Legendre transform of  $\log Q$  with respect to  $\beta$ . By the involution property of the Legendre transform,  $\log \Omega$  is the transform of  $\log \Omega$  with respect to the conjugate variable  $E$ :

$$\log Q = \log \Omega - \beta E. \quad (5.95)$$

It follows from the properties of the Legendre transformation that the differential of  $\log Q$  is

$$d \log Q = -\bar{E} d\beta + \epsilon dV + \gamma dN. \quad (5.96)$$

Here we have used  $\bar{E}$  instead of  $E$  to indicate that when the controlled variable is  $\beta$ , energy is not constant among the microstates of the ensembles. Whenever energy appears in the equations of the canonical ensemble we understand it to refer to the average energy.

From Eq. (5.95) we identify the grand canonical partition function as the Legendre transform of the microcanonical partition function with respect to energy and number of particles:

$$\log \Xi = \log \Omega - \beta E - \gamma N, \quad (5.97)$$

and its differential is

$$d \log \Xi = -\bar{E} d\beta + \epsilon dV - \bar{N} d\gamma. \quad (5.98)$$

Several other ensembles can be defined and it can be shown that their partition functions are related by Legendre transform to the microcanonical partition function. These equations produce a network of relationships between the various quantities that appear in them and are summarized in Table 5.1. These results are not meant to be memorized, they follow by inspection of the fundamental equation for  $\log \Omega$  (the only one to memorize) and its Legendre transforms.

**Equilibrium and stability** Having established that the equilibrium state of an  $(E, V, N)$  system maximizes  $\log \Omega$  with respect to internal partitionings of its energy, volume and number of particles, we may obtain the equilibrium conditions by applying this maximization principle. The derivation follows the same steps as in the classical case. Here we use a slightly modified approach in keeping with the notion of the ensemble. We begin with a system of fixed  $(E, V, N)$  and subdivide it into  $K$  regions of volume  $V_k$  with energy  $E_k$  and number of particles  $N_k$ . The corresponding derivatives of  $\log \Omega_k$  within these volume elements are  $\beta_k$ ,  $\epsilon_k$  and  $\gamma_k$ , respectively, with  $k = 1, \dots, K$ . The total value of  $\log \Omega$  is the sum of the parts, and this is to be maximized with respect to exchanges of energy, volume and number of particles between any two regions of the system. We

**Table 5.1:** Relationships among properties of the microcanonical, canonical, and grand canonical ensembles.

---


$$\begin{aligned}
 \beta &= \left( \frac{\partial \log \Omega}{\partial E} \right)_{VN} \\
 \epsilon &= \left( \frac{\partial \log \Omega}{\partial V} \right)_{EN} = \left( \frac{\partial \log Q}{\partial V} \right)_{\beta N} = \left( \frac{\partial \log \Xi}{\partial V} \right)_{\beta \mu} \\
 \gamma &= \left( \frac{\partial \log \Omega}{\partial N} \right)_{EV} = \left( \frac{\partial \log Q}{\partial N} \right)_{\beta V} = \left( \frac{\partial \log \Xi}{\partial N} \right)_{\beta V} \\
 \bar{E} &= - \left( \frac{\partial \log Q}{\partial \beta} \right)_{VN} = - \left( \frac{\partial \log \Xi}{\partial \beta} \right)_{V\gamma} \\
 \bar{N} &= - \left( \frac{\partial \log \Xi}{\partial \gamma} \right)_{\beta V}
 \end{aligned}$$


---

construct the function to be maximized using Lagrange multipliers,

$$\begin{aligned}
 \max_{E_k, V_k, N_k} \sum_k \log \Omega_k \\
 - \lambda_1 \left( \sum_k E_k - E \right) - \lambda_2 \left( \sum_k V_k - V \right) - \lambda_3 \left( \sum_k N_k - N \right).
 \end{aligned}$$

Setting the derivative with respect to  $E_k$  equal to zero we have

$$\left( \frac{\partial \log \Omega_k}{\partial E_k} \right) - \lambda_1 E_k = \beta_k - \lambda_1 = 0,$$

from which we conclude that  $\beta_k$  has the same value in every volume element. The same result is obtained for  $\epsilon$  and  $\gamma$  and thus we conclude that equilibrium requires  $\beta$ ,  $\epsilon$  and  $\gamma$  to be *uniform* throughout the entire system.

The classical stability criteria are obtained from the concave curvature of  $\log \Omega$ , which requires its second derivatives to be negative. Since the first derivatives of  $\log \Omega$  are all positive, the potential functions are convex, and from their curvature we may obtain additional stability criteria. These results will not be shown here because they follow exactly the same steps as in classical thermodynamics.

## 5.6 Connection to classical thermodynamics

It should be clear by now that all of the new quantities that were introduced in the statistical treatment have their counterpart in classical thermodynamics. We began the development on the premise that the probabilistic functional defined in Eq. (5.2) is the same as the familiar thermodynamic entropy. Energy, volume and number of particles have of course the same meaning in thermodynamics. All other analogies are obtained by comparing Eqs. (5.94) and (1.25). The

following identifications follow by inspection of the two equations:

$$\beta = 1/kT, \quad (5.99)$$

$$\epsilon = P/kT, \quad (5.100)$$

$$\gamma = -\mu/kT. \quad (5.101)$$

The thermodynamic equivalents of the canonical and grand canonical partition functions follow from their Legendre relationships to the microcanonical partition function. Thus we have:

$$\log \Omega = S/k, \quad (5.102)$$

$$\log Q = -A/kT, \quad (5.103)$$

$$\log \Xi = PV/kT. \quad (5.104)$$

In general, partition functions are Legendre transforms of the microcanonical partition function represent various thermodynamic potentials. We have already seen in the discussion of classical thermodynamics that entropy and its Legendre transforms are property generating functions in the sense that all thermodynamic properties may obtained from the generating function and its derivatives. The same is true here. For example, if we take  $Q(\beta, V, N)$  to be the generating function, the equation of state is given by the relationship

$$\epsilon = \left( \frac{\partial \log Q}{\partial V} \right)_{\beta N},$$

which produces an equation that relates  $P, T, V$  and  $N$ .

### 5.6.1 Statistical versus classical thermodynamics

It is clear from the results of the previous section that the statistical treatment reproduces the entire network of thermodynamic relationships of the classical treatment. This is enormous progress, from the classical treatment, but still only half-way to to the full treatment of actual physical systems. What we have gained, relative to classical thermodynamics is a firm connection to a probabilistic treatment of internal states. The macroscopic state defines the set of microscopic states that are possible. The set of these microstates along with their probabilities define the thermodynamic ensemble and give rise the all familiar relationships of thermodynamics. Alternative specifications of the macroscopic state produce alternative ensembles. In all cases, the probability of microstate is of the form

$$p_i = \frac{w_i}{W}, \quad (5.105)$$

where  $w_i$  is the statistical weight of the macroscopic state and  $W$  is the partition function of the ensemble and is fully fixed by the normalization condition of probabilities

$$W = \sum_i w_i. \quad (5.106)$$

The thermodynamic weight is a function of microstate: once the microstate is specified,  $w_i$  is readily calculated. The statistical weights are summarized in

**Table 5.2:** Statistical weights of microstates in the microcanonical, canonical and grand canonical ensembles.

ensemble	weight
microcanonical	1
canonical	$e^{-\beta E_i}$
grand canonical	$e^{-\beta E_i - \gamma N_i}$

Table 5.2 for the three ensemble discussed here. The macroscopic state may be viewed as a set of boundary conditions that define the microstates of the ensemble; every microstate within these boundary conditions belongs to the ensemble and contributes to the measurable properties of the physical system. Microstates themselves are inaccessible: we cannot prepare a system by pinpointing its exact microstate. From a mathematical standpoint the macroscopic state sets the constraints under which to maximize entropy. These constraints are of two types: if an extensive state variable  $X$  is fixed for all microstates, the constraint is built-in into the microstates of the ensemble; if an extensive property is allowed to fluctuate between microstates such that its mean is fixed, then an explicit constraint appears in the maximization of entropy. We used this methodology to obtain the canonical and grand canonical ensemble and we can repeat it to obtain additional ensembles (see problem 5.9).

The statistical treatment replaces the second law with a new postulate according to which, the equilibrium state of the system is determined by the conditions that maximize the entropy functional,

$$S = -k \sum p_i \log p_i.$$

In the context of the statistical treatment, equilibrium refers to the probability distribution of microstates that maximizes this functional.<sup>31</sup>

What is missing to apply this theory to physical systems? We still need to establish what we mean by microstate of a real thermodynamic system in a way that is consistent with the laws of physics, as we understand them. This will be done in chapter 6.

## 5.7 Fluctuations

An isothermal system with volume  $V$  and number of particles  $N$  in thermal equilibrium with a bath, and a system of same volume and number of particles scooped out the same bath and immediately placed in thermal isolation from the surroundings, are equivalent: all their measurable properties are identical. Their corresponding ensembles, however, appear very different from each other. The canonical ensemble contains all microstates that can be constructed with  $N$  particles in volume  $V$ , having energies that range from 0 to  $\infty$ , whereas the microcanonical ensemble contains only a very small subset of microstates with a specific energy.<sup>32</sup> The microcanonical distribution of energy is by definition delta function at  $E = E^*$ .<sup>33</sup> We have established that the canonical distribution of energies has the same mean (see Eq. ??), but is it also a delta function?

In pursuit of this question we first recall Eq. (5.30), which gives the mean energy in the canonical ensemble from the first derivative of  $\log Q$ :

$$\bar{E} = - \left( \frac{\partial \log Q}{\partial \beta} \right)_{VN}. \quad 5.30$$

<sup>31</sup>In section 5.8 we discuss a probabilistic derivation of the entropy functional that leads to the identification of the equilibrium probability distribution as the *most probable distribution* among all possible distributions that we may choose to describe the system.

<sup>32</sup>The term *microcanonical* was coined by Gibbs who viewed it as a differential slice at energy  $(E, E + dE)$  carved out of the continuous distribution of energies of the canonical ensemble.

<sup>33</sup>Actually, a Kronecker delta, since we have taken the approach that microstates are discrete.

Let us calculate the second derivative:

$$\frac{\partial^2 \log Q}{\partial \beta^2} = \frac{\partial(-\bar{E})}{\partial \beta} = -\frac{\partial}{\partial \beta} \left( \sum_i \frac{E_i e^{-\beta E_i}}{Q} \right) =$$

$$\underbrace{\sum_i \frac{E_i^2 e^{-\beta E_i}}{Q}}_{\bar{E}^2} + \underbrace{\sum_i E_i e^{-\beta E_i}}_{\bar{E}} \underbrace{\frac{1}{Q} \left( \frac{\partial Q}{\partial \beta} \right)}_{-\bar{E}}.$$

The right-hand side is the variance of the energy distribution, thus we have:

$$\text{var}(E) = \left( \frac{\partial^2 \log Q}{\partial^2 \beta} \right). \quad (5.107)$$

The square root of the variance is the standard deviation and represents the magnitude of the average energy fluctuation from its mean value. We can formulate a scaling argument to estimate the magnitude of this fluctuation relative to the mean. The log of the canonical partition function is extensive and so are both its first and second derivatives with respect to  $\beta$ , because  $\beta$  is intensive.<sup>34</sup> Accordingly, the variance is extensive (proportional to the number of particles  $N$  when all intensive variables are fixed) and its square root scales as the square root of  $N$ :

$$\delta E = \sqrt{\text{var}(E)} \sim \sqrt{N}.$$

The mean energy is extensive and scales as  $N$ . The ratio of the standard deviation to the mean energy scales then as

$$\delta E / \bar{E} \sim 1 / \sqrt{N}.$$

The magnitude of energy fluctuations relative to the mean energy decreases inversely with the square root of the number of particles and for a system large enough they become vanishingly small. For all practical purposes, the canonical distribution of energies in typical system of macroscopic dimensions is a Dirac delta function. When a microcanonical ensemble is turned into canonical by making thermal contact with a bath at the same temperature, it gains access to a very large number of microstates but populates very few of them such that the vast majority of microstate sampled by the system are found in the vicinity of the mean energy. While nothing prevents the system from spontaneously freezing by transferring all its energy to the bath, this particular microstate has an exceedingly low probability; it is so improbable that for all practical purposes is impossible. The physical reason for this is *dynamics*. When a fluctuation arises in a stable system, it sets up a transport process in the direction of restoring the initial state. It is possible for a fluctuation to grow but as its magnitude increases the restoring force generally becomes much stronger. This makes it very difficult to create large fluctuations.

**Energy distribution and the maximum term method** When we say *energy distribution*, we mean the probability to find a microstate with specified energy  $E$ . This is not the same as the canonical probability in Eq. (5.27), though the two probabilities are related. To establish their relationship, we note that the number

<sup>34</sup>From the perspective of dimensional analysis a derivative is a ratio of two dimensional quantities. The ratio of two extensive properties is intensive; dividing an extensive property with one that is intensive produces an extensive result. These results, which are easily verified for regular division, also apply to differentiation.

of microstates with energy  $E$ , volume  $V$  and number of particles  $N$  is  $\Omega(E, V, N)$ . All of these microstates are present in the canonical ensemble  $(\beta, V, N)$  and all have the same probability,  $\exp(-\beta E)/Q$ . Accordingly, the probability to find a canonical microstate with specified energy  $E$  is

$$\text{Prob}(E|\beta, V, N) = \frac{\Omega(E, V, N)}{Q(\beta, V, N)} e^{-\beta E}, \quad (5.108)$$

where the notation  $\text{Prob}(E|\beta, V, N)$  reads “probability to find energy  $E$  given a canonical ensemble  $(\beta, V, N)$ .” Upon summing both sides of this equation over all energies in the canonical ensemble, the left-hand side gives 1 and the result can be written as<sup>35</sup>

$$Q(\beta, V, N) = \sum_E \Omega(E, V, N) e^{-\beta E}. \quad (5.109)$$

<sup>35</sup>The summation limits are from  $E = 0$  to  $E = \infty$ . They are omitted for simplicity.

We could have obtained this equation directly by noting that the summation over states in Eq. (5.28) can be organized by energy, and since there are  $\Omega(E, V, N)$  microstates with energy  $E$ , the above result follows. In the approach we used here we gained the added benefit of obtaining an expression for the probability distribution of the energy in Eq. (5.108). Taking the log of both sides in Eq. (5.109) we obtain

$$\log Q(\beta, V, N) = \log \sum_E \Omega(E, V, N) e^{-\beta E}. \quad (5.110)$$

Let us calculate the maximum term in the above summation. Since the log is a monotonic function of its argument we might as well maximize the log of the summation terms. Taking the derivative of  $\log \Omega \exp(-\beta E)$  with respect to  $E$  and setting it equal to zero we obtain

$$\frac{\partial(\log \Omega - \beta E)}{\partial E} = 0.$$

Noting that the derivative of  $\log \Omega$  with respect to  $E$  is  $\beta$  (Table 5.1), the solution to this equation is  $E = \bar{E}$ . Therefore, the maximum term of the summation is the term that corresponds to the mean energy.

Recall now that in Eq. (??) we obtained a relationship between the canonical partition function, the microcanonical partition function and the mean energy, which we may express in the form

$$\log Q(\beta, V, N) = \log \Omega(\bar{E}, V, N) e^{-\beta \bar{E}}. \quad (5.111)$$

Comparison of the last two results leads us to conclude that the log of the summation in Eq. (5.110) has the limiting value

$$\log \sum_E \Omega(E, V, N) e^{-\beta E} \rightarrow \log \Omega(\bar{E}, V, N) e^{-\beta \bar{E}}, \quad (5.112)$$

which amounts to replacing the log of the sum with the log of its maximum term,  $\Omega(\bar{E}, V, N) \exp(-\beta \bar{E})$ .<sup>36</sup> This is known as the maximum term method, a shortcut that may be used to simplify the log of summations that run over extensive properties that fluctuate very narrowly around their mean value.<sup>37</sup>

<sup>36</sup>We treat this as a limit because to derive Eq. ?? we assumed a system large enough such that fluctuations could be neglected.

<sup>37</sup>This shortcut is acceptable for the log of the summation but not for the summation itself!

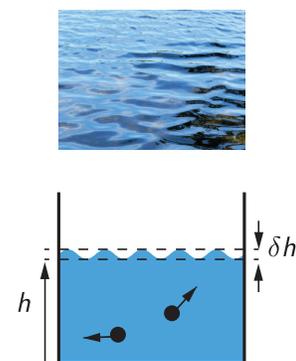
**Fluctuations and stable equilibrium** Equation (5.112) offers some additional insight into the nature of fluctuations and the relationship between ensembles. The intuitive meaning of that equation is this. Upon increasing  $E$  at fixed  $V$  and  $N$ ,  $\Omega$  increases, and enormously so. On the other hand, the statistical weight of microstates is  $\exp(-\beta E)$  and this decreases exponentially with energy. Accordingly, the product  $\Omega \exp(-\beta E)$  has a maximum and this is located at the mean energy of the canonical ensemble. This maximum is very sharp, as a result the microstates of the canonical ensemble, although free to roam in the entire domain of microstates from zero energy to infinity, they are found to be mostly around the same microstates as those of the corresponding microcanonical ensemble. An analogy can be proposed with a macroscopic fluid in a container (Fig. 5.11). The macroscopic state here is defined by the amount of liquid and the shape of the container. The equilibrium shape of the free surface is flat but its instantaneous shape is perturbed by waves. The potential energy then is not uniform across the system; if we subdivide the system into vertical slices of thickness  $\delta x$  (we imagine this as a two dimensional problem), the potential energy of the slice is

$$E_i = \rho g h_i,$$

where  $h_i$  is the height of the slice,  $\rho$  is the fluid density, which we take to be constant, and  $g$  is the gravitational acceleration. Fluctuations of energy are directly proportional to the height  $h_i$ , whose fluctuations are visible to the naked eye. We will work then with  $h_i$  instead of  $E_i$ . If the container has a flat lid such that there is no air above the liquid, the free surface is forced to be flat and all microstates of this system will have the same height. By microstate we refer to the exact position and velocity of every molecule in this fluid. This gives an analogy for the microcanonical ensemble. If we take the lid off and assume that the walls of the container are tall enough, the fluid will gain access to many more microstates since it is not contained to a flat surface any more. Some of these microstates have wild shapes, for example, the liquid might up one wall, form a hole, or produce a mini tsunami wave that travels from wall to wall. And yet we know from experience that that out of all these microstates the system will explore only those that represent small perturbations on the equilibrium shape of the free surface. There is a physical reason for this. If we imagine fluctuations as a sine wave perturbation on the flat surface, the forces that develop point in the direction of restoring equilibrium: the high peak of the wave produces a larger hydrostatic pressure relative to the low peak and the resulting force tends to equalize pressures by erasing the fluctuation. Against this force acts the momentum of the fluid, and this temporarily may allow the fluctuation to grow, but as it continues to grow, the opposing force grows in magnitude and eventually the balance is in favor of the restoring force. The net result is that with or without the lid, the liquid is essentially at the same state, never too far from the equilibrium shape of the free surface.

## 5.8 A combinatorial view of entropy

We introduced the entropy functional in a somewhat unintuitive fashion. We set out to find a measure of “uncertainty” of a probability distribution, specified a set of properties that such measure ought to have, and determined that



**Figure 5.11:** A fluid analogy to ensembles.

the mathematical expression that satisfies the stated properties is the familiar entropy functional:

$$S = -k \sum p_i \log p_i.$$

We will provide a different justification for this result based on a model rather than the abstract notion of uncertainty.

The problem is to determine the unknown probability distribution of a stochastic variable  $X$  based on a few macroscopic measurements, for example, the mean  $\bar{X}$ . For simplicity we will assume that  $X$  takes positive integer values. There is an infinite number of probability distributions whose average is  $\bar{X}$ . We will choose our distribution to be the *average* among all probability distributions that have the same average. In doing so we acknowledge that if we have no reason to pick one distribution over the other, we must allow for the possibility that all feasible distributions must contribute to the one we choose. This is consistent with the principle of maximum uncertainty in the sense that anything that can happen contributes to the macroscopic observable, except that we do not have to introduce such a fuzzy concept as uncertainty. What we need is the probability of every possible probability distribution.<sup>38</sup> To do this we construct the ensemble of all possible distributions. We do this as follows. We toss a die with a large number of sides  $N$  times and keep the result only if the sum of the  $N$  tosses is  $X$ . The ordered sequence of outcomes is a “configuration”. Suppose in a configuration we have  $n_1$  ones,  $n_2$  twos, and so on. The  $n_i$  represent the frequency distribution of outcomes in the configuration and satisfy the conditions,

$$\sum n_i = N, \quad \sum X_i n_i = X. \quad (5.113)$$

The normalized frequency  $p_i = n_i/N$  is a probability distribution whose mean is  $X/N$ . The ensemble consists of *all* distributions  $p_i$  that we can form by this procedure. If we let  $X$  and  $N$  be big at fixed  $X/N$ , we obtain the ensemble of all distributions  $p_i$ ,  $i = 1, \dots, \infty$ , whose mean is  $\bar{X}$ . A distribution  $\{n_i\}$  can appear in any possible permutation of the outcomes that represent the distribution. This number is given by the multinomial coefficient

$$\frac{N!}{n_1 n_2! \dots}$$

The multinomial coefficient of distribution  $\{n_i\}$  is the multiplicity of the distribution in the ensemble. Accordingly, the probability of distribution is proportional to its multiplicity factor. Now that we know the probability of each distribution we can calculate the average distribution. This task is simplified enormously if we recognize that for large  $X$  and  $N$ , the multinomial factor peaks very sharply and the mean distribution is the same to the most probable distribution. We will not prove this fact but the general steps are very similar to those that lead to the maximum term method in section 5.3.3.<sup>39</sup> Suffices then to determine the distribution with the maximum multinomial coefficient. We do this by maximizing the log of the multinomial,

$$\log \frac{N!}{n_1 n_2! \dots} \approx - \sum n_i \log \frac{n_i}{N} = -N \sum p_i \log p_i$$

with respect to  $p_i$  subject to the constraints in Eq. (5.113), which we now express in terms of  $p_i = n_i/N$ :

$$\sum p_i = 1, \quad \sum X_i p_i = \bar{X}.$$

<sup>38</sup>This may sound like a brain twister but we are actually referring to the probability of *probability*, namely, the probability to select a particular probability distribution among all distributions that fit the data.

<sup>39</sup>In our analysis in section (5.3.3) we found that the distribution of energies is a Gaussian function from which we concluded that the mean energy and the most probable energy are the same. The situation is similar here with respect to the distribution of the multinomial factors.

Using the familiar by now method of Lagrange multipliers, the solution is

$$\text{most probable distribution: } p_i = \frac{e^{-\beta x_i}}{Q}.$$

The result is the canonical probability. If we use only the first constraint, the result will be the uniform distribution; if we add a new constraint for a second variable  $Y_i$  with fixed mean  $\bar{Y}$ , we will obtain the grand canonical. Therefore, this alternate approach generates the correct probabilities of microstate as the method based on the maximization of entropy. This is not surprising, of course. We recognize that the maximization of entropy in our original approach is equivalent to the maximization of the multinomial coefficient, which in turn amounts to maximizing the probability of  $p_i$  in the ensemble of all distributions with the same mean. Mathematically, the maximization of entropy and the maximization of the multinomial coefficient are equivalent. Nonetheless, the combinatorial approach may seem more intuitive because it obtains the unknown probability through a stochastic process, the tossing of a super-big die. Entropy in this approach appears naturally, it is the log of the multiplicity (statistical weight) of a distribution – and the distribution with the maximum weight wins. It also helps us place a simple probabilistic interpretation of entropy. The multinomial coefficient, whose log is the entropy functional, represents the number of ways that a particular distribution may be realized, that is, the number of permutations in the outcomes of the die toss with the same distribution. Uncertainty in this view is associated with the number of different ways a distribution can materialize in the sense that we have higher level of *ignorance* as to the specific sequence of tosses that produces a given distribution.

The combinatorial view interprets thermodynamics as a calculus that applies to *probability distributions*. The identification of entropy as a functional of the probability distribution tells us that entropy is not a “physical” property like mass, energy, etc., but a property of stochastic systems in general.

**Calculus of variations** In calculus we often deal with unknown functions that are determined by solving a differential equation. In this problem the unknown function is a probability distribution but its solution is obtained not from a differential equation, but by solving a maximization problem, namely, by identifying the function that maximizes a particular functional (in this case, entropy). The branch of mathematics that studies this type of problems is variational calculus.<sup>40</sup> It studies the extrema of scalars that depend on functions (here, the unknown probability distribution) in the limit that these functions become continuous,

$$p_i \rightarrow f(x).$$

In light of this limiting behavior, the statement that in the thermodynamic limit the ensemble converges to the most probable distribution must be understood in context. It is not strictly true, as we stated, that a *single* distribution becomes overwhelmingly more probable than all others. In reality, when the probability distribution approaches the continuous limit, the neighborhood around any distribution contains distributions that represent differential variations of the distribution of interest. In the discrete domain these distributions are distinctly different from each other and their number is very large, but in the continuous

<sup>40</sup>The stochastic variable in the problem we have been discussing in this section is not just a variable but an entire probability distribution. If you think about it, we have been talking about the probability of a probability distribution, and about the distribution of distributions. This mind twister becomes easier to resolve if we simply view distributions as mathematical objects that can also be treated as variables and serve as inputs to other functions (actually, functionals).

limit their difference becomes a differential. In this case the distribution of distributions is very sharp relative to its maximum term.

## 5.9 A brief history of the entropy functional

The association of entropy with the expression in Eq. (5.2) has deep roots in thermodynamics. Its most obvious link is to the ideal entropy of mixing: if we interpret the mol fraction  $x_i$  of component  $i$  as the probability to find a particle of type  $i$  in the solution, then the entropy of  $x_i$  is the ideal entropy of mixing per particle:

$$\frac{\Delta S_{\text{mix}}}{N} = -k \sum x_i \log x_i = S[x_i].$$

Boltzmann obtained this form of entropy for an ideal gas by assuming that the state of a molecule can be discretized to take discrete values in space and energy with discrete probability  $p_i$ . These examples connect thermodynamic entropy to the entropy functional but in rather limited cases, ideal solution and ideal gas. Gibbs provided the most general connection between the classical and statistical versions of entropy by showing that the entropy functional has all the hallmarks of classical entropy. More importantly, he showed that the probability of microstate in each ensemble maximizes the entropy functional with respect to all distributions that may be assigned to the microstates of the ensemble. That is, the maximization of entropy signifies the selection of the *correct* probability distribution. Gibbs published his work in 1902 (Gibbs, 1902). Forty-six years later, Shannon came across the same functional from an entirely different perspective (Shannon, 1948). Shannon, who worked for Bell Labs, was searching for a measure of the information content of a signal and he found it in the entropy functional. He borrowed the name “entropy” from thermodynamics because of the identical mathematical form of his result, even though his derivation bears no connection to thermodynamics. Shannon’s work became the foundation of Information Theory, whose applications extend to areas well beyond signal processing. In 1957, Jaynes suggested a formal connection between Shannon’s view of entropy and that of Gibbs’s (Jaynes, 1957). The central problem in statistical thermodynamics is the assignment of probabilities to the microstates of a given ensemble. If we adopt Shannon’s entropy as a measure of uncertainty, Jaynes suggested, then the maximization of entropy amounts to determining the distribution with the maximum uncertainty, namely, the distribution that is most faithful to the available information, where by information we mean any *macroscopic constraints* that are placed in the data. This view provides a unified approach that works with all ensembles and gives the correct answer for the probability distribution and all thermodynamic properties that are calculated from it. It also provides an interpretation of entropy that is less unintuitive than the classical view. The combinatorial view of entropy provides a more refined interpretation that some may find even more intuitive: it is the “multiplicity” –or statistical weight– of a probability distribution in an ensemble that contains all probability distributions that fit the constraints placed on the macroscopic state. Maximizing entropy amounts to selecting the probability distribution that is most probable among all distributions of the ensemble.

## 5.10 Additional reading

1. The combinatorial derivation of ensembles is discussed in [Hill \(1986\)](#) (chapters 1 and 2) and in [\(McQuarrie, 2001\)](#) (chapters 2 and 3).
2. The unified derivation of the various ensembles is due to E. T. Jaynes. Jaynes drew the connection between Shannon's information theory ([Shannon, 1948](#)) and statistical mechanics, and formulated the principle of maximum entropy as a general method of inference. His 1957 paper is a classic ([Jaynes, 1957](#)).
3. The classical source for information theory and the interpretation of entropy as a measure of uncertainty is Shannon's original paper, ([Shannon, 1948](#)) but its focus is signal processing, which may pose some barriers to readers from other backgrounds. A more accessible presentation is given by ([Kapur, 1989](#)).

## 5.11 Problems

**Problem 5.1.** a) Calculate the entropy of a fair coin.

b) Calculate the entropy of a coin that is rigged so that heads appears twice as many times as tails.

c) Is it true that the entropy of a rigged coin is always less than the entropy of a fair coin?

*Note:* If this problem seems to you to be unrelated to thermodynamics, take the coin to be an electron that can be in one of two spin states  $i = 1$  and  $i = 2$  with probability  $p_i$  that is controlled by an external magnetic field.

**Problem 5.2.** Begin with the premise that the probability of a microstate  $i$  depends solely on the energy of the microstate and show that the probability of a microstate in the canonical ensemble is

$$p_i = \frac{e^{-\beta E_i}}{\sum_i e^{-\beta E_i}}$$

Hint: Consider two systems in thermal contact and argue that probability to find the system 1 at energy  $E_1$  and system 2 at energy  $E_2$  is

$$p(E_1 + E_2) = p(E_1)p(E_2)$$

where  $p(E)$  is the probability to find an  $N, V, T$  system at energy  $E$ .

**Problem 5.3.** Start with  $M$  distinguishable buckets and  $N$  indistinguishable balls. The buckets are divided into two groups that contain  $M_1$  and  $M_2 = M - M_1$  buckets, respectively. Your job is to divide the  $N$  balls between the two systems.

a) Show that the partitioning that maximizes the product  $\Omega_1 \Omega_2$  satisfies the condition  $N_1 / M_1 = N_2 / M_2$ .

b) Use the full Stirling asymptotic formula to express the the maximum value of  $\log \Omega_1 \Omega_2$  in terms of  $M_1$ ,  $M_2$  and  $x$ , where  $x$  is the common value of the ratio  $M_1 / N_1 = M_2 / N_2$ .

Hint: Show first that the binomial coefficient can be written in the form

$$\log \frac{(a+b)!}{a!b!} = -(a+b)(x \log x + y \log y) + \log \sqrt{2\pi xy(a+b)}$$

where  $x = a/(a+b)$  and  $y = b/(a+b)$ .

c) Show that

$$\log \Omega_{1+2} - \log \Omega_1^* - \log \Omega_2^* \geq 0$$

## 5.11. Problems

where  $\Omega_{1+2}$  is the number of configurations of the combined system.

d) Obtain the limiting relationship between  $\log\Omega_{1+2}$  and  $\log\Omega_1^* + \log\Omega_2^*$  for large  $M_1, M_2$ .

e) Discuss how this problem of balls and buckets relates to thermodynamics of common matter.

**Problem 5.4.** Each of the expressions below represents a probability. Explain clearly what this probability is ( $i$  refers to microstate):

$$\begin{array}{ll}
 \text{(a)} \quad \frac{e^{-\beta E_i}}{\sum_E \Omega(E, V, N) e^{-\beta E}} & \text{(b)} \quad \frac{\Omega(E, V, N) e^{-\beta E}}{\sum_E \Omega(E, V, N) e^{-\beta E}} \\
 \text{(c)} \quad \frac{e^{-\beta E_i - \gamma N_i}}{\sum_N Q(\beta, N, V) e^{-\gamma N}} & \text{(d)} \quad \frac{Q(\beta, N, V) e^{-\gamma N}}{\sum_N Q(\beta, N, V) e^{-\gamma N}} \\
 \text{(e)} \quad \frac{Q(\beta, N, V) e^{-\gamma N}}{\sum_N e^{-\gamma N} \sum_E \Omega(E, V, N) e^{-\beta E}} & \text{(f)} \quad \frac{e^{-\beta E}}{\Xi(\beta, V, \gamma)} \sum_N \Omega(E, V, N) e^{-\gamma N}
 \end{array}$$

**Problem 5.5.** a) Express the grand canonical partition function as a summation that involves the canonical partition function. Show that when fluctuations in energy and in the number of particles are negligible, this result expresses a familiar equation in thermodynamics (take the logarithm of the result and identify the various quantities by their thermodynamic names).

b) Express the grand canonical partition function as a summation that involves the microcanonical partition function. Show that when fluctuations in energy and in the number of particles are negligible, this result expresses a familiar equation in thermodynamics (use the same hint as above).

**Problem 5.6.** a) Show that the probability to find a microstate with energy  $E$  and number of particles  $N$  in the grand canonical ensemble, is

$$p(E, N) = \Omega(E, V, N) \frac{e^{-\beta E - \gamma N}}{\Xi}.$$

where  $\Omega(E, V, N)$  is the microcanonical partition function.

b) Explain how the above result is different from the grand canonical probability of microstate,

$$p_i = \frac{e^{-\beta E_i - \gamma N_i}}{\Xi}.$$

c) Write the normalization condition for the probability  $p(E, N)$ . Take the logarithm, apply the method of the maximum term, and show that the result is

$$\log\Omega = \log\Xi - \beta \left( \frac{\partial \log\Xi}{\partial \beta} \right)_{\gamma V} - \gamma \left( \frac{\partial \log\Xi}{\partial \gamma} \right)_{\beta V}$$

d) Show the equalities,

$$\beta = \left( \frac{\partial \log\Omega}{\partial E} \right)_{VN} \quad \gamma = \left( \frac{\partial \log\Omega}{\partial N} \right)_{EV}.$$

where  $\beta$  and  $\gamma$  are the parameters in the probability of microstate in the grand canonical ensemble.

**Problem 5.7.** Show that if the probability of microstate is given by

$$p_i = \frac{e^{-\alpha_1 F_{1i} - \alpha_2 F_{2i} \dots}}{\Psi}$$

the mean and variance of property  $F_j$  are given by

$$\langle F_j \rangle = -\frac{\partial \ln \Psi}{\partial \alpha_j}, \quad \text{var}(F_j) = \frac{\partial^2 \ln \Psi}{\partial \alpha_j^2}$$

**Problem 5.8.** Consider the ensemble defined by the fixed  $E$  and  $V$ , but the number of particles is allowed to fluctuate about  $\bar{N}$ .

- Obtain the probability of microstate and identify all variables by their thermodynamic equivalents.
- Identify the independent variables of the partition function.
- Obtain the mean number of particles in terms of the partition function and its derivatives.

**Problem 5.9.** Consider the ensemble of discrete microstates of  $N$  identical particles in a closed system that interacts with its surroundings by exchanging of energy (via diathermal walls) and volume (via a movable piston).

- Show that the probability of microstate  $i$  is

$$p_i = \frac{e^{-\beta E_i - \delta V_i}}{\Delta}$$

where  $\Delta$  is the partition function.

- Explain why  $\Delta$  is a function of  $\beta$ ,  $\delta$  and  $N$  but not of  $E_i$  or  $V_i$ .
- Obtain the mean energy  $\bar{E}$  and the mean volume  $\bar{V}$  in terms of the partition function.
- Relate  $\Delta$ ,  $\beta$  and  $\delta$  to known thermodynamic quantities.
- Determine the equation of state if  $\Delta$  is given by

$$\Delta(\beta, \delta, N) = \left( \frac{e^{-a\delta}}{\Lambda^3 \delta} \right)^N$$

where  $a$  and  $\Lambda$  are constants.

**Problem 5.10.** Consider a system defined by its energy  $E$ , number of particles  $N$ , and mean volume  $\bar{V}$ . Construct the corresponding ensemble that contains of  $n_{\text{tot}}$  systems at all possible microstates  $i$ . All microstates have the same  $E$  and  $N$  but different volume  $V_i$ , such that

$$\sum_{i=1}^{n_{\text{tot}}} n_i V_i = n_{\text{tot}} \bar{V}$$

where  $n_i$  is the number of systems in microstate  $i$ , and  $\sum n_i = n_{\text{tot}}$ .

- Show that the probability of microstate  $i$  is of the form

$$p_i = a e^{-b V_i}$$

where  $a$  and  $b$  are functions of  $E$ ,  $N$  and  $\bar{V}$ .

- Express the derivative

$$\left( \frac{\partial \log a}{\partial b} \right)_{E, N}$$

in terms of common thermodynamic properties.

- Relate  $a$  and  $b$  to common thermodynamic properties.

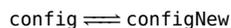
**Problem 5.11.** Consider a system whose energy, number of particles, and volume are allowed to fluctuate around their mean values,  $\bar{E}$ ,  $\bar{N}$ ,  $\bar{V}$ , respectively. Construct the corresponding ensemble of microstates  $i$  such that the number of particles is  $N_i$ , the energy is  $E_i$  and the volume is  $V_i$ .

- Write the probability of microstate  $i$  in terms of  $N_i$ ,  $E_i$ ,  $V_i$ . Identify any variables in this probability with their corresponding thermodynamic property.
- Which Legendre transform corresponds to this ensemble?
- Show that the partition function of the ensemble is 1.

**Problem 5.12.** (Monte Carlo simulation of exchange reaction) Setup a Monte Carlo simulation of the process in problem 4.8 as follows. Represent the configuration as a list,  $\text{config} = \{x_1, x_2, \dots\}$ . Pick two elements at random, say  $x_i$  and  $x_j$ , and replace them with two new positive integers  $x'_i$  and  $x'_j$  such that  $x_i + x_j = x'_i + x'_j$ . Call the new configuration,  $\text{configNew}$ . Replace the old configuration with the new if the following criterion is satisfied:

$$\text{rnd} < \frac{\text{Prob}(\text{configNew})}{\text{Prob}(\text{config})}$$

otherwise, retain the old configuration, where  $\text{rnd}$  is a random number between 0 and 1. This simulates the reversible “reaction”



Repeat this process a large number of times and keep track of the average distribution. Compare the average distribution  $\bar{\mathbf{n}}$  to the one you calculated in problem 4.8 by showing both distributions on the same graph.

**Problem 5.13.** (Monte Carlo simulation of exchange reaction) Consider  $M = 50$  monomers that are distributed into  $N = 20$  clusters. A configuration is an ordered list of  $N$  clusters with total mass  $M$ . Each configuration is equally probable. The distribution of configuration is the list  $(n_1, n_2, \dots)$ , where  $n_i$  is the number of clusters with mass  $i$ .

- Show that the most probable distribution  $\tilde{n}_i$  is

$$\frac{\tilde{n}_i}{N} = \frac{e^{-\beta i}}{q}$$

where  $\beta$  and  $q$  are given by the equations,

$$q = \sum_{i=1}^n e^{-\beta i}, \quad \frac{M}{N} = \frac{\sum_{i=1}^N i e^{-\beta i}}{q}$$

- Calculate the numerical value of  $\beta$  and  $q$ .
- Sample configurations from this ensemble by Monte Carlo and use the results to calculate the mean distribution. Compare the results with the most probable distribution you obtained above by showing both distributions on the same graph. This MC simulation can be conducted by the “exchange reaction” method in which we select two clusters at random and replace them with two new clusters with the same mass. (A mathematical notebook will be provided.)

**Problem 5.14.** Consider the ensemble of microstates that represent a closed system that exchanges energy and volume with its surroundings while the number of particles is fixed.

- Obtain the probability distribution, and identify the partition function,  $\Delta$ , in terms of known thermodynamic variables.
- Obtain a relationship between  $\Delta$  and the canonical partition function. *Hint:* Arrange the sum-over states in terms of energy first.

## 5.11. Problems

**Problem 5.15.** a) A college professor who taught graduate stat thermo to a class of 25 very bright students decided to assign grades entirely randomly, such that a student has equal probability to receive any of the five grades,  $A, B, C, D$  or  $F$ . We denote the grade distribution by  $(n_A, n_B, n_C, n_D, n_F)$ , where  $n_X$  is the number of students who got letter grade  $X$ . Determine the most probable grade distribution, the most probable letter grade, and the mean GPA in the most probable distribution based on the scale  $A = 4, B = 3, C = 2, D = 1$ , and  $F = 0$ .

b) Several students complained that grading was unfair and the professor decided to curve the grades to a mean GPA of 3.5. He did this by picking at random one grade distribution among all possible distributions with mean GPA 3.5. He then assigned the grades to each student randomly. Determine the most probable grade distribution, the most probable letter grade, and the mean GPA in the most probable distribution.

PS: When the students finally confronted the professor for disregarding individual effort in assigning grades, his answer was: “you are all so bright, you are indistinguishable to me.”

**Problem 5.16.** The problem of distributing  $M$  balls into  $K$  buckets randomly so that no bucket is empty has an analytic solution: the probability to obtain  $k_1$  buckets with one ball,  $k_2$  buckets with 2 balls and so on is

$$P(k_1, k_2, \dots) = \frac{K!}{k_1! k_2! \dots} \left/ \binom{M-1}{K-1} \right.$$

a) Determine the distribution with the maximum probability (most probable distribution) using the method of Lagrange multipliers.

b) Make a plot of the maximum probability as a function of  $M$  at fixed  $M/K = 10$ .

c) Determine the asymptotic form of the most probable distribution in the limit  $M \rightarrow \infty, K \rightarrow \infty$  at fixed  $M/K = \bar{n}$ .

**Problem 5.17.** In this variation of the problem of  $M$  balls into  $K$  buckets, suppose that there are  $\omega_i$  ways to produce a “cluster” with  $i$  balls. For example, the balls could be monomers with  $f$  functional groups, and the cluster is a polymer network that contains  $i$  monomers. Assuming that all  $\omega_i$  ways to form the cluster of size  $i$  are equally probable:

a) Obtain the probability to obtain a polymer size distribution with  $k_1$  monomers,  $k_2$  dimers, and so on.

b) Obtain the most probable distribution in the limit  $M, K \rightarrow \infty$  at fixed  $M/K = \bar{n}$ .

**Problem 5.18.** As a model for random copolymerization, consider the ensemble of polymer chains formed from a large pool of  $N_A$  monomers of type  $A$  and  $N_B$  monomers of type  $B$  such that

$$N_A + N_B = N_{\text{tot}} = \text{const.}; \quad N_A / (N_A + N_B) = \phi_A = \text{const.}$$

We distribute the monomers among  $K$  chains ( $K = N_{\text{tot}}/\nu$ ) and record the resulting distribution  $\mathbf{p} = \{p_0, p_1, \dots\}$  such that  $p_i$  is the fraction of chains that contain  $i$  monomers of type  $A$ . The ensemble consists of all possible ways that the  $N_A + N_B$  monomers can be distributed among  $K$  chains.

a) Determine the multiplicity  $w(\mathbf{n})$  of distribution  $\mathbf{n}$ .

b) Obtain the *most probable distribution* in the ensemble.

c) Obtain the partition function,

$$Q = \sum_{\mathbf{n}} w(\mathbf{n})$$

d) Obtain the entropy of the ensemble, defined as

$$S = \ln Q$$

Note: express all of your results in terms of  $\phi_A, \phi_B$  and  $\nu$ .

## Statistical Mechanics

To apply the probabilistic calculus of chapter 5 to real physical systems (gases, liquids and solids composed of pure or mixed components) we must to adopt a physical model of matter that specifies the microstate and relates it to the physical laws that govern its evolution. We may do this at the quantum or at the classical level. As most systems of engineering interest are well within the classical domain, or development will be based on classical mechanics. Classical mechanics is in fact the limiting behavior of quantum mechanics under conditions that make the discreteness of the quantum world vanish into a continuum, and the wave-like nature of matter dissolves into the concrete properties of familiar matter. For a system of  $N$  particles of mass  $m$  in volume  $V$ , the classical treatment applies when the de Broglie wavelength of the particle is much smaller than the mean interparticle distance:

$$\frac{h}{\sqrt{2\pi mkT}} \ll \left(\frac{V}{N}\right)^{1/3} \quad (6.1)$$

where  $h = 6.62607 \times 10^{-34} \text{ m}^2 \text{ kg/s}$  is the Planck constant. Except at very low temperatures or very high densities, classical mechanics provides adequate description of the matter for the purposes of describing its evolution and deriving its thermodynamic behavior.

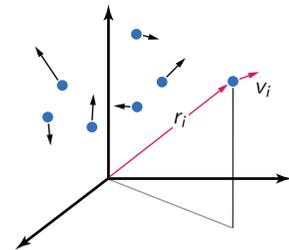
In this chapter we develop the thermodynamics of a *mechanical* system of interacting particles, namely, a collection of particles that evolve in time according to the laws of classical mechanics. Our goal is to define the microstate of the system in terms of continuous mechanical variables and to develop the corresponding continuous expressions for the probability of microstate and the partition function.

### 6.1 Newtonian mechanics

The motion of a classical point particle is governed by Newton's laws, commonly expressed in the form of two differential equations, one for the position of the particle, and one for its velocity:

$$\frac{dr_x}{dt} = v_x, \quad \frac{dv_x}{dt} = \frac{F_x}{m}. \quad (6.2)$$

Similar equations apply to the  $y$  and  $z$  coordinates. The state of a single particle is determined by its instantaneous position  $r = (r_x, r_y, r_z)$  and velocity



**Figure 6.1:** The microstate of  $N$  point masses is defined by the positions and velocities of all particles.

## 6.2. Hamilton's equations

$v = (v_x, v_y, v_z)$ . If this state is known at time  $t_0$  the state at any other time  $t$  is obtained by integrating the equations of motion. This integration can be done either forward or backward in time and allows us to obtain the future as well as the past history of the particle.<sup>1</sup>

When  $N$  particles are present, the force on a particle generally has two contributions, one from all other particles via various types of interactions, for example electrostatic, van der Waals, etc., and one from external objects, for example gravitational forces or forces exerted by the walls of the container.<sup>2</sup> Assuming that all forces are conservative<sup>3</sup> the interaction is described by a potential function,

$$E_P(r_1, r_2, \dots; A_1, A_2, \dots) \quad (6.3)$$

that depends on the positions  $r_i$  of the  $N$  particles, and on external coordinates  $A_i$  that represent the external objects (the walls of the container, for example). For simplicity the external coordinates will be omitted from the potential and will be assumed to be constant. On the other hand, the coordinates of the particles are not constant as particles move through space under the forces exerted on them. The force on particle  $i$  is

$$F_{ix} = -\frac{\partial E_P}{\partial r_{ix}},$$

with similar equations for the  $y$  and  $z$  components of the force. The equations for the force along with the equations of motion in Eq. (6.2) provide a closed set of equations that can be integrated in time. The only additional information we need is the instantaneous positions and velocities of all particles at some time  $t_0$ .<sup>4</sup> The positions and velocities of all particles define the state of this system. The energy of the system is the sum of kinetic and potential energy,

$$E = E_K + E_P = \sum_{i=1}^N \frac{m(v_{xi}^2 + v_{yi}^2 + v_{zi}^2)}{2} + E_P(r_1, r_2, \dots).$$

The energy of the system is immediately known once the state is specified, which is to say, energy is a state function. In general, any mechanical property that can be calculated from the instantaneous positions and velocities of all particles is a state function. Pressure and density are two examples of such properties.

### 6.2 Hamilton's equations

An alternative mathematical description of Newton's equations of motion is Hamilton's form. In this approach we use the momentum  $p = mv$  in place of the velocity. For the total energy of  $N$  particles we now have<sup>5</sup>

$$E_K = \sum_{i=1}^N \frac{p_{xi}^2 + p_{yi}^2 + p_{zi}^2}{2m} + E_P(r_1, r_2, \dots) \doteq \mathcal{H}(\mathbf{p}^N, \mathbf{r}^N). \quad (6.4)$$

The total energy expressed as a function of position and momentum is called the Hamiltonian function of the system and has some special properties that make it very useful. First we calculate the derivative of  $E$  with respect to position

<sup>1</sup>To perform this integration we need to know any external interactions (for example gravitational forces) as a function of time.

<sup>2</sup>While the walls of the container normally do not exert any force in the bulk of the system, they do exert a force on those particles that collide with the wall. This amounts to a potential that is infinite at the wall and zero everywhere else.

<sup>3</sup>A force is conservative if the work is independent of the path and only a function of the end states. Conservative forces can be represented as derivatives of a potential function that depends only on position.

<sup>4</sup>We also need information about the external coordinates. These may be held constant, as assumed here, or they may be known functions of time. Allowing the external coordinates to vary with time amounts to conducting a *process*, for example expansion or compression of the system volume.

<sup>5</sup>For simplicity we take the masses of all particles to be the same. It is straightforward to write these equations for a collection of dissimilar particles.

component  $x_i$ . Since only the potential energy depends on  $x_i$  the result is the negative of the force component  $F_{xi}$ ,

$$\frac{\partial E}{\partial r_{xi}} = \frac{\partial E_P}{\partial r_{xi}} = -F_{xi} = -\dot{p}_{xi}. \quad (6.5)$$

To obtain the result on the far right we used the fact that force is the time derivative of momentum. Next we calculate the derivative of  $E$  with respect to  $p_{ix}$ . Since  $p_{ix}$  appears only in one term of the kinetic energy, the result is

$$\frac{\partial E}{\partial p_{xi}} = \frac{p_{xi}}{m} = \dot{r}_{xi}. \quad (6.6)$$

Equations (6.5) and (6.6) give the time derivatives of all components of the state vector  $(p_1, p_2, \dots; r_1, r_2, \dots)$  and represent the equations of motion in Hamilton's form.

Hamilton's equations have a symmetry with respect to position and momentum: a position component and its corresponding momentum component form a conjugate pair such that the time derivative of the one component is given by the derivative of the total energy with respect to the other component (with a plus or minus sign). An additional relationship is obtained by taking the derivative of Eq. (6.5) with respect to  $p_{xi}$  and equating it to the derivative of Eq. (6.6) with respect to  $p_{xi}$ . We obtain,

$$\frac{\partial \dot{p}_{ix}}{\partial r_{ix}} + \frac{\partial \dot{r}_{ix}}{\partial p_{ix}} = 0, \quad (6.7)$$

and similarly for the components in the  $y$  and  $z$  direction. This provides us with yet another relationship between conjugate components. The importance of the Hamiltonian formulation is that it applies to mechanical particles with any degree of complexity. The particles in our formulation are point masses with three translational degrees of freedom each. For a dumbbell particle formed by two point particles connected with a rigid bond, in addition to the translational degrees of freedom we have two rotational degrees along the axis of the dumbbell and perpendicular to it. To specify the state of this particle we need the angular displacements and the corresponding angular momenta in addition to the coordinates and momentum components of the particle center of mass. It turns out that if we treat the angular displacements as generalized coordinates, these coordinates and their corresponding momenta satisfy Eqs. (6.5), (6.6) and (6.7). This remains true as we add complexity to the structure of the particles, for example, by allowing bonds to vibrate or by connecting more particles or chains of particles to the dumbbell. Regardless of the details of structure, the Hamiltonian is

$$\mathcal{H}(p; q) = \sum_i \frac{p_i^2}{2m} + E_P(p_1, p_2, \dots; q_1, q_2, \dots), \quad (6.8)$$

where  $i$  refers to degrees of freedom rather than number of particles,  $q_i$  is the generalized coordinate of the  $i$ th degree of freedom, and  $p_i$  is its conjugate momentum. In the convention of the Hamiltonian notation the energy is  $\mathcal{H}$  and the generalized coordinates are  $q$ ; the generalized momenta are still  $p$

as in the standard notation.<sup>6</sup> Here we will retain the original symbols  $E$  for energy and  $r$  for space coordinates since all of our examples in this book will be of unconnected point particles. The important point is that the Hamiltonian formulation is the basis for the study of mechanical systems with any number of degrees of freedom, regardless of the structural details of the constituent particles.

### 6.3 Microstate and the continuous phase space

The Hamiltonian formulation suggests that the natural set of variables to describe the microstate of a collection of classical particles is the set of space coordinates and momenta of all particles. We use the symbol  $\Gamma$  to represent this multidimensional vector,

$$\Gamma = (r_{x1}, r_{y1}, r_{z1}, \dots; p_{x1}, p_{y1}, p_{z1}, \dots).$$

In a system of  $N$  point particles, this vector contains  $6N$  components, 3 coordinate components and 3 momentum components per particle.  $N$  is typically of the order of Avogadro's number, which makes the dimensionality of this vector very high. The canonical phase space of a system with fixed volume  $V$  and number of particles  $N$  contains all microstates whose particle coordinates fall within the container of the system and whose momentum is anything between  $-\infty$  and  $+\infty$ . This is a space with dimensionality  $6N$ . Any point inside this space represents a possible microstate of an isothermal system. By analogy to 3-dimensional space where the volume enclosed in the region between  $(x, y, z)$  and  $(x + dx, y + dy, z + dz)$  is  $dV = dx dy dz$ , a volume element in phase space is

$$d\Gamma = dr_{x1} dr_{y1} dr_{z1} \dots dr_{p1} dr_{p1} dr_{p1} \dots$$

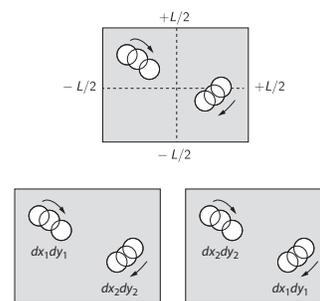
The volume element  $d\Gamma$  represents the number of microstates within the region  $(\Gamma, \Gamma + d\Gamma)$  of phase space. The dimensions of this product are (distance  $\times$  momentum) <sup>$3N$</sup>  but the number of microstates should be represented by a pure number. We can fix this by incorporating a dimensional constant  $c_N$  with dimensions (distance  $\times$  momentum) <sup>$-3N$</sup> . Luckily a universal physical constant already exists that has the correct dimensions: Planck's constant,

$$h = 6.626 \times 10^{34} \text{ m}^2 \text{ kg/s}.$$

Then, the number of microstates in volume element  $d\Gamma$  is

$$\text{number of microstates in } (\Gamma, \Gamma + d\Gamma) = \frac{d\Gamma}{h^{2N}}, \tag{6.9}$$

which is now dimensionless. We may think of  $h^{2N}$  as the elementary size of microstates in a mechanical system with  $N$  particles. It is not a coincidence that  $h$  represents the natural unit of particle microstate in the continuum phase space of classical mechanics. In quantum mechanics, where microstates are discrete and can be enumerated directly, Planck's constant appears naturally. Since the continuity of classical mechanics is only an "illusion" that sets in when the system becomes too large relative to the discrete separation of microstates,



**Figure 6.2:** (Top) Microstates of two particles in the vicinity of  $(x_1, y_1)$  and  $(x_2, y_2)$  (for implicitly the space is two dimensional and the momentum is not shown). (Bottom) When all microstates in the box are integrated with respect to all coordinates from  $-L/2$  to  $+L/2$ , the microstate at the top (and every other microstate) will be counted twice, once with the coordinates  $(x_1, y_1)$  referring to the particle at the top left part of the box, and once with  $(x_1, y_1)$  referring to the coordinates of the in the bottom right. This counting is appropriate for *distinguishable* particles. If the particles are *indistinguishable*, this over counts by a factor of  $2!$ .

Planck's constant continues to be present in the final result. The important point is that despite the continuity of classical phase space we may still talk of the "number of microstates." This number is dimensionless and so large that we will treat it as a continuous variable.

#### 6.4 Probability of microstate and Liouville's theorem

Microstates are functions of time, they travel in phase space as the coordinates and momenta of the particles change under the laws of motion. The movement of a microstate in phase space is analogous to the motion of a fluid in a three dimensional space – a good analogy that helps us visualize this space of high dimensionality. Even though the macroscopic state of an equilibrium system is invariant in time, the microscopic state is in continuous flux. It is not possible to tell, based on macroscopic observations alone, which microstate is present at any given time  $t$ . We may only speak of the *probability* to find the system at some microstate. Since microstates are continuous, we use the probability density,  $\rho(\Gamma)$ , to represent the probability of microstate. The probability density is such that

$$\rho(\Gamma)d\Gamma$$

is the probability to find the microstate in the region of phase space between  $\Gamma$  and  $\Gamma + d\Gamma$ . A way to visualize this probability is via an ensemble of macroscopic systems. We imagine a very large number of systems with identical macroscopic properties (e.g., fixed temperature, volume, number of particles). If we take a microscopic snapshot in time we will observe a variety of microstates. If the ensemble is large enough, then we expect we would see every microstate that can possibly exist and that every microstate would be present in proportion to its probability. In this ensemble, the frequency of a microstate (the fraction of systems that are found in that microstate) is equal to the probability. This notion of an ensemble of microstates that represent a single macroscopic state is due to Gibbs. If the ensemble contains  $\mathcal{N}$  systems, then the number of systems whose microstate is in the vicinity of  $\Gamma$  is

$$d\mathcal{N} = (\mathcal{N}\rho(\Gamma))d\Gamma$$

This is completely analogous to the relationship between material density, mass and volume.<sup>7</sup>

A snapshot of the ensemble gives us the instantaneous collection of every microstate that can exist under the macroscopic specification of state, and its probability. If we turn the snapshot into a movie, we will observe continuous motion as every microstate evolves under the laws of motion, like a fluid that flows through phase space. The fluid analogy is actually very apt. Since  $\rho$  gives the density of microstates in phase space (proportional to the number of systems with microstates near  $\Gamma$ ) and the number of systems is fixed during the movie, we may write a continuity equation on the probability density  $\rho$ . This equation is known as the Liouville equation and is derived next.

We will obtain the governing equation of  $\rho$  by complete analogy to the continuity equation of a solute with local concentration  $c(x, y, z)$ . We will first develop the equation for solute in three dimensions and then we will extend the result

<sup>7</sup>If  $c(x, y, z)$  is the local concentration of a solute in a fluid at  $(x, y, z)$ , the mass of solute in a volume element at that point is

$$dm = c(x, y, z)dxdydz = c(x, y, z)dV.$$

$\mathcal{N}$  is analogous to mass,  $\mathcal{N}\rho(\Gamma)$  analogous to concentration  $c(x, y, z)$  and  $\rho(\Gamma)$  analogous to concentration normalized by the total amount of solute in the system.

to a phase space with  $6N$  dimensions. We consider a volume element  $dx dy dz$  and the transport of solute in and out by convection of the fluid whose velocity at  $(x, y, z)$  is  $(\dot{x}, \dot{y}, \dot{z})$ . The flux in the  $x$  direction is  $v_x c$ , where  $v_x = \dot{x}$  is the velocity component in the  $x$  direction. The net flux within a differential element  $(x, x + dx)$  is

$$(\dot{x}c|_x - \dot{x}c|_{x+dx}) dx dy = -d(\dot{x}c) dx dy$$

and the continuity equation reads

$$\frac{d}{dt}(c dx dy dz) = -d(\dot{x}c) dx dy$$

or finally,

$$\frac{\partial c}{\partial t} = -\frac{\partial(\dot{x}c)}{\partial x}. \quad (6.10)$$

This generalizes readily to any number of dimensions. To do this, we apply Eq. (6.10) with  $c \rightarrow \rho$ , and with  $x$  replaced all position and momentum coordinates. It will be simpler to temporarily switch to the convention that the subscript  $i$  in  $p_i$  and  $r_i$  refers to the  $i$  degree of freedom, rather than to particle  $i$ . In this nomenclature the result is

$$\frac{\partial \rho}{\partial t} = -\sum_{i=1}^{3N} \left( \frac{\partial(\dot{p}_i \rho)}{\partial p_i} + \frac{\partial(\dot{r}_i \rho)}{\partial r_i} \right),$$

with the summation going over all degrees of freedom from  $i = 1$  to  $3N$ , since we have 3  $p$ 's and 3  $r$ 's per particle. Expanding the derivatives inside the summation and using Eq. (6.7), the previous result becomes

$$\frac{\partial \rho}{\partial t} = -\sum_i \left( \dot{p}_i \frac{\partial \rho}{\partial p_i} + \dot{r}_i \frac{\partial \rho}{\partial r_i} \right). \quad (6.11)$$

This equation expresses Liouville's theorem and governs the motion of the ensemble of microstates through phase space. It may be expressed in the alternative form<sup>8</sup>

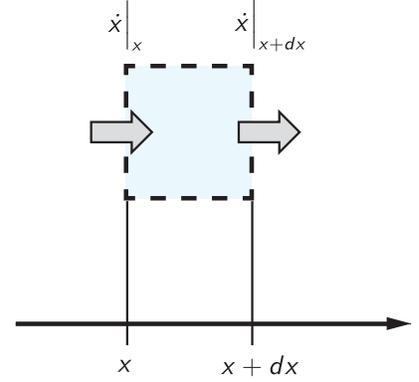
$$\frac{D\rho}{Dt} \equiv \frac{\partial \rho}{\partial t} + \sum_i \left( \dot{p}_i \frac{\partial \rho}{\partial p_i} + \dot{r}_i \frac{\partial \rho}{\partial r_i} \right) = 0. \quad (6.12)$$

Here,  $D\rho/Dt$  is the total derivative and gives the change of  $\rho$  along a volume element that follows the flow. The fact that the total derivative is zero means that the ensemble moves as an incompressible fluid.<sup>9</sup>

**Statistical equilibrium** The system is in *statistical equilibrium* if the probability density at fixed  $\Gamma$  is independent of time. Setting the time derivative in Eq. (6.11) equal to zero we obtain

$$\sum_{i=1}^{3N} \left( \dot{p}_i \frac{\partial \rho}{\partial p_i} + \dot{r}_i \frac{\partial \rho}{\partial r_i} \right) = 0. \quad (6.13)$$

This is the stationary Liouville equation and represents a necessary (though not sufficient) condition for the equilibrium probability distribution of microstate.



**Figure 6.3:** Geometry for the calculation of flux of solute or microstates along the  $x$  axis of a volume element in real space or phase space.

<sup>8</sup>The term  $\partial \rho / \partial t$  represents the rate of change at fixed point in phase space. The term  $\dot{p}_i \partial \rho / \partial p_i$  results from the motion in phase space by  $dp_i$  in time  $dt$ . A similar contribution arises from all  $q_i$ . The sum of all such contributions is the total derivative of  $\rho$  and represents the derivative along the trajectory of a point in phase space.

<sup>9</sup>The continuity equation of an incompressible fluid is

$$v_x \frac{\partial \rho}{\partial x} + v_y \frac{\partial \rho}{\partial y} + v_z \frac{\partial \rho}{\partial z} = 0.$$

While an incompressible fluid has the same density throughout space, our ensemble has different densities at different points of phase space (see Eq. 6.14 for example). Nonetheless, along the path of a volume element in phase space, the probability density remains constant.

**Special solutions to the Liouville equation** If a function  $\rho(\Gamma)$  is to represent the probability density of microstates, it must satisfy the stationary Liouville equation. Let us consider a probability density that is an explicit function of energy, of the form  $\rho = \rho(E)$ . Since energy is a function of microstate, this function is ultimately a function of microstate and a possible candidate for the unknown probability of microstate. But does it satisfy the stationary Liouville equation? The answer is yes, as we can easily show. Inserting  $\rho(E)$  into Eq. (6.13) and using the chain rule to calculate derivatives with respect to  $p_i$  and  $r_i$  we have

$$\sum_{i=1}^{3N} \left( \dot{p}_i \frac{\partial \rho}{\partial E} \frac{\partial E}{\partial p_i} + \dot{r}_i \frac{\partial \rho}{\partial E} \frac{\partial E}{\partial r_i} \right) = \frac{\partial \rho}{\partial E} \sum_{i=1}^{3N} \left( \dot{p}_i \frac{\partial E}{\partial p_i} + \dot{r}_i \frac{\partial E}{\partial r_i} \right) = 0,$$

whose right-hand side is zero by virtue of Eq. (6.7). The stationary Liouville equation is indeed satisfied, therefore distributions of the form  $\rho = \rho(E)$  are admissible as possible probability distributions of continuous mechanical microstates. This is an important conclusion. In chapter 4, as well the closely related microcanonical and grand canonical distributions are explicit functions of energy, they do satisfy the stationary Liouville equation and can be applied to the continuous phase space of mechanical particles that interact according to the laws of classical mechanics. In other words, the ensemble theory developed in chapter 4 for unspecified discrete microstates applies entirely to mechanical systems provided that we interpret microstate to refer to a vector that gives the positions and momenta components of all particles. What is left to do is adapt the discrete probability distributions to the continuous phase space of a mechanical system that contains  $N$  classical particles.

**Necessary and sufficient conditions for the probability distribution  $\rho$**  The stationary Liouville equation is a necessary condition for any probability density  $\rho$  that is proposed as a possible probability for a physical system of classical particles that obeys the familiar laws of mechanics. It is not sufficient since, as we have found out, *any* probability density that is a function of energy satisfies the stationary Liouville equation. This means that any of the functions below satisfy that equation (there are many more):

$$\rho_1 = ae^{-bE}, \quad \rho_2 = ae^{-(E-b)^2/c}, \quad \rho_3 = aE^b e^{-(cE)^d} \dots$$

where  $a, b, \dots$  are appropriate constants. Any of them could be suggested as probability distributions of an isothermal system, as long as the constants are selected to give the desired value of  $\bar{E}$ . Yet, *only* the exponential function makes “thermodynamic sense,” in the sense that it is the *only* distribution that reproduces the correct thermodynamic relationships. In chapter 4 we developed an unambiguous methodology to identify the correct probability distribution out of all possible distributions. According to this methodology, we must choose the distribution that maximizes the entropy functional under any constraints that are placed by the macroscopic state. There is no need to repeat this maximization here, we may merely translate the discrete results of chapter 4 to the continuous phase space of classical mechanics (this is done in the next sections).



**Figure 6.4:** J. Willard Gibbs as a student at Yale University, which he attended from 1854 to 1858.

**How Gibbs did it** Scientific discovery is hardly ever a straight line from observation to theory and statistical mechanics is no exception. It is very instructive to review the steps by which Gibbs arrived at the results we are discussing in this chapter. Gibbs's premise is that the thermodynamic properties of a mechanical system can be calculated as an average over every possible microstate that the system may adopt. This poses two new questions. First, how to define the microstate of the system; and second, how to obtain the correct probability distribution. By setting up the problem in terms of Hamilton's equations, it becomes clear that the natural way to describe the mechanical microstate of  $N$  particles is via the space coordinates and momenta of the particles, *not* velocities.<sup>10</sup> The Liouville equation led Gibbs to realize that the unknown probability distribution of microstate could be just a function of energy. This represents an enormous simplification –indeed, a breakthrough– since it narrows significantly the scope of the search for the unknown distribution, from distributions of  $6N$  variables, to distributions with one independent variable. Even so, without a method to sort out among the infinite number of distributions that are functions of energy, the task of finding the right one remains daunting. Gibbs's solution was to *guess* the answer. His first guess is the exponential distribution, which he calls canonical. He then goes on to show that this distribution is in full agreement with every known thermodynamic result,<sup>11</sup> discovers that entropy can be expressed by the now familiar functional,<sup>12</sup>

$$- \int \rho (\log \rho) d\Gamma$$

and shows that the canonical probability distributions maximizes this functional among all other possible distributions. The maximization of the entropy functional provides a statistical interpretation of the second law, but more significantly, it shows that this maximization *identifies* the correct distribution among all possible distributions.<sup>13</sup> Indeed our approach in chapter 4 was to start with the maximization of the entropy functional as the founding postulate of statistical thermodynamics. In this approach we obtain the same distributions as Gibbs without the need to guess, and without the need to deal with the messy calculus of a continuous, highly dimensional phase space.

## 6.5 Canonical distribution in continuous phase space

The canonical probability distribution we obtained in chapter 5 depends solely on the energy of the system, it is therefore an acceptable solution to the Liouville equation. Moreover, we showed in chapter 5 that this probability distribution reproduces, via the canonical partition function, all known thermodynamics. It is therefore a thermodynamically consistent probability distribution. We express the canonical distribution in the continuous phase space of  $N$  classical particles in the form

$$\rho_C(\Gamma) = \frac{e^{-\beta E(\Gamma)}}{h^{3N} Q(\beta, V, N)} \quad (6.14)$$

Here,  $E(\Gamma)$  is the energy of microstate  $\gamma$ ,  $Q(\beta, V, N)$  is the canonical partition function (dimensionless) and  $h$  is Planck's constant. This probability distribution

<sup>10</sup>The distinction may seem irrelevant but it is not. In a system of two types of particles, the region of phase space defined by  $\delta x \delta v$  contains *more* microstates of the heavier particle than of the lighter one. If we used velocity to indicate the microstate we would be counting microstates incorrectly. This distinction is even more important when we consider additional degrees of freedom such as rotation or vibration. These degrees are also represented in phase space by their generalized momenta and coordinates.

<sup>11</sup>Gibbs develops the canonical probability in chapter IV of his book, then takes another five chapters to prove the full compatibility of this distribution with all known thermodynamics (Gibbs, 1902).

<sup>12</sup>For dimensional consistency  $\log \rho$  should be replaced with  $\log(\rho/h^{3N})$ . This dimensional constant has no effect on the maximization.

<sup>13</sup>Gibbs did not express this conclusion in such stark terms, rather he viewed it as a corollary of his theory. It was Jaynes who suggested, several decades later, that the constrained maximization of the entropy functional may be viewed as the principle for the determination of any unknown probability distribution, whether in statistical mechanics or in any other situation (Jaynes, 1957).

satisfies the normalization condition

$$\int_{\text{phase space}} \frac{e^{-\beta E(\Gamma)}}{h^{3N} Q(\beta, V, N)} d\Gamma = 1,$$

with the integral taken over all phase space. Since  $h^{3N}$  has dimensions of  $\Gamma$ , this integral is dimensionless, as it should. Solving the normalization condition for  $Q$  we obtain the partition function as an integral of the Boltzmann factor over all microstates:<sup>14</sup>

$$Q(\beta, V, N) = \int_{\text{phase space}} e^{-\beta E(\Gamma)} \frac{d\Gamma}{h^{3N}} \quad (6.15)$$

<sup>14</sup>By pairing  $\Gamma$  with  $h^{3N}$  we emphasize the dimensional homogeneity of this equation. Notice that the ratio  $d\Gamma/h^{3N}$  is the number of microstates in  $d\Gamma$ .

We have determined that the log of the canonical partition function is  $-A/kT$  and that  $\beta = 1/kT$ . The above result can then be expressed as

$$e^{-A/kT} = \frac{1}{h^{3N}} \int e^{-E/kT} dr_{1x} dr_{1y} \cdots dp_{x1}, \cdots \quad (6.16)$$

This is one of the most important results of statistical mechanics. Its significance lies in the fact that it connects the macroscopic world (on the left-hand side) with the microscopic world (right-hand side). Equally significant is the elegance of this result, which is free of the chaotic and largely unresolvable details of motion of an astronomically enormous number of particles that interact with each other.

As in the discrete case, the mean energy is obtained from the derivative of the partition function with respect to  $\beta$ :

$$\bar{E} = - \left( \frac{\partial \log Q}{\partial \beta} \right)_{VN}$$

This is easy to confirm starting with the definition of the mean energy in combination with the exponential distribution in Eq. (6.14).

## 6.6 Microcanonical ensemble

The number of microstates in phase space is a function of energy – the more energy there is, the more microstates can be constructed. Using the two-dimensional analogy, all microstates with the same energy form a “contour” in phase space. If we imagine the phase space as a three dimensional space, this contour is a surface; its area is a measure of the number of microstates at that energy. If energy is increased we have many more microstates and so contours of higher energy move “outward” and enclose contours of lower energy. The number of microstates is represented by “volume” but since microstates of fixed energy form a *surface* and a surface has no volume, we cannot speak of the number of microstates at energy  $E$ . Instead, we speak of the number of microstates in a thin shell of thickness  $\delta E$  around contour  $E$  (see Fig. 6.5). This number is

$$\omega(E, V, N) \delta E = \int_{(E, E+\delta E)} c_N d\Gamma. \quad (6.17)$$

where  $\omega$  is the *density* of microstates per unit energy. The integral on the right-hand side gives the volume of phase space in the thin shell  $(E, E + dE)$ . The microcanonical ensemble consists of all microstates inside the energy shell  $(E, E + dE)$ . All microstates inside this shell have the same energy, and according to Eq. (6.14), the same probability. This probability is

$$\rho_{\mu C}(\Gamma) = \frac{c_N}{\omega(E, V, N)\delta\epsilon} \quad (6.18)$$

for microstates with energy in  $(E, E + \delta\epsilon)$ , and zero otherwise. Using Eq. (6.17) we confirm that the distribution satisfies the condition,

$$\int_{(E, E+\delta\epsilon)} \omega(E, V, N) d\Gamma = 1,$$

it is therefore properly normalized. Here,  $\delta\epsilon$  is an arbitrary but fixed value of energy that endows a finite volume to the energy contour. The presence of  $\delta\epsilon$  is necessary for dimensional homogeneity but has no effect on thermodynamics. To see why, suppose that  $F(\Gamma)$  is an extensive property of the microstate. Its average over the microcanonical ensemble is

$$\bar{F} = \int F(\Gamma) \rho_{\mu C}(\Gamma) d\Gamma = \frac{c_N}{\omega(E)\delta\epsilon} \int_{(E, E+\delta\epsilon)} F(\Gamma) d\Gamma. \quad (6.19)$$

As long as the shell is thin, all extensive quantities within the shell are proportional to its thickness  $\delta\epsilon$ . The integral on the far right is then proportional to  $\delta\epsilon$  and this cancels the  $\delta\epsilon$  in the denominator. This makes  $\bar{F}$  independent of the thickness of the shell and provides the justification for the arbitrary selection of  $\delta\epsilon$ .

### 6.6.1 Microcanonical entropy

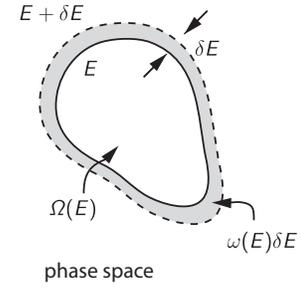
To connect the density of microstates to thermodynamics we return to Eq. (6.15), which gives the canonical partition function as an integral in phase space, and rearrange the integration to go across energy contours. The number of microstates on contour  $E$  is  $\omega(E, V, N)dE$  and all of these contribute the factor  $e^{-\beta E}$  to the summation. The result is

$$Q(\beta, V, N) = \int_E \omega(E, V, N) e^{-\beta E} dE.$$

This is analogous to Eq. (5.37) that we obtained in the discrete phase space. As in the discrete case, the product  $\omega(E, V, N)e^{-\beta E}$  is a narrowly peaked function that asymptotically becomes Gaussian function,

$$\omega(E, V, N) e^{-\beta E} \rightarrow \omega(\bar{E}, V, N) e^{-\beta \bar{E}} \exp\left\{-\frac{(E - \bar{E})^2}{2\sigma^2}\right\}$$

where  $E_*$  is the energy that maximizes  $\omega(E, V, N)e^{-\beta E}$ , also equal to the mean energy of the canonical ensemble, and  $\sigma^2 = kT^2 C_V$ . Using this asymptotic form



**Figure 6.5:**  $\Omega(E)$  is the the number of microstates inside energy contour  $E$ ;  $\omega(E)$  is the density of microstates per unit energy on contour  $E$ .  $\omega$  is to  $\Omega$  what “density” is to “volume.”

the canonical partition function becomes

$$\begin{aligned} Q(\beta, V, N) &= \int_E \omega(E, V, N) e^{-\beta E} dE \\ &\rightarrow \omega(\bar{E}, V, N) e^{-\beta \bar{E}} \int_{-\infty}^{\infty} \exp\left\{-\frac{(E - \bar{E})^2}{2\sigma^2}\right\} dE \\ &= \omega(\bar{E}, V, N) e^{-\beta \bar{E}} \sqrt{2\pi k T^2 C_V}. \end{aligned}$$

For the log of the partition function we now have

$$\log Q(\beta, V, N) = \log \omega(\bar{E}, V, N) - \beta \bar{E} + \log \sqrt{2\pi k T^2 C_V}.$$

The quantity  $\sqrt{2\pi k T^2 C_V}$  has dimensions of energy and scales as  $\log N$ . This quantity qualifies as a differential  $\delta\epsilon$  relative to  $\bar{E}$ , which scales as  $\sim N$ . We may rewrite the above result in the form

$$\log Q(\beta, V, N) = \log(\omega(\bar{E}, V, N)\delta\epsilon) - \beta \bar{E} \quad (6.20)$$

where

$$\delta\epsilon = \sqrt{2\pi k T^2 C_V}.$$

This result is entirely analogous to Eq. (5.37) and establishes the relationship between the canonical and the microcanonical ensemble. Since the mean energy  $\bar{E}$  is the negative of the derivative of  $\log Q$  with respect to  $\beta$ , Eq. (6.20) identifies  $\log \omega \delta\epsilon$  as the Legendre transform of  $\log Q$  with respect to  $\beta$ . But since  $\log Q = -A/kT$ , the Legendre transform in question is  $S/k$ , we finally have

$$\frac{S}{k} = \log(\omega(\bar{E}, V, N)\delta\epsilon). \quad (6.21)$$

This identifies  $\log \omega(E, V, N)\delta\epsilon$  as entropy (divided by  $k$ ) and  $\omega(E, V, N)\delta\epsilon$  as the microcanonical partition function. The result contains the bothersome constant  $\delta\epsilon$  but as our derivation shows, the log of  $\delta\epsilon$  grows logarithmically with the size of the system so that, numerically at least, we have

$$\log \omega(E, V, N) + \log \delta\epsilon \approx \log \omega(E, V, N).$$

For all practical purposes we may throw this constant away. Nonetheless, it is not good practice to take the log of a dimensional quantity, even if the final result is numerically correct.<sup>15</sup> We will leave  $\delta\epsilon$  in our equations with the understanding that it has no effect whatsoever on any thermodynamic property.

### 6.6.2 An alternative expression for the microcanonical entropy

It is possible to provide a relationship for the entropy of the microcanonical ensemble that does not involve  $\delta\epsilon$ . Let  $\Omega(E, V, N)$  be the number of microstates enclosed by energy  $E$ , this includes all microstates with any energy less than or equal to  $E$ . This number is obtain this by integrating the phase space from  $\infty$  energy up to energy  $E$ :<sup>16</sup>

<sup>15</sup>You should learn to be bothered a lot more by the *lack* of dimensional homogeneity than by the presence of any extraneous variables that are needed to maintain proper homogeneity.

$$\Omega(E, V, N) = \int_{(-\infty, E)} c_N d\Gamma.$$

We can easily see that the relationship between  $\omega$  and  $\Omega$  is

$$\omega(E, V, N) = \left( \frac{\partial \Omega(E, V, N)}{\partial E} \right)_{V, N}. \quad (6.22)$$

Notice that unlike  $\omega$ ,  $\Omega$  is dimensionless. It turns out that  $\Omega$  increases rapidly with the size of the system and scales as<sup>17</sup>

$$\Omega(E, V, N) \sim E^N. \quad (6.23)$$

Using this scaling, for  $\omega(E, V, N)$  we find

$$\omega(E, V, N) = \frac{\partial \Omega(E, V, N)}{\partial E} = aNE^{N-1} = a \frac{N\Omega(E, V, N)}{E}.$$

where  $a$  is some constant that makes the scaling an exact equality. Combining with Eq. (6.21) we have

$$\frac{S}{k} = \log(\omega(E, V, N)\delta\epsilon) = \log\Omega(E, V, N) + \log\left(\frac{aN}{E}\right).$$

The term  $\log(aN/E)$  on the far right grows logarithmically with the size of the system and may be dropped relative to  $\log\Omega$ , leading to the final result,

$$\frac{S}{k} = \log(\omega(E, V, N)\delta\epsilon) = \log\Omega(E, V, N). \quad (6.24)$$

This says that the entropy of the microcanonical ensemble is equal to the log of the number of microstates with energy  $(E, E + \delta\epsilon)$ , also equal to the cumulative number of microstates from the ground state up to the energy of interest.<sup>18</sup> The result may seem counterintuitive since there are clearly many more microstates at all energies up to  $E$  compared to microstates at just  $E$ . The important point here is that entropy refers to the *logarithm* of these numbers. While  $\Omega \gg \omega\delta\epsilon$ , the difference between their logarithms is small relative to the logarithms themselves.<sup>19</sup>

## 6.7 Grand canonical ensemble

To complete the development of ensembles in statistical mechanics we formulate the probability distribution in the grand canonical ensemble. The microstates of the grand canonical ensemble consist of all canonical microstates that can be constructed with *any* number of particles  $N$  placed in volume  $V$ . The microstate is specified by position and momentum components of all particles, but since  $N$  is now allowed to vary, we will denote this vector by  $\Gamma_N$  to indicate the number of particles it contains. The grand canonical probability then is

$$\rho_{gC}(\Gamma_N) = \frac{e^{-\beta E - \gamma N}}{h^{3N} \Xi(\beta, V, \gamma)}$$

<sup>17</sup>We will not provide proof of this scaling. In section 6.8.4 where we discuss a system of  $N$  harmonic oscillators we will confirm this scaling.

<sup>18</sup>Gibbs identified the entropy of the microcanonical ensemble with the log of the cumulative number of microstates,  $\log\Omega$ , not with  $\log\omega\delta\epsilon$  (see Problem 6.6).

<sup>19</sup>Logarithms have an amazing ability to equalize quantities that in absolute scale are enormously different from each other. Compare, for example,  $10^{1000}$  and  $10^{1010}$ . The two numbers differ by a factor of  $10^{10}$  but their logarithms differ only by 1%.

where  $E = E(\Gamma_N)$  is the energy of the microstate and  $N$  is the number of particles it contains. The normalization condition is expressed by an integral over all  $\Gamma_N$ , tucked inside a summation over all  $N$ :

$$\sum_{N=0}^{\infty} \int_{\Gamma_N} \frac{e^{-\beta E - \gamma N}}{\Xi(\beta, V, \gamma)} \frac{d\Gamma_N}{h^{3N}} = 1,$$

which, solved for  $\Xi$ , provides the defining equation for the grand canonical partition function:

$$\Xi(\beta, V, \gamma) = \sum_{N=0}^{\infty} \int_{\Gamma_N} e^{-\beta E - \gamma N} \frac{d\Gamma_N}{h^{3N}}. \quad (6.25)$$

Here, the integration is over the phase space of  $N$  particles with  $N = 0, 1, \dots$ . The integration at fixed  $N$  gives the canonical partition function (see Eq. (6.15)), therefore an alternative expression for the grand canonical partition function is

$$\Xi(\beta, V, \gamma) = \sum_N e^{-\gamma N} Q(\beta, V, N).$$

We obtained the same result in the discrete analysis. Of course, this is because with respect to the number of particles the phase space is discrete even when we are dealing with classical particles.

## 6.8 The linear harmonic oscillator

The concepts discussed here are difficult to visualize, in large part because of the high dimensionality of phase space. To help formulate a mental picture,<sup>20</sup> the examples below demonstrate some of these concepts using one of the simplest possible examples from mechanics, the linear harmonic oscillator, whose phase space is two dimensional (one position component, one momentum component) and can be represented on a plane. We discuss this problem in the following sections and calculate various ensemble properties to demonstrate how statistical mechanics works in practice.

<sup>20</sup>Gibbs's book (Gibbs, 1902) is 207 pages of straight text and equations without a single drawing. Gibbs builds the complete mental picture of the problem inside his head and challenges the reader to do the same.

### 6.8.1 Phase space

A mass  $m$  attached to a linear spring with constant  $k$  oscillates along the  $x$  axis about its equilibrium position at  $x = 0$ . This simple system consists of a single particle that moves inside a one-dimensional harmonic potential  $E_p(x) = kx^2/2$  that produces a force  $F_x = -kx$ . The microstate of this system is defined by the position  $x$  and momentum  $p$  of the oscillator. This produces the two-dimensional phase space  $(x, p)$ , which is represented graphically by a plane with axes  $x$  and  $p$ . Any point on this plane is a possible microstate: it specifies the instantaneous velocity and position of the point mass. If we are not given the initial conditions, the oscillator may be anywhere on this plane. Simply put, the phase space is the set of all possible microstates of the system.

The oscillator moves in phase space according to the equations of motion, which here we combine into a second order differential equation for the position of the mass,

$$m\ddot{x} = -kx.$$

Its solution is

$$x(t) = x_0 \cos(\alpha t + \phi_0),$$

with  $\alpha = \sqrt{k/m}$ ;  $x_0$  and  $\phi_0$  are integration constants.<sup>21</sup> The momentum is

$$p(t) = m\dot{x}(t) = -m\alpha x_0 \sin(\alpha t + \phi_0),$$

and involves the same integration constants as  $x$ . The total energy of the oscillator is

$$E = \frac{p^2}{2m} + \frac{kx^2}{2} = \frac{kx_0^2}{2}$$

and is constant. This equation defines an ellipse on the  $(p, x)$  plane that represents a contour of fixed energy. The oscillator moves clockwise on the perimeter of this ellipse and its motion can be visualized by making a parametric plot of  $p(t)$  against  $x(t)$  using  $t$  as a parameter.

### 6.8.2 Microcanonical oscillator

The calculation of the density of microstates  $\omega$  requires an integration at the edge of a contour. First we express the components of phase space in polar coordinates,  $x = r \cos\theta$ ,  $p = r \sin\theta$ , where  $r$  is the distance from the origin to point  $(x, p)$ , and  $\theta$  is the polar angle from the  $x$  axis. The equation of the contour in polar coordinates is,

$$\frac{r^2 \sin^2 \theta}{2m} + \frac{kr^2 \cos^2 \theta}{2} = E.$$

The radial distance  $r$  as a function of  $\theta$  is obtained by solving this equation for  $r$ ,

$$r = \frac{(2mE)^{1/2}}{\sin^2 \theta + mk \cos^2 \theta} = A(\theta)E^{1/2}, \quad (6.26)$$

where  $A$  contains everything except  $E$  in the above equation. We now want to calculate the number of microstates  $\omega(E)$  on a contour of fixed energy  $e$ . "Number of microstates" refers to an *area* in phase space but the microcanonical phase is a *line*. To calculate the number of microstates we must imagine a thin area around the contour of interest with thickness  $r dr d\theta$  in polar coordinates. The number of microstates in this thin area is  $\omega(E)dE$  and is equal to the area of the phase space:

$$\omega(E)dE = c \int_0^{2\pi} r dr d\theta,$$

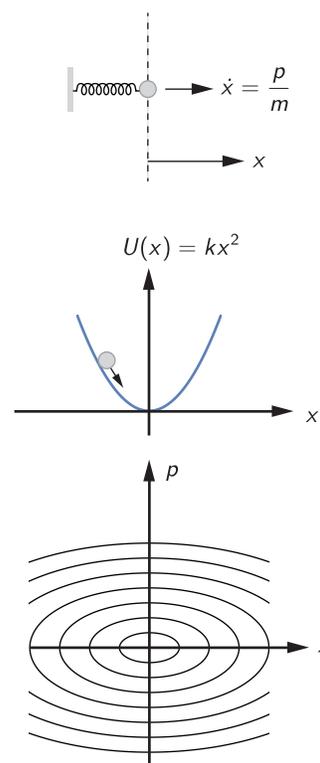
with  $r$  given from Eq. (1) as a function of  $\theta$ . Substituting the expression for  $r$  we have,

$$\omega(E)dE = c dE \int_0^{2\pi} (A(\theta)E^{1/2}) \left( \frac{1}{2} E^{-1/2} A(\theta) \right) d\theta,$$

and finally

$$\omega(E) = c \int_0^{2\pi} \frac{A^2(\theta)}{2} d\theta = \omega_0. \quad (6.27)$$

<sup>21</sup>The constant  $x_0$  is the maximum distance of the oscillator from the origin and is fixed by the total energy;  $\phi_0$  is fixed by the position of the oscillator at  $t = 0$ .



**Figure 6.6:** Phase space of linear harmonic oscillator showing contours of constant energy.

Everything on the right-hand side is constant, therefore we conclude that the number of microstates on a contour of energy  $E$  is independent of  $E$ .

The fact that the number of microstates around a contour is independent of energy may sound counter intuitive when one considers the fact that the contour perimeter increases with energy. To understand this result, recall that we are calculating the number of microstates in a thin region of phase space around a contour. The perimeter of the contour scales as  $r \sim E^{1/2}$  and increases with  $E$  but its thickness scales as  $dr \sim E^{-1/2}$  and decreases with energy. The two effects cancel exactly so that the number of microstates in the vicinity of the energy contour is the same for all contours. This is a special result because we are dealing with a single oscillator. When we consider many oscillators, as we do in section 6.8.4, we find that  $\omega$  increases very rapidly with energy.

### 6.8.3 Canonical oscillator

The microcanonical oscillator is constrained to be on a single energy contour of phase space. Suppose now we allow the oscillator to be anywhere in phase space with probability proportional to  $e^{-\beta E}$ . This is equivalent to allowing the oscillator to exchange energy with a “bath” at temperature  $1/\beta$ . There is no physical bath in this problem, what we have instead is a stochastic process that adds or removes energy from the oscillator. By making this probability exponential to the energy of the microstate we produce behavior identical to that of a thermodynamic system in equilibrium with a bath.<sup>22</sup>

**Canonical probability** The microstates of the oscillator in this problem constitute a canonical ensemble. Accordingly, the probability to find the oscillator in the vicinity of microstate  $(p, x)$  is

$$\rho_c(p, x) = \mathcal{C} e^{-\beta E(p, x)}$$

where  $E$  is the energy of the oscillator,

$$E(p, x) = \frac{p^2}{2m} + \frac{kx^2}{2}.$$

The constant  $\mathcal{C}$  is obtained from the normalization condition,

$$\frac{1}{\mathcal{C}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dp dx e^{-\beta(p^2/2m + kx^2/2)} = \int_{-\infty}^{\infty} dp e^{-\beta p^2/2m} \int_{-\infty}^{\infty} dx e^{-\beta kx^2/2}.$$

The integrals are elementary and we find,

$$\mathcal{C} = \frac{2\pi\sqrt{m/k}}{\beta}.$$

For the canonical probability density we finally obtain

$$\rho_c(x, p) = \frac{\beta}{2\pi\sqrt{m/k}} \exp\left(-\frac{\beta p^2}{2m} - \frac{\beta kx^2}{2}\right). \quad (6.28)$$

<sup>22</sup>There is an alternative reading of this problem that takes the oscillator to be at *any* possible microstate as long as the total energy of the system is satisfied. In this view there is no bath or physical exchange of energy, only complete uncertainty about the precise state of each oscillator beyond its mean energy.

The unknown parameter  $\beta$  is evaluated from the mean energy:

$$\bar{E} = \frac{\beta}{2\pi\sqrt{m/k}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( -\frac{p^2}{2m} - \frac{kx^2}{2} \right) \exp\left( -\frac{\beta p^2}{2m} - \frac{\beta kx^2}{2} \right) dp dx.$$

This integral is also evaluated easily by separating the  $p$  and  $x$  integrals and the final result is

$$\bar{E} = 1/\beta.$$

The second moment is

$$\overline{E^2} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E^2 \rho(p, x) dp dx = 2/\beta^2,$$

and the variance follows as

$$\text{var}(E) = \overline{E^2} - \bar{E}^2 = 1/\beta^2.$$

The typical energy fluctuation is  $\delta E = \text{var}(E)^{1/2} = 1/\beta$ , the same as the mean energy.

**Canonical partition function** The canonical partition function is obtained from the inverse of the normalization constant  $\mathcal{C}$ ,

$$Q = \frac{c}{\mathcal{C}} = c \left( \frac{2\pi}{\beta} \sqrt{m/k} \right), \quad (6.29)$$

where  $c = 1/h^2$  ( $h$  is Planck's constant) is a dimensional constant that makes  $Q$  dimensionless (we will retain this as an unevaluated factor). The mean energy of the canonical ensemble is (see Eq. 5.30). Applying this to the canonical partition function of the oscillator we find

$$\bar{E} = -\frac{\partial \log Q}{\partial \beta} = -\frac{\partial}{\partial \beta} \log \left( \frac{1}{\beta} \right) = \frac{1}{\beta}.$$

This is the same result that was obtained by direct integration of the probability density  $\rho_c$  but now the calculation involves a simple differentiation rather than double integration.

**Energy distribution** We define the energy distribution  $f(E)$  such that  $f(E)dE$  is the probability to find the oscillator in a microstate with energy  $E$ . This is closely related to the density of microstates  $\omega(E)$ . The number of microstates with energy  $E$  is  $\omega(E)dE = \omega_0 E$  and the probability of microstate is proportional to  $\exp(-\beta E)$ ; therefore, the energy distribution is proportional to the product of the two,

$$f(E) = C\omega_0 e^{-\beta E} = C' e^{-\beta E}.$$

where  $C' = C\omega_0$  combines all proportionality factors. We obtain  $C'$  by the normalization condition. Integrating over all  $E$  from 0 to  $\infty$  we find  $C' = \beta$  we find,

$$f(E) = \beta e^{-\beta E}. \quad (6.30)$$

The mean and variance of this distribution are obtained easily and the result is

$$\bar{E} = 1/\beta, \quad \overline{E^2} = 1/\beta^2.$$

These are exactly the same as those obtained from the probability density of microstates,  $\rho(p, x)$ . This should not be surprising because both distributions cover the same phase space but in different ways; integrals involving  $\rho$  are done in the two dimensional plane  $(p, x)$  whereas integrals involving  $E$  are conducted by moving outward from one energy contour to the next.

The exponential distribution is very broad: it peaks at the lowest possible energy ( $E = 0$ ) and its standard deviation is equal to the mean. This behavior is a special result because we are dealing with a single oscillator. In the next variation, discussed in the section, we study a system with  $N$  oscillators and find that the energy distribution becomes very sharp as  $N$  increases.

#### 6.8.4 Weakly coupled harmonic oscillators

The system now consists of  $N$  independent linear oscillators coupled via a single bath. The total energy of this system is not fixed but fluctuates around an equilibrium value  $\bar{E}$ .<sup>23</sup> We want to obtain the probability to find the system at a microstate with energy  $E$ . This is still a canonical system and its microstate is described by the  $2N$ -dimensional vector  $\mathbf{r} = (p_1, \dots; x_1, \dots)$  that gives the position and momentum of the individual oscillators. As before, the probability of microstate is given by the canonical distribution,

$$\rho_c = e^{-\beta E},$$

where  $E$  is now given by the total kinetic and potential energy of the  $N$  oscillators,

$$E = \sum_{i=1}^N \left( \frac{p_i^2}{2m} + \frac{kx_i^2}{2} \right).$$

By analogy to the two-dimensional analysis, the volume of phase space in the vicinity of contour  $E$  is

$$\omega(E)dE \propto r^{n-1} dr$$

where  $n = 2N$  is the dimensionality of phase space,  $n - 1 = 2N - 1$  is the dimensionality of the hyper surface that defines the energy contour, and  $r = |\mathbf{r}|$  is the distance of a point on the contour from the origin. The length of this vector scales with energy as  $r \sim E^{1/2}$  (this follows from the fact that contour is an ellipse even in this high-dimensional space), and  $dr$  scales as  $E^{-1/2} dE$ . Then,

$$\omega(E)dE \sim E^{(n-1)/2-1/2} dE = E^{n/2-1} dE$$

and finally

$$\omega(E) \sim E^{N-1}. \quad (6.31)$$

Notice that the density of microstates is of the form postulated in Eq. (6.23). As with the single oscillator, the energy distribution  $f(E)$  is proportional to  $\omega(E)e^{-\beta E}$ ,

$$f(E) = C' E^{N-1} e^{-\beta E},$$

<sup>23</sup>We refer to this type of interaction as *weak coupling*. The oscillators are not directly interacting with each other but they all interact with a common bath that redistributes energy among them.

and the constant  $C'$  is obtained by the normalization condition,

$$\frac{1}{C'} = \int_0^\infty E^{N-1} e^{-\beta E} dE = \frac{(n-1)!}{\beta^N}.$$

Thus the final form of the energy distribution is

$$f(E) = \frac{\beta^N}{(N-1)!} E^{N-1} e^{-\beta E}. \quad (6.32)$$

The mean energy, second moment, and standard deviation are easily calculated; we find,

$$\begin{aligned} \bar{E} &= \frac{N}{\beta}, \\ \overline{E^2} &= \frac{N}{\beta^2}, \\ \delta E &= (\overline{E^2} - \bar{E}^2)^{1/2} = \frac{\sqrt{N}}{\beta}. \end{aligned}$$

The mean energy increases in proportion to  $N$ , which confirms it is extensive. The standard deviation  $\delta E$ , which represents the magnitude of a typical fluctuation, increases as  $\sim \sqrt{N}$ . The ratio  $\delta E/\bar{E}$  is

$$\frac{\delta E}{\bar{E}} = \frac{1}{\sqrt{N}},$$

and vanishes as the number of oscillators increases. Thus we recover the expected behavior in the thermodynamic limit, according to which the vast majority of microstates are clustered in the vicinity of the mean energy.

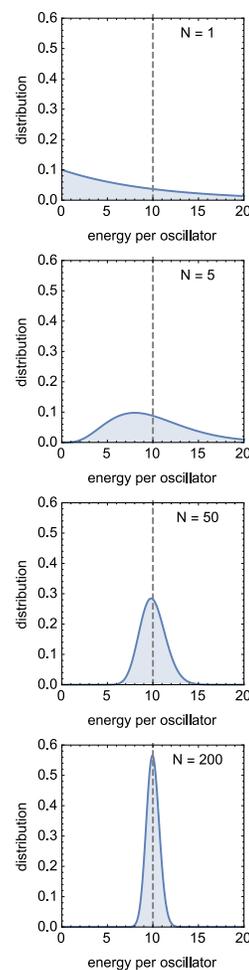
It is more instructive to calculate the probability density of the energy per oscillator  $\epsilon = E/N$ , an intensive property. Denoting this distribution as  $h(\epsilon)$  we have

$$h(\epsilon)d\epsilon = f(E)dE \Rightarrow f(\epsilon) = Nf(E),$$

then, using Eq. (6.32) with  $E = N\epsilon$ , the final result is

$$h(\epsilon) = \frac{N\beta^N (N\epsilon)^{N-1} e^{-\beta N\epsilon}}{(N-1)!}. \quad (6.33)$$

Figure 6.7 shows the energy distribution per oscillator for different numbers of oscillators at fixed  $\beta = 0.1$  (this fixes the mean energy per oscillator to  $\bar{E}/N = 10$ ). With  $N = 1$  we recover the exponential distribution of the single oscillator. As the number of oscillators is increased, the distribution sharpens and in the limit  $N \rightarrow \infty$  it becomes a Gaussian distribution whose variance goes to zero. This behavior is characteristic of systems composed of a large number of particles: the energy fluctuates in a very narrow region about its mean and in the thermodynamic limit all stochasticity vanishes and energy becomes a deterministic variable.



**Figure 6.7:** Energy distribution per oscillator in a system of  $N$  harmonic oscillators in one-dimensional motion coupled via a bath ( $\beta = 0.1$ ). In all cases the mean energy per oscillator is  $\bar{E}/N = 10$ .

## 6.9 Thermodynamics and the equations of motion

The equations of motion thread microstates into a time sequence such that a microstate at time  $t$  is the evolution of a corresponding microstate in the previous time step  $t - \delta t$  and leads to the next microstate at  $t + \delta t$ . This time sequence of microstates, literally a movie of microstates (Fig. 6.8), describes the dynamic evolution of the system through time and defines its trajectory in phase space. If  $F(t) = F(\Gamma(t))$  is a mechanical property that depends on the instantaneous microstate, its average value is given by the time average

$$\bar{F} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_t^{t+T} F(t) dt.$$

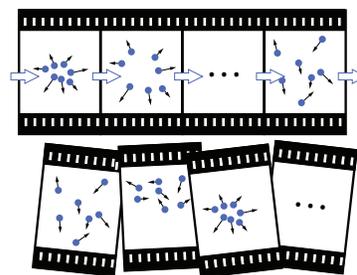
To actually calculate properties by this method we would need to solve the equations of state over a time interval  $T$  long enough to produce estimates of reasonable accuracy. This can be done numerically, but even with the high computing power of today's computers, the calculation becomes very slow quickly as the size of the system is increased. Ensemble theory allows us to bypass the equations of motion entirely. Instead, we collect all possible microstates and calculate the value of the mechanical property  $F$  as a weighted average over this ensemble with the weights given by the probability of microstate:

$$\bar{F} = \sum_i p_i F_i = c_N \int \rho(\Gamma) F(\Gamma) d\Gamma,$$

The ensemble average is not trivial to calculate, but is certainly much easier than solving the equations of motion. The equivalence between time and ensemble averages is the essential postulate of statistical mechanics. It implies that the movie and the ensemble contain exactly the same frames, and that the probability of microstate in the ensemble is equal to the fraction of time that the dynamic system spends on that microstate. This equivalence is known as the *ergodic* hypothesis. A process is ergodic if every microstate of the ensemble is connected via a kinetic trajectory to every other microstate.

Ergodicity comes into the picture when we attempt to derive thermodynamics starting with the equations of motion. There is a serious conceptual obstacle to this proposition. Given an initial state the equations of motion have a unique solution that extends forward and backward in time. Uniqueness of solution means that only one trajectory passes through any given point in phase space. It is difficult to see how a system that starts at a fixed initial state—a *point* in phase space—ends up spreading to the entire *volume* of phase space. This problem is endemic to the equations of mechanics and implies that any attempt to connect mechanics and thermodynamics must involve assumptions (postulates) that are external to mechanics.

**Boltzmann entropy and the  $H$  theorem** Kinetic theory is Boltzmann's attempt to derive thermodynamics from the equations of motion. It is beyond the scope of this book to discuss kinetic theory in any detail but one particular element of it, Boltzmann's  $H$  theorem, comes very close to our treatment that it deserves a closer look. Boltzmann considers a collection of  $N$  mechanical particles and defines the one-particle probability distribution  $f(\mathbf{r}, \mathbf{v})$  such that



**Figure 6.8:** The equations of motion produce a time sequence of microstates that evolve in time. The ensemble is the static collection of all frames that can be constructed given the macroscopic state. The system is ergodic if the frames of the ensemble can be pieced together into a single time sequence that satisfies the laws of motion.

$f d\mathbf{r}^3 d\mathbf{v}^3$  is the mean number of particles in the region of space  $(\mathbf{r}, \mathbf{r} + d\mathbf{r})$  whose velocity falls in  $(\mathbf{v}, \mathbf{v} + d\mathbf{v})$ . Boltzmann derived a transport equation that governs the evolution of  $f$  under binary collisions (Fig. 6.9):

$$\frac{\partial f_1}{\partial t} = \int d\Omega \int d\mathbf{v}_2 \sigma(\Omega) |\mathbf{v}_1 - \mathbf{v}_2| (f'_1 f'_2 - f_1 f_2),$$

where  $f_1 = f(\mathbf{r}, \mathbf{v}_1)$ ,  $f_2 = f(\mathbf{r}, \mathbf{v}_2)$ ,  $f'_1 = f(\mathbf{r}, \mathbf{v}'_1)$ ,  $f'_2 = f(\mathbf{r}, \mathbf{v}'_2)$ , and  $\sigma(\Omega)$  is the scattering cross section in solid angle  $d\Omega$  and is determined by the details of the intermolecular interactions. Boltzmann derived the transport equation under two crucial assumptions. The first assumption is that the system is a dilute gas and allows one to consider only binary collisions. The second assumption is known as *molecular chaos* and allows one to neglect correlations between particles so that the probability to find a pair particles entering a collision trajectory is equal to the product of the probability to find each individual particle in their respective states. The latter assumption endows the transport equation with the property of *irreversibility*, i.e., it evolves in the direction of establishing an equilibrium distribution, the Maxwell-Boltzmann distribution (see Problem 6.10).

To show the irreversible nature of the transport equation, Boltzmann defined the functional  $H[f]$  as

$$H[f] = \int d\mathbf{v} f(\mathbf{r}, \mathbf{v}) \log f(\mathbf{r}, \mathbf{v})$$

(we have omitted the dimensional constant in this equation) and showed that if  $f$  satisfies the transport equation, then  $H$  satisfies

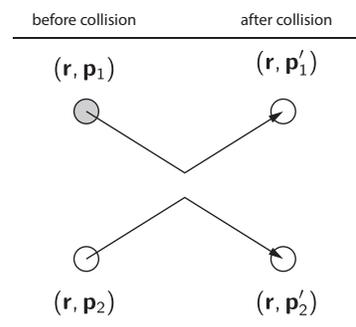
$$\frac{dH}{dt} \leq 0,$$

with the equal sign when  $f$  becomes time invariant Huang (1963, Chap. 4). The negative of  $H$  exhibits a hallmark of entropy, and thus

$$S_{\text{Boltzmann}} = -H[f] = - \int d\mathbf{v} f \log f. \quad (6.34)$$

Boltzmann's entropy functional is the same as that of Gibbs but while Gibbs's functional is applied to the probability of microstate (probability to find each particle at a particular state), Boltzmann's functional is applied to the one-particle probability (probability to find a single particle at a particular state irrespectively of the state of all other particles). That is, Boltzmann's entropy represents the special case of Gibbs's entropy in the ideal-gas limit.

**The Liouville equation and thermodynamics** The ensemble theory bypasses kinetics entirely by considering a static set of microstates regardless of how these might be threaded into a kinetic sequence. The equation that governs the flow of microstates through phase space is the Liouville equation, which like Boltzmann's transport equation is a differential equation for the evolution of a probability distribution. Unlike Boltzmann's transport equation, however, the Liouville equation is reversible in time. An important result of the Liouville



**Figure 6.9:** A binary collision can be represented as the reaction of two particles in  $\mu$  space,  $A(\mathbf{r}, \mathbf{v}_1) + A(\mathbf{r}, \mathbf{v}_2) \rightleftharpoons A(\mathbf{r}, \mathbf{v}'_1) + A(\mathbf{r}, \mathbf{v}'_2)$  with forward rate  $R^+ = f_1 f_2 \sigma(\Omega) |\mathbf{v}_1 - \mathbf{v}_2| d\Omega d\mathbf{v}_2$  and reverse rate  $R^- = f'_1 f'_2 \sigma(\Omega) |\mathbf{v}'_1 - \mathbf{v}'_2| d\mathbf{v}_2$ , where  $\sigma(\Omega) d\Omega$  is the scattering cross section in solid angle  $d\Omega$  and depends on the details of intermolecular interactions. The reverse rate can be expressed in terms of  $\mathbf{v}_2$  because the pair  $(\mathbf{v}'_1, \mathbf{v}'_2)$  is uniquely determined by  $(\mathbf{v}_1, \mathbf{v}_2)$  and  $\mathbf{v}_1$  is held constant. For equally sized particles  $|\mathbf{v}_1 - \mathbf{v}_2| = |\mathbf{v}'_1 - \mathbf{v}'_2|$  by virtue of energy and momentum conservation. The transport equation is obtained as the rate of producing state  $(\mathbf{r}, \mathbf{v}_1)$  via the reverse reaction minus the rate of depleting it via the forward reaction. For a more detailed derivation see Huang (1963, Chap. 3).

theorem is that the volume of phase space is preserved. If we tag a finite region of phase space surrounded by a closed surface at time  $t$  and follow its evolution in phase space, its volume will remain constant at all times (Fig. 6.10). The shape of the element may be stretched and twisted during its motion, but its volume will be unchanged. If  $\Delta\Omega$  is the volume of this element at time  $t$  and  $\Delta\Omega'$  its volume at some later time  $t'$ , then  $\Delta\Omega' = \Delta\Omega$ . This also means that the value of the entropy functional,

$$-\int \rho \log \frac{\rho}{c_N} d\Gamma$$

is constant at all times.<sup>24</sup> This is true for *any* probability distribution  $\rho$ , whether it represents equilibrium or not. The Liouville equation is capable of describing isentropic changes only and cannot describe the irreversible approach to equilibrium. This limitation is built into the equations of motion, which move reversibly forward and backward in time. Neither can the Liouville equation fix the equilibrium probability distribution of microstate. This is fixed by the extra-mechanical postulate that  $\rho$  must be chosen as the distribution that maximizes the entropy functional under the given macroscopic constraints. The Liouville equation then must be viewed as a necessary but not sufficient condition on the equilibrium probability distribution.

**Irreversibility as incomplete knowledge** If the Liouville equation preserves entropy at all times, how should we interpret the second law, which requires entropy to increase in the approach to equilibrium? With reference to Fig. 6.10, we start from initial state  $A$  and track an element of phase space initially at time  $t$  with volume  $\Omega = \Omega(t)$  and initial probability distribution  $\rho$ . We subject the ensemble to an adiabatic process by varying the external coordinates until at time  $t'$ , at which point the external coordinates become constant again. The equilibrium distribution at the final state,  $\rho'$ , maximizes the entropy functional with respect to any other distribution over the same microstates. Let  $\rho'_L$  be the distribution that is obtained from the Liouville equation at time  $t'$  using  $\rho(t)$  as the initial condition. As we will see,  $\rho'$  and  $\rho'_L$  are not necessarily the same. By property of the Liouville equation the entropy functionals on  $\rho$  and  $\rho'_L$  are equal,

$$S_A \equiv -\int_{\Omega} \rho \log \frac{\rho}{c_N} d\Gamma = -\int_{\Omega'} \rho'_L \log \frac{\rho'_L}{c_N} d\Gamma.$$

As a solution of the Liouville equation,  $\rho'_L$  is a *possible* probability distribution at time  $t'$ . Its entropy then is *less or equal* to the canonical distribution  $\rho'$  that corresponds to the final state ( $T', V', N$ ). That is

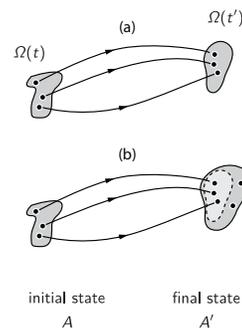
$$S(\rho'_L) \leq S(\rho').$$

Since  $S(\rho')$  is the final entropy  $S_{\text{final}}$ , and  $S(\rho'_L)$  is equal to the initial entropy  $S_{\text{initial}} = S(\rho)$ , we have

$$S_{\text{final}} \geq S_{\text{initial}}. \quad (6.35)$$

This is the standard statement of the second law. The geometric interpretation of this analysis is that the volume of phase space is *not always conserved* – It is conserved *only* if the process is *reversible*. Then, and then only, is there a

<sup>24</sup>See Problem 6.7.



**Figure 6.10:** (a) A Hamiltonian system preserves the volume of phase space during any process (Problem 6.7); this amounts to a reversible adiabatic process and conserves entropy. (b) In real processes the volume of phase space increases as a result of degrees of freedom that are not explicitly accounted for in the Hamiltonian; this amounts to entropy generation. Think of it as “leakage” of information, a decrease in our degree of certainty about the system, as more things are possible than previously thought. The conclusion is that the end state of an adiabatic process is *not* uniquely determined by the initial state *unless* the process is reversible.

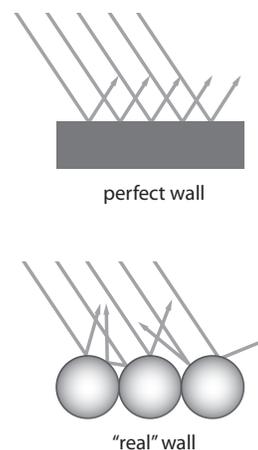
one-to-one correspondence between the microstates in the initial and final state: every initial microstate is linked to one final microstate via a dynamic trajectory described by the Liouville equation, and vice versa. If the external driving forces are reversed, then all microscopic trajectories are reversed and the system retraces its path to the initial state. If the process is not reversible the one-to-one mapping is lost as the final system contains more microstates than the initial state. Where do they come from?

The time dependent Liouville equation is a partial differential equation for the probability density  $\rho$ . To solve it we need an initial condition in time, and a set of boundary conditions that determine how the system interacts with the outside world. Gibbs prescribed the following recipe:

1. *Initial conditions*: use every possible microstate under the given boundary conditions.
2. *Boundary conditions*: take them to be fixed, or allow them to be deterministic functions of time.

This prescription views the boundary of a closed system as a deterministic external potential that confines the system in space. In reality, this boundary is *not* deterministic. It consists, just as the system itself, of a very large number of particles, each with its own degrees of freedom. By modeling the boundary as a deterministic potential, we have effectively *coarsened* its degrees of freedom into an average potential. These degrees of freedom do not appear in the Hamiltonian explicitly, yet they influence the dynamic evolution in a way that conflicts with the solution of the coarsened Liouville equation. Consider the behavior of particles of the system near the wall. A “perfect” wall, i.e., one that is described by a deterministic function  $A_i$ , reflects particles in ways that depend entirely on the properties of the particle; a real wall, on the other hand, produces a trajectory whose details depend on the specific interactions between particles of the system and particles of the wall (Fig. 6.11). To solve the problem then we eliminate the wall details but by doing that we create a new problem: the final state is not fully predictable from its state at an earlier time because some physical details are missing. This amounts to a “leakage of microstates” into the phase space, as more things are possible at future time  $t'$  than we had anticipated based on our earlier knowledge at time  $t$ : entropy increases.

The conundrum seems insurmountable: a system undergoing a process interacts with its surroundings and this means that if we want to solve the equations of motion for the system we must actually solve them for the entire universe! Gibbs’s method to resolve this indeterminacy was the stroke of a genius: forget about solving the equations of motion. Assume instead that *every* possible microstate is equally probable because if it is not the evolution of a parent microstate at some earlier time, it must have “leaked” into the phase space via all those physical details that were left out of the Liouville equation. This translates to a straightforward recipe: when physics does not give enough information to know which distribution to pick, pick the one that maximizes entropy. For those who adopt information theory as their primary interpretation of entropy,<sup>25</sup> this is the only guess that makes sense: if the problem has an infinite number of possible answers that differ on how they treat microscopic details that are hopelessly beyond our ability to evaluate, assume that all of them are present and average over them to obtain any property of the system.



**Figure 6.11:** Trajectories of point particles that hit a wall made of hard spheres scatter over a *range* of angles, even though the incident angle is the same for all incoming particles. When we treat the wall as a perfect plane we underestimate the number of future microstates because we take all trajectories with the same incident angle to bounce off in the same direction. This is an example of coarsening. The problem is not only that the actual shape of the wall is complicated, but it also changes in time as the wall molecules themselves move around. This picture should not be read literally to mean that the molecular constitution of walls renders all processes irreversible. Rather that the omission of physical details from *our models* is a source of unanticipated microstates. Whether these microstates are large enough in number to cause a measurable increase of entropy depends on the details of the process.

<sup>25</sup>This school of thought is known as Bayesian and takes its name from Bayes’ rule in probability theory. A champion of this approach was E. T. Jaynes, who used Shannon’s ideas to draw a connection between statistical mechanics and information theory (see Additional Reading).

## 6.10 Additional reading

- A highly readable account of the development of statistical mechanics, its controversies, and the unresolved philosophical questions it raises is given by [Uffink \(2006\)](#).
- Coarsening and the emergent nature of entropy are discussed in [Sethna \(2006\)](#). The discussion of inequality of the second law in relationship to the reversible Liouville equation comes from [Jaynes \(1965\)](#) (see section V of that paper).
- A brief technical account of ergodic theory and its relevance to statistical mechanics is given by [Bühler \(2006\)](#). This book is terse but to the point. A more accessible and informative discussion of ergodicity is given in [Sethna \(2006\)](#), along with examples of non-ergodic systems in astrophysics.
- The discussion of the second law in section 6.9 is from [Jaynes \(1965\)](#) (the paper will make more sense if you first study chapter 7).
- The Liouville equation is discussed in most standard textbooks on statistical mechanics, especially those with a physics bent. Pathria's book is good reference on this subject ([Pathria and Beale, 2011](#)).
- The original source on statistical mechanics is [Gibbs \(1981\)](#). Much of the discussion in this chapter draws from Gibbs's book. The original text is rather difficult to follow, especially for someone with no previous background in statistical mechanics. That's because Gibbs's purpose is not to teach but to derive a theory that will stand against all possible criticism. Nonetheless the basic ideas are both powerful and quite simple to convey. A few problems at the end of the chapter are based on Gibbs's derivations and provide enough hints to help you plow through them.

## 6.11 Problems

**Problem 6.1.** (Microcanonical oscillator – theory and simulation) The motion of a 1-dimensional harmonic oscillator with mass  $m$  inside a potential  $U(x) = kx^2/2$  is given by the equation

$$x(t) = x_0 \cos(\omega t)$$

where  $x_0$  is the maximum distance travelled from equilibrium, and  $\omega = \sqrt{k/m}$ .

- Obtain the momentum  $p(t)$  as a function of time.
- Calculate the mean squared displacement  $\overline{x^2}$  from the time average

$$\overline{x^2} = \frac{1}{T} \int_0^T x^2(t) dt,$$

where  $T = 2\pi/\omega$  is the period of the oscillator.

- The total energy of the oscillator is

$$\frac{p^2}{2m} + \frac{kx^2}{2} = E$$

and defines an ellipse that represents a contour of constant energy in phase space. The volume of the microcanonical ensemble is the area of a thin shell between two contours at  $E$  and  $E + \Delta E$ , as in Fig. 6.12. The value of  $\Delta E$  is not important as long as it is constant and small enough that changes between the contours can be calculated from first-order derivatives.

Show that the probability density is uniform inside the microcanonical shell.

*Hint:* Take the probability density to represent the material density of a fluid moving through the shell between the two contours. The problem is asking you to show that the

density of this fluid is constant throughout the entire volume. The continuity equation for an incompressible fluid in three dimensions is

$$\frac{d\dot{x}}{dx} + \frac{d\dot{y}}{dy} + \frac{d\dot{z}}{dz} = 0.$$

You need show that the motion of the oscillator satisfies this equation in the two dimensional space  $(x, p)$ .

c) The previous part establishes the fact that any microstate in the microcanonical volume is equally probable. Construct a Monte Carlo simulation to sample microstates from that volume and calculate the mean square displacement  $\overline{x^2}$ . Perform the calculations for  $x_0 = 1$ ,  $k = 2$ ,  $m = 8$ . To pick points on the surface of the ellipse with constant density, choose a random number  $\text{rnd}$  with uniform probability in the interval  $(0, 2\pi)$ , set  $t = \theta/\omega$  and use your equations for  $x(t)$  and  $p(t)$  to obtain the coordinates of the state (since you are only asked to calculate  $\overline{x^2}$  you don't really need the value of  $p$  in this calculation).

d) Use your simulations results to make a plot of the probability to find the oscillator in position  $x$ . Derive the theoretical distribution and compare with results of the simulation.

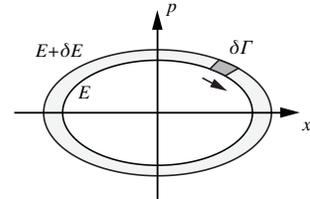


Figure 6.12: Problem 6.1.

**Problem 6.2.** (Canonical harmonic oscillators – theory) Show that the entropy of a system of  $N$  canonical oscillators can be calculated by either of the two equations below:

$$\frac{S}{k} = \log(\omega(E)\delta\epsilon) = \log\Omega(E),$$

where  $\log\Omega(E)$  is the cumulative number of microstates with energy less than or equal to  $E$ .

**Problem 6.3.** (Canonical oscillators – simulation) In this problem we will build a Monte Carlo simulation for a small ensemble of  $N$  oscillators that exchange energy so that the total energy is fixed. We will keep track of the microstate in a two dimensional list  $((x_1, p_1), (x_2, p_2), \dots, (x_N, p_N))$  that contains the coordinates in phase space of each oscillator. We will sample the ensemble as follows:

1. Pick two oscillators at random, say  $i$  and  $j$ , and exchange energy between them so that the energies after the exchange are  $E'_i = \epsilon$ ,  $E'_j = E_i + E_j - \epsilon$ , where  $E_i$  and  $E_j$  are the energies before the exchange. Pick the value of  $\epsilon$  randomly in the interval  $(0, E_i + E_j)$ .
2. Having determined the energies of the oscillators after the exchange, determine their coordinates in phase space by selecting at random a microstate on the corresponding energy contour (use the method described in Problem 6.1).

Use  $N = 100$ ,  $m = 1$  kg,  $k = 1$  J/m<sup>2</sup>,  $E_{\text{tot}} = 1000$  J and run the simulation for 10000 Monte Carlo steps.

- a) Take the first  $N_A = 50$  oscillators and view them as a canonical subsystem that interacts with the remaining oscillators. Obtain the total energy  $E_A$  within this subsystem and make a plot as a function of Monte Carlo steps. Discuss the features of this graph. How does the mean energy of this subsystem compare with  $N_A E_{\text{tot}}/N$ , which is the value one would expect based on homogeneity?
- b) Use data from the simulation to obtain the distribution (probability density function) of  $E_A$ . Make a plot of this distribution and superimpose a Gaussian distribution with mean and variance equal to the values you obtained from the distribution. How do the two distributions compare? Is this what would have expected?
- c) Obtain the temperature of this system (in K) from the simulation.
- d) Why do we pick  $\epsilon$  with uniform probability between 0 and  $E_i + E_j$ ? Shouldn't we be using a Boltzmann factor to weigh this probability?

**Problem 6.4.** (Canonical oscillators – equation of state)  $N$  one-dimensional harmonic oscillators are coupled via a bath at fixed temperature  $\beta$ . The oscillators are arranged in space so that their equilibrium points line on a vertical line in the  $z$  direction at distance  $L$  from each other (Fig. ??). The motion takes place on the  $xy$  plane with each oscillator moving on a line that is randomly oriented with respect to  $x$ . We may define the volume of this system as the volume of a cylinder with height  $NL$  and radius equal to the square root of  $\bar{r}^2$ , the ensemble average square displacement of all oscillators (since the energy of individual oscillators can be anywhere between 0 and  $\infty$ , a fraction of them will be outside this volume).

- Express the volume of the system in terms of  $m$  (mass of oscillators),  $k$  (spring constant),  $\beta$ ,  $N$  and  $L$ .
  - Obtain the pressure of the system (hint: express the canonical partition function in terms of  $\beta$ ,  $V$  and  $N$ )
  - Obtain the equation of state.
  - Obtain the constant volume heat capacity.
- You may treat the oscillators as distinguishable or indistinguishable as you please.

**Problem 6.5.** (Gibbs's derivation of canonical thermodynamics) This problem walks you through the steps of Gibbs's derivation of the connection between the canonical ensemble and macroscopic thermodynamics.

The potential energy of a system of  $N$  particles is generally a function of the particle coordinates,  $q_i$ , as well as of a set of external coordinates,  $a_1, a_2, \dots$ , that determine the physical volume occupied by the system. The negative derivative of the potential with respect to a particle coordinate is the force on the particle in the direction of that coordinate:

$$\left(\frac{\partial U}{\partial q_i}\right) = \left(\frac{\partial \mathcal{H}}{\partial q_i}\right) = -F_i$$

Similarly, the negative derivative with respect to an external coordinate is the force exerted by the system at the location of that coordinate:

$$\left(\frac{\partial U}{\partial a_i}\right) = \left(\frac{\partial \mathcal{H}}{\partial a_i}\right) = -A_i$$

(For example, the spring constants of the oscillators in Problem 6.4 play the role of the external forces, and the volume of the system that of the external coordinates.)

- Using  $Q = \exp(-\psi/\Theta)$ ,  $\Theta = 1/\beta$ , with  $\psi$  a parameter, write the canonical probability in the form,

$$\rho_c(\Gamma) = c_N e^{(\psi - \mathcal{H})/\Theta}$$

then show that

$$\frac{\psi - \bar{\mathcal{H}}}{\Theta} = \int \rho_c(\Gamma) \log \frac{\rho_c(\Gamma)}{c_N} d\Gamma,$$

where  $\bar{\mathcal{H}}$  is the mean value of the Hamiltonian in the canonical ensemble (mean energy).

- Using  $\psi$  and  $\Theta$ , write Eq. (??) for the partition function in the form

$$e^{-\psi/\Theta} = c_N \int e^{-\mathcal{H}/\Theta} d\Gamma.$$

Take the differential of both sides treating the left-hand side as a function of  $\Theta$  and  $\psi$ , and the right hand side as a function of  $\Theta$  and the external coordinates. Assume there is only one external coordinate,  $a$ , with corresponding force  $A$ . Show that the result is

$$d\psi = -\frac{\psi - \bar{H}}{\Theta} d\Theta - \bar{A} da,$$

where  $\bar{A}$  is the mean value of force  $A$ .

c) Combine the results from parts a and b to eliminate  $\psi$  and show that the final result can be expressed as

$$d\bar{\mathcal{H}} = -\Theta d\bar{\eta} - \bar{A} da,$$

where  $\bar{\eta} = (\psi - \bar{\mathcal{H}})/\Theta$ . Identify the thermodynamic equivalents of all terms in this equation.

(The details of this derivation can be found in [Gibbs \(1902\)](#), pp 42–45.)

**Problem 6.6.** (Gibbs's derivation of the microcanonical ensemble) This problem outlines Gibbs's method for establishing the relationship between the microcanonical ensemble and macroscopic thermodynamics.

a) To avoid the singularities associated with microstates of constant energy, we write the microcanonical probability in the form,

$$\rho(\Gamma) = \frac{e^{-(E-E')^2/\epsilon^2}}{W},$$

where  $W$  and  $\epsilon$  are constants. By making  $\epsilon$  arbitrarily small, the Gaussian function on the right-hand side becomes very sharp around  $E'$  and the corresponding ensemble contains only microstates with energy exactly equal to  $E'$ . The constant  $W$  is given by the normalization condition,

$$W = \int e^{-(E-E')^2/\epsilon^2} d\Gamma.$$

Take the derivative of this expression with respect to external coordinate  $a_i$  noting that  $\partial E/\partial a_i = -A_i$ , where  $A_i$  is the force associated with this coordinate and show that the result is

$$\frac{\partial W}{\partial a_i} = \int \left( A_i \frac{d\omega}{dE} + \omega \frac{dA_i}{dE} \right) e^{-(E-E')^2/\epsilon^2} dE.$$

where  $\omega = \partial\Omega/\partial E$  is the density of microstates per unit energy.

b) Express the normalization constant in the alternative form,

$$W = \int e^{-(E-E')^2/\omega^2} \omega dE$$

and take again the derivative with respect to  $a_i$  noting that  $\omega$  is the only factor on the right-hand side that depends on  $a_i$  (why?). Show that the result is

$$\frac{\partial W}{\partial a_i} = \int \frac{d\omega}{dE} e^{-(E-E')^2/\omega^2} dE.$$

c) Combine the results from (a) and (b) to show that

$$A_i \frac{d\omega}{dE} + \omega \frac{dA_i}{dE} = \frac{d\omega}{da_i}.$$

d) Show that the result in (c) may also be written as

$$\frac{\partial(\omega A_i)}{\partial E} = \frac{\partial^2 \Omega}{\partial a_i \partial E},$$

and conclude that

$$\frac{\partial \log \Omega}{\partial a_i} = \frac{\omega A_i}{\Omega}.$$

e) In (d) we obtained the partial derivative of  $\Omega$  with respect to  $a_i$ ; we also know that its partial derivative with respect to  $E$  is  $\omega$ . Construct the differential of  $\log \Omega(E, a_1, a_2, \dots)$  with variables  $E, a_1, a_2, \dots$ , and show that the final result is

$$dE = \frac{\Omega}{\omega} d \log \Omega - \sum_i A_i da_i.$$

## 6.11. Problems

Conclude that  $\log \Omega = S/k$  and  $\Omega/\omega = kT$ , where  $k$  a dimensional constant.

f) Show that the result for the temperature obtained in the previous part is equivalent to

$$T = \frac{\partial E}{\partial \log \Omega}$$

g) The differential in part (e) can also be written as

$$dE = \frac{d\Omega}{\omega} - \sum_i A_i da_i,$$

from which one might argue that the proper assignments should be  $\Omega \rightarrow S$  and  $1/\omega \rightarrow T$ . What is your response to this criticism of the theory?

(The details of this derivation can be found in [Gibbs \(1902\)](#), pp 125–128.)

**Problem 6.7.** (Gibbs entropy and the Liouville equation) Consider an ensemble of particles described by a Hamiltonian function with  $n$  degrees of freedom, and with probability density  $\rho(\Gamma; t)$  that also depends on time (i.e., the system is not in statistical equilibrium).

a) If  $f = f(\rho)$  is some function of  $\rho$ , show that it satisfies the equation

$$\frac{\partial f}{\partial t} = - \sum_{i=1}^n \left[ \frac{\partial(f\dot{p}_i)}{\partial p_i} + \frac{\partial(f\dot{q}_i)}{\partial q_i} \right]$$

where  $p_i$  and  $q_i$  are the generalized momenta and coordinates corresponding to the  $i$ th degree of freedom.

b) Use Gauss's theorem to show that the integral of  $f$  over the entire [phase space is constant, i.e.,

$$\frac{\partial}{\partial t} \int f(\rho) dp_1 \cdots dq_1 \cdots = 0.$$

c) Conclude that the Gibbs entropy,

$$S/k = - \int \rho \log \frac{\rho}{h^n} dp_1 \cdots dq_1 \cdots$$

is constant at all times ( $h$  is Planck's constant and makes  $\rho$  dimensionless).

d) Subject the system to an adiabatic process (compression or expansion) by allowing the external coordinates to be functions of time. Determine the effect of this process on the Gibbs entropy.

e) Discuss the implications of your answer in the previous question in relation to reversibility and in particular as to whether the Liouville equation is capable of describing irreversible process.

**Problem 6.8.** Use the canonical partition function to show

$$\left( \frac{\partial \bar{E}}{\partial V} \right)_{\beta, N} + \beta \left( \frac{\partial \bar{P}}{\partial \beta} \right)_{N, V} = - \langle P \rangle$$

where  $\langle P \rangle$  is the mean pressure in the canonical ensemble. *Hint:* Recall that the pressure of microstate  $i$  can be expressed as

$$P_i = - \left( \frac{\partial E_i}{\partial V} \right)_N$$

[This problem should be discussed in the context of reversibility and conservation of microstates]

## 6.11. Problems

**Problem 6.9.** (Gibbs inequality) a) The *Gibbs inequality* states that for any two probability distributions  $\rho_1(x)$  and  $\rho_2(x)$  defined over the same domain, the following inequality is always true:

$$\int \rho_1(x) \log \frac{\rho_2(x)}{\rho_1(x)} dx \geq 0,$$

with the equal sign only of  $\rho = \rho'$ . Prove this inequality.<sup>26</sup> *Hint:* Add the identity

$$\int (\rho_2(x) - \rho_1(x)) dx = 0$$

to the Gibbs inequality and show that the quantity inside the integral is never negative.

b) Consider two possible probability distributions for the microstate of a system of a  $N$  particles in volume  $V$  with the following properties:

$\rho_1$  is an explicit function of the microstate  $\Gamma$ .

$\rho_2$  is a function of the energy of the microstate,  $E(\Gamma)$ .

Both distributions produce the same energy distribution, i.e., the number of microstates in the interval  $(E, E + dE)$  is the same for both distributions:

$$\int_E^{E+dE} \rho_1 d\Gamma = \int_E^{E+dE} \rho_2 d\Gamma. \quad (\text{A})$$

Use the Gibbs inequality to show that  $\rho_2$  has *higher* entropy than  $\rho_1$ . *Hint:* Multiply both sides of equation (A) with  $\log \rho_2$ , let  $\rho_2$  inside the integral (why?) and show that

$$\int \rho_1 \log \rho_2 d\Gamma = \int \rho_2 \log \rho_2 d\Gamma, \quad (\text{B})$$

with the integration going over all phase space. Finally combine with the Gibbs inequality with  $x = \Gamma$  to prove  $S(\rho_2) > S(\rho_1)$ .

c) Discuss what this means for statistical mechanics. In particular, consider the fact that the Liouville equation does not require the probability of state to be a function of energy – it could be, but it does not have to. Nonetheless, Gibbs considered only distributions that are functions of energy. How does the inequality derived in part (b) justify Gibbs's approach?

**Problem 6.10.** Show that the distribution

$$f(v) = C e^{-Av^2}$$

with

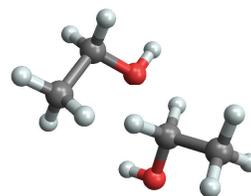
$$C = n \left( \frac{3m}{4\pi\epsilon} \right)^{3/2}, \quad A = \frac{3m}{4\epsilon}$$

where  $\epsilon$  is the mean kinetic energy satisfies the stationary transport equation.

<sup>26</sup>This result was obtained by Gibbs as an intermediate step in the derivation of Theorem II in chapter IX (Gibbs, 1902). The inequality plays an important role in Information Theory (the integral on the right-hand side is called relative entropy or directed divergence).

## Classical Molecular Systems

A molecular system is a collection of interacting particles, where by “particles” we refer to constituent elements of molecules. We will take particles to be atoms, although the definition can be expanded to include electrons, protons or any other unit within the molecule. The simplest model for a molecule is a point particle with no internal structure. Such particle has three degrees of freedom corresponding to translation in three dimensions. A simple diatomic molecule may be represented by two point particles (atoms) linked via a bond. We commonly represent bonds as rigid rods (Fig. 7.1), in reality they are more like very stiff springs that allow vibrations. The simple diatomic molecule has three translational degrees of freedom, which refers to the motion of the center-of-mass in space, two rotational degrees in two directions perpendicular to the axis (rotation about the axis produces no motion since we treat the atoms as points), and one vibrational degree of freedom. This makes for 6 degrees of freedom, the same as for a pair nonbonded particles. The fact that the bond restricts the motion of the diatomic molecule does not remove the degree of freedom. The situation is analogous to a single particle in a very steep potential well. The particle has three degrees of freedom but the strong force keeps its motion inside a narrow region around the bottom of the potential. If we replace the spring with a rigid rod we remove the vibrational degree of freedom and the total number of degrees of freedom reduces to 5. The rigid bond is a limiting case of an infinitely stiff spring. In this sense a diatomic molecule may be represented as two particles (atoms) interacting via a very deep potential. A system composed of molecules with complex structure may be viewed now as a collection of interacting atoms. Neighboring atoms in the same molecule interact via strong, deep potentials, while more distant atoms and atoms that belong to different molecules interact much more weakly. The Hamiltonian of this system is calculated by summing the kinetic energies of all constituent particles and adding the interaction energy among all particles in the system. We may continue to refine the scale of interactions and account for the electrons and nuclei that constitute the building “particles” of atoms. This would require the use of quantum mechanics and we will not do it here. Rather, the point is that, in principle, the energy of a molecular system is known if all the its constituent elements are taken into consideration. Fortunately, it is not necessary to work with a model of extreme atomic detail. Under temperatures typical of systems of chemical engineering interest, the time scales of inter-atomic interactions



**Figure 7.1:** Simple molecular model of ethanol.

are much smaller than those of intra-atomic and intra-molecular interactions that the corresponding motions can be separated and superimposed as if they were totally independent. We will adopt a simple model that treats molecules as point particles that interact via a spherically symmetric intermolecular potential. The model is crude but adequate to demonstrate how statistical mechanics works with real systems. More realistic models are built by adding detail to the intermolecular potential but this only affects the degree complexity of the calculation – the big picture remains the same: once we have the Hamiltonian of a molecular system, we can perform thermodynamic calculations using statistical mechanics.

## 7.1 Intermolecular potential

Molecules interact through a variety of forces. They include ionic forces, interactions between permanent and induced dipoles, hydrogen bonding and others. It is not possible to describe all of these interactions by a single mathematical function. There are, however, some general characteristics that we will use to construct simplified models. Regardless of the nature of the intermolecular force, the interaction between two molecules may be represented by the intermolecular potential  $\phi_{12}$ . The intermolecular potential is the energy of a pair of particles at distance  $r$  from each other. Since we are treating particles as spherically symmetric, their interaction depends only on the distance between them. The force between the particles is given by the derivative of the potential,

$$F = -\frac{\partial\phi_{12}}{\partial r}. \quad (7.1)$$

If we change the distance from  $r$  to  $r'$ , the associated work is

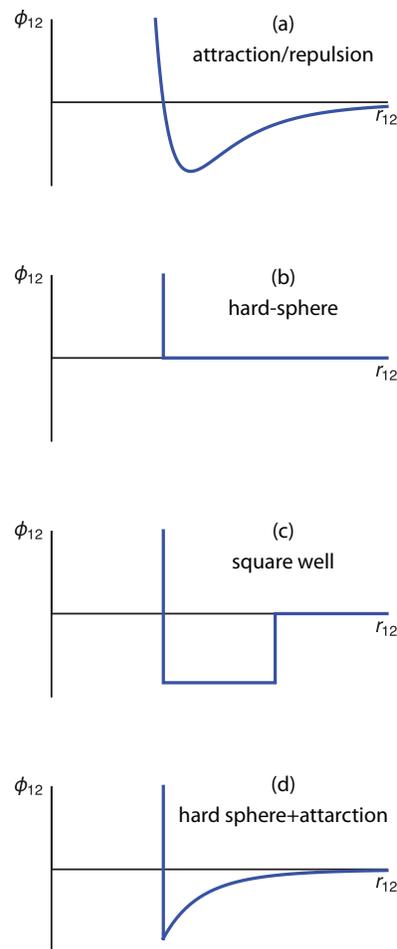
$$W_{r \rightarrow r'} = \int_r^{r'} F dr = - \int_r^{r'} d\phi_{12} = -(\phi_{12}(r') - \phi(r)).$$

At infinite separation there is no interaction as all physical forces have finite range, i.e.,  $\phi_{12}(\infty) = 0$ . Setting  $r' = \infty$  in the above equation we obtain,

$$W_{r \rightarrow \infty} = \phi_{12}(r), \quad (7.2)$$

which states that the intermolecular potential is the work that is produced by the system when the two particles are moved from distance  $r$  to  $\infty$ . Conversely,  $\phi_{12}(r)$  is the work that must be consumed by the surroundings to assemble the pair from infinite separation to distance  $r$ .

The force between two molecules may be attractive or repulsive, depending on their distance. At distances close enough that the electron clouds overlap, the force is very strongly repulsive that gives molecules the appearance of a solid core and prevents two molecules from occupying the same point in space. The potential in this region is a very steep function of  $r$  with negative slope. At further distances there is an attraction that develops from the concerted fluctuations of the electron clouds in the two molecules (London dispersion force), or from interaction between permanent or induced dipoles. This attraction forces are collectively known as van der Waals force, and scale as  $-1/r^6$  (the minus sign



**Figure 7.2:** Examples of model potentials. (a) Typical behavior with strong repulsion at short distances, attraction at intermediate separations. (b) Hard-sphere potential: infinite repulsion at  $r = r^*$ , zero interaction at longer distances. (c) Idealized potential: hard-sphere repulsion with square attractive well; (d) hard-sphere repulsion superimposed to smooth attraction.

## 7.1. Intermolecular potential

indicates that the potential increases with  $r$ , i.e., the force is attractive). The combination of repulsion at short range and attraction at longer range typically produces a minimum in the potential at some intermediate distance (7.2a). The steep repulsion is often modelled as a hard sphere (7.2b):

$$\phi_{12}^{\text{HS}}(r) = \begin{cases} \infty, & 0 < r < \sigma, \\ 0, & r > \sigma \end{cases} \quad (7.3)$$

This potential is infinite at separation distances shorter than  $\sigma$  and zero everywhere else. The closest possible distance between two particles is  $\sigma$ , equal to the diameter of the particles.<sup>1</sup> In this model particles do not interact unless they touch: they move in straight lines and collide elastically upon touching. Attraction can be added in a number of forms. In the square-well model, the minimum of the potential is represented by a rectangular well (7.2c). Alternatively, it may be represented by a smooth line that increases as  $\sim -1/r^6$  (7.2d).

<sup>1</sup>The closest distance between two solid spheres of radius  $R$  is  $2R$ .

A simple mathematical form that describes in a continuous and differentiable form the attractive/repulsive interaction is the Lennard-Jones potential:

$$\phi(r) = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right]. \quad (7.4)$$

The repulsive part is represented by a steep descent that scales as  $1/r^{12}$  and the attractive part by the van der Waals scaling  $\sim 1/r^6$  and for this reason it is also known as the 6-12 potential. The minimum occurs at  $r = 2^{1/6}\sigma = 1.122\sigma$  and its depth is equal to  $\epsilon$ . The core of the particle is soft but very stiff, that is, we can bring the particles to any distance from each other, but the repulsion becomes rapidly very strong when the separation distance is less than about  $\sigma$ . Thus we may take  $\sigma$  to represent the radius of the particle core. The minimum of the potential is located very close to the particle radius. The attractive portion extends a couple particle diameters and is practically zero past  $r \gtrsim 2\sigma$ . The Lennard-Jones potential is commonly used to describe interactions between non-polar atoms. Its numerical simplicity compared to more physically realistic potentials makes it a popular choice even for more complicated molecules and the  $\epsilon$  and  $\sigma$  parameters of many common compounds can be found in the literature.

The interaction between  $N$  molecules is calculated by taking the pair interactions to be additive. Thus the total potential energy is the sum of all pair interactions,

$$U(\mathbf{r}^N) = \sum_{\text{all pairs}} \phi_{12}(|r_i - r_j|), \quad (7.5)$$

and the Hamiltonian of the system is

$$\mathcal{H} = \frac{p_{ix}^2 + p_{iy}^2 + p_{iz}^2}{2m} + \sum_{\text{all pairs}} \phi_{12}(|r_i - r_j|). \quad (7.6)$$

Therefore, if the intermolecular potential is known, the Hamiltonian of the system is fully specified.

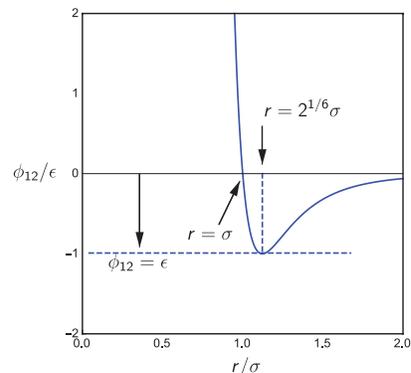


Figure 7.3: Lennard-Jones potential.

## 7.2 Canonical partition function

In chapter 6 we expressed the partition function of a collection of particles as an integral over continuous phase space (Eq. ??). We now write this equation more explicitly as

$$Q(T, V, N) = \frac{1}{N!h^{3N}} \int e^{-\mathcal{H}(p,r)/kT} d^{3N}p d^{3N}r. \quad (7.7)$$

Here, in addition to the dimensional constant  $h^{3N}$  we have the factor  $N!$ . To understand the reason for the presence of this factor, we must first explain that the integral on the RHS is taken with *all* momentum coordinates spanning the entire phase space from  $-\infty$  to  $+\infty$ , and *all* position coordinates spanning the region of physical space occupied by the volume of the system, i.e., the integration limits are

$$\int d^{3N}p d^{3N}r = \int_{-\infty}^{+\infty} dp_{1x} dp_{1y} dp_{1z} \int_{-\infty}^{+\infty} dp_{2x} dp_{2y} dp_{2z} \cdots \int_V dr_{1x} dr_{1y} dr_{1z} \int_V dr_{2x} dr_{2y} dr_{2z} \cdots$$

That is, every particle is placed in every position of the available phase space. This leads to overcounting because it counts as distinct configurations that are identical except with respect to which particle is placed in each position. If all particles are molecules of the same kind (pure component), this results in overcounting by exactly  $N!$ .<sup>2</sup> Thus the presence of  $N!$  in the denominator. It is possible to construct the integration in such a way that overcounting is avoided. The problem is that to achieve this we must use different integration limits for each particle,<sup>3</sup> which, however, is an undesired complication since otherwise identical particles are mathematically treated differently. It is easier to apply uniform integration limits for *all* particles and apply the correction for overcounting at the end. In doing this, we are momentarily treating the particles as distinguishable and apply the correction for distinguishability afterwards.

## 7.3 Configuration integral

Writing the Hamiltonian as the sum of kinetic and potential energy, the partition function becomes

$$Q = \frac{1}{N!h^{3N}} \int_{-\infty}^{\infty} e^{-\sum p_i^2/2mkT} d^{3N}p \int_V e^{-U/kT} d^{3N}r \quad (7.8)$$

This separation of the integral is possible because the kinetic energy depends only on the momentum vectors and the potential energy depends only on the position vectors. The momentum integral can further be separated into a product of integrals, each along a component of the momentum of particle  $i$ . For the  $x$  component of particle  $i$  we have<sup>4</sup>

<sup>2</sup>Here we assume pure component. The extension to mixtures will be discussed later.

<sup>3</sup>The way to do this is to place the first particle in every possible position in phase space; the second particle in all positions except those already taken by particle 1; and so on. In this approach the integration limits for each particle exclude the region of phase space that has already been integrated. Each particle then has its own integration limits.

### 7.3. Configuration integral

$$\int_{-\infty}^{\infty} e^{-p_{ix}^2/2mkT} dp_{ix} = (2\pi mkT)^{1/2} \quad (7.9)$$

There are  $3N$  such terms so that the partition function now becomes

$$Q = \frac{1}{N!} \left( \frac{2\pi mkT}{h^2} \right)^{3N/2} \int_V e^{-\mathcal{Q}/kT} d^{3N}r \quad (7.10)$$

We define the *configuration integral*  $Z(T, V, N)$  as the integral of the potential energy term over the volume occupied by the system,<sup>5</sup>

$$Z_N = \int_V e^{-\mathcal{Q}/kT} dr_1 dr_2 \cdots dr_N. \quad (7.11)$$

The canonical partition function then becomes

$$Q_N = \frac{1}{N!} \frac{Z_N}{\Lambda^{3N}} \quad (7.12)$$

where  $\Lambda$  is defined as

$$\Lambda = \left( \frac{h^2}{2\pi mkT} \right)^{1/2} \quad (7.13)$$

$\Lambda$  has units of length and is the thermal DeBroglie wavelength of the particle.<sup>6</sup> Classical mechanics applies when the de Broglie length is much smaller than the mean interparticle distance,

$$\left( \frac{h^2}{2\pi mkT} \right)^{1/2} \ll \left( \frac{V}{N} \right)^{1/3}.$$

This condition is generally satisfied unless temperature is very low or the particle density very high.

Returning to Eq. (7.12), we have expressed the canonical partition function in terms of two contributions: the contribution of kinetic energy, which is expressed by the factor  $1/\Lambda^{3N}$ , and the contribution of the potential energy, which is expressed by the configuration integral. The main problem in applied statistical mechanics is to obtain the configuration integral for a given intermolecular potential.

**On distinguishability** Suppose we play a game with two dice, one black and one white. Are the dice distinguishable or indistinguishable? Don't rush to say yes! It all depends on the rules of the game we are playing –the physical color of the dice has nothing to do with it. If the game is to obtain a 3 on the white die and a 4 in the black die, the dice are distinguishable; if the game is to obtain a 3 and a 4 regardless of the color of the result, the dice are *indistinguishable*. For the purposes of the game, the dice are indistinguishable if by exchanging their values (say from white = 3 and black = 4 to white = 4 and black = 3) the new result is identical according to the rules of the game.

How does this relate to statistical mechanics? We perceive oxygen and hydrogen as distinct species because we can identify them experimentally and can separate them from each other. On the other hand we cannot distinguish

<sup>5</sup>From here on we use the simplified notation  $Z_N$  and  $Q_N$  to refer the configuration integral and partition function of  $N$  particles in volume  $V$  and temperature  $T$ .

<sup>6</sup>The wavelength of a quantum particle is  $h/p$ , where  $h$  is Planck's constant and  $p = \sqrt{2mE_K}$  is the momentum of the particle. If we take the kinetic energy to be  $E_K = kT$ , a result that we will derive later, the corresponding wavelength of the particle is

$$\lambda = \frac{h}{\sqrt{2mE_K}} = \left( \frac{h^2}{2mkT} \right)^{1/2}.$$

This result differs from Eq. (7.13) by a factor  $\pi$  in the denominator. That is,  $\Lambda$  and  $\lambda$  are proportional to each other with a proportionality constant of the order of 1. In our derivation of  $\lambda$  we took the kinetic energy to be  $kT$ , which is not strictly correct; still, the correct value of the order of  $kT$  so that this detail only affects the precise value of the proportionality constant between  $\Lambda$  and  $\lambda$ .

## 7.4. Molecular partition function

two hydrogen atoms of the same isotope from each other because they have identical properties and no experiment can distinguish between the two or separate them from each other. If two physical particles are *identical* with respect to *any conceivable experiment*, they are *indistinguishable*.<sup>7</sup> In working with indistinguishable particles it is always easier to first treat all particles as distinguishable and then apply any necessary corrections due to indistinguishability. We demonstrate this with an example using the two dice.

**Example 7.1.** *Counting states*

In a game with two dice with  $N$  faces each, a “state” consists of the two numbers that turn up in a toss. Count the number of states first by taking the dice to be distinguishable, then by taking them to be indistinguishable.

**Solution**

*Case 1 Distinguishable dice* If we represent the state as an ordered pair,  $(i, j)$ , where the order indicates the die, it is easy to see that the total number of outcomes is  $N^2$ , corresponding to the total number of possible outcomes. Each of these states are distinct because no two pairs are equal.

*Case 2 Indistinguishable dice* In this case the order of the numbers in the pair does not matter. To count the states it is convenient to construct the table of all possible outcomes treating the dice as *distinguishable*. This is shown in Table 7.1. Except for states on the diagonal, all other states appear twice as  $(i, j)$  and as  $(j, i)$ . The number of distinct states is equal to the number of states in the diagonal plus one-half the number of all other states:

$$N + \frac{N^2 - N}{2} = \frac{N(N+1)}{2}$$

For large  $N$  the result scales as  $N^2/2$ . We could have obtain this result with a simple argument as follows. Treating the dice as distinguishable, the total number of states is  $N^2$ . Due to indistinguishability the multiplicity of all states is  $2!$  since permutations within a pair give the same state. Therefore the correct number of states for indistinguishable particles is  $N^2/2!$ . The calculation neglects the fact that the multiplicity of the diagonal elements is 1, however, the fraction of the elements in the diagonal is  $1/N$  and becomes vanishingly small when  $N$  is large.

The calculation of the canonical partition function follows the same logic: first calculate the configuration integral treating all particles as indistinguishable, then divide by the multiplicity factor  $N!$  to correct for indistinguishability.

### 7.4 Molecular partition function

Consider a single molecule in volume  $V$ . A very simple simple result is obtained if the molecule is represented as a point mass without internal structure. In this case the interaction potential is 0, since there are no other particles to interact with, and the configuration integral is simply,  $Z_1 = V$ . Therefore, for a single, structureless molecule,

$$Q_1 = \frac{V}{\Lambda^3} \tag{7.14}$$

	1	2	3	4	5	6
1	⊙	•	•	•	•	•
2	•	⊙	•	•	•	•
3	•	•	⊙	•	•	•
4	•	•	•	⊙	•	•
5	•	•	•	•	⊙	•
6	•	•	•	•	•	⊙

**Table 7.1:** Table of outcomes for two dice in Exercise 7.1. Each element represents an ordered pair of outcomes. If the dice are indistinguishable the number of states is equal to the total number of elements in the table. For indistinguishable dice the off-diagonal elements appear twice, their number therefore must be divided by 2.

## 7.4. Molecular partition function

---

Real molecules are made of nuclei and electrons assembled into atoms connected via bonds to other atoms. Molecules with internal structure have internal degrees of freedom that contribute to the Hamiltonian of the molecule. These contributions come from nuclear and electronic contributions, and, for multiatom molecules, from rotational and vibrational motion. To a good approximation, these degrees of freedom –with the exception of rotation– make additive contributions to the Hamiltonian of the molecule:

$$\mathcal{H} = \mathcal{H}_{\text{trans+rot}} + \underbrace{\mathcal{H}_{\text{vib}} + \mathcal{H}_{\text{nuc}} + \mathcal{H}_{\text{el}}}_{\text{internal}} \quad (7.15)$$

Here, the first term in the right hand side contains interactions among molecules (translational/rotational contribution) and the remaining terms refer to interactions *within* the molecule. Additivity means that each contribution can be treated as an independent process. Nuclear and electronic effects can be separated from the rest because they are not affected strongly by the orientation of the molecule or by the presence of other molecules in the vicinity. The same is true for vibrational contributions which are mainly determined by the bonds that hold atoms together. Rotational contributions, on the other hand, are coupled to the translational term because the potential interaction between molecules depends not only on the position of the molecules but also on their relative orientation.

Each Hamiltonian term in Eq. is a function of its own  $p$  and  $q$  coordinates that describe the corresponding degrees of freedom. Whenever the Hamiltonian can be written in such additive form, the corresponding partition function becomes a product of the individual partition functions:

$$Q_1 = Q_{\text{trans+rot}} \underbrace{Q_{\text{vib}} Q_{\text{nuc}} Q_{\text{el}}}_{Q_i} = Q_{\text{trans+rot}} + Q_{\text{int}} \quad (7.16)$$

where  $Q_{\text{int}}$  is the partition function of the *internal* degrees of freedom of the molecule. The translational-rotational partition of function of a single molecule is still given by Eq. (7.14). Therefore, the single-particle partition function is

$$Q_1 = \frac{V}{\Lambda^3} Q_{\text{int}} \quad (7.17)$$

which is more general than Eq. (7.14): In this form the partition function contains the contribution of translational energy ( $\Lambda$ ), the contribution of intermolecular interactions ( $Z_N$ ), including both translational and rotational effects, and the contribution of internal degrees of freedom ( $Q_{\text{int}}$ ). For a collection of  $N$  particles, the corresponding partition function is

$$Q_N = \frac{Z_N}{N! \Lambda^{3N}} Q_{\text{int}}^N \quad (7.18)$$

which is a more general form of Eq. (7.12). Here,  $Z_N$  contains interactions between particles, i.e., interaction between atom  $i$  in particle  $j$  and atom  $k$  in particle  $l$  through an intermolecular potential. In writing this equation we are assuming that  $Q_{\text{int}}$  accounts for interactions within a single molecule (i.e. interactions between the electrons in the molecule). Interactions between the

electrons of different molecules are taken to be part of the intermolecular potential and these are accounted for in the configuration integral,  $Z_N$ .

We note that  $Q_{\text{int}}$  is a function of temperature but not of  $N$ , since it refers to a single molecule, or  $V$ , since the subatomic particles that make up the molecule are constrained to be within the volume of the molecule (i.e., placing a molecule in a larger box does not affect the size of the molecule). Accordingly, the internal partition function makes additive contributions to properties (since thermodynamic properties are related to the logarithm of the partition function) and is relevant only when the absolute value of the property is needed. One such property is heat capacity. Clearly, internal degrees of freedom represent additional modes of energy (and thus additional capacity to store energy). With respect to other thermodynamic properties, the additive contribution is unimportant when only property changes are of interest.

## 7.5 Probability density of microstates

The canonical probability to find the system in microstate  $i$  is

$$p_i = \frac{e^{-\beta E_i}}{Q} \quad [5.27]$$

This expression assumes microstates to be discrete. The equivalent expression in the continuous phase space of  $N$  particles, is

$$dP_N = \frac{1}{N!h^{3N}} \frac{e^{-\beta H}}{Q_N} d\Gamma \quad (7.19)$$

where  $H$  is the Hamiltonian function and  $d\Gamma = d\mathbf{r}^N d\mathbf{p}^N$ . Using Eq. (7.12) and the definition of  $\Lambda$  this becomes

$$\frac{dP_N}{d\Gamma} = \frac{1}{(2\pi mkT)^{3N/2}} \frac{e^{-H/kT}}{Z_N} = W_N(\mathbf{r}^N, \mathbf{p}^N) \quad (7.20)$$

we define  $W_N(\mathbf{r}^N, \mathbf{p}^N)$  as

$$W_N(\mathbf{r}^N, \mathbf{p}^N) = \frac{1}{(2\pi mkT)^{3N/2}} \frac{e^{-H/kT}}{Z_N} \quad (7.21)$$

This is equal to  $dP_N/d\Gamma$ , i.e., it represents the probability to find the system within a volume  $d\Gamma$  of the phase space, i.e., the probability that the  $N$  particles are at positions  $(\mathbf{r}_1, \mathbf{r}_2 \dots)$  with momentum  $(\mathbf{p}_1, \mathbf{p}_2 \dots)$ . We refer to  $W_N$  as the joint probability density of  $N$  particles in the phase space, or simply, the  $N$ -particle distribution function. It is a simple exercise to show that this distribution function satisfies the normalization

$$\int_{\Gamma} W_N(\Gamma) d\Gamma = 1 \quad (7.22)$$

The average of any mechanical property  $F(\Gamma)$  can now be expressed in terms of the  $N$ -particle density function:

$$\bar{F} = \int_{\Gamma} F(\Gamma) W_N(\Gamma) d\Gamma \quad (7.23)$$

In these derivations we have dropped the internal partition function  $Q_{\text{int}}$ . The internal partition function characterizes the distribution of internal states of the particle and as we noted previously, these states depend on their own set of internal molecular coordinates which are independent of the position and momentum of the particles. With this in mind,  $W_N$  is understood to be the probability to find the  $N$ -particle system in the vicinity of microstate  $\Gamma$  *independently* of the internal microstate of the particles. Mathematically, this is equivalent to integrating over all internal coordinates.

---

*Exercise:* Express the mean potential energy in the canonical ensemble as a function of  $Z_N$  and its derivatives.

Show how the mean potential energy can be calculated from the equation of state. As a demonstration, calculate the mean potential energy for ideal-gas equation of state, and the van der Waals equation.

---

### 7.5.1 Single-particle probability density

Suppose that we tag a particle, we will call it particle 1, and follow its motion as the  $N$ -particle system moves through phase space. We wish to calculate the probability density function to find particle 1 in the vicinity of microstate  $(\mathbf{r}_1, \mathbf{p}_1)$  independently of the position and momentum of the other  $N-1$  particles. This is obtained by integrating out in Eq. (7.21) the coordinate and momentum vectors of all particles except for particle 1:

$$w_1 = \int_{\Gamma_{-1}} W_N d\Gamma_{-1} \quad (7.24)$$

where  $\Gamma_{-1}$  stands for  $(\mathbf{r}_2, \mathbf{r}_3, \dots; \mathbf{p}_2, \mathbf{p}_3, \dots)$ . Integrating over the  $3(N-1)$  momentum coordinates we obtain<sup>8</sup>

$$w_1 = \frac{e^{-(p_{1x}^2 + p_{1y}^2 + p_{1z}^2)/2mkT} \int_V e^{-U/kT} d\mathbf{r}_2 d\mathbf{r}_3 \dots}{(2\pi mkT)^{3/2} Z_N} \quad (7.25)$$

<sup>8</sup>Recall that each momentum component integrates to  $(2\pi mkT)^{1/2}$  and there are  $3(N-1)$  such components.

The integral on the right-hand side is easily related to the configuration integral. To obtain this relationship, we begin with the configuration integral and separate the integration with respect to the position of particle 1:

$$Z_N = \int_V d\mathbf{r}_1 \int_V e^{-U/kT} d\mathbf{r}_2 d\mathbf{r}_3 \dots \quad (7.26)$$

The inner integral represents interactions over all configurations in space for a fixed position of particle 1. This interaction, however, does not depend on the position of particle 1, therefore, the two integrals are separable and the result is

$$Z_N = V \int_V e^{-U/kT} d\mathbf{r}_2 d\mathbf{r}_3 \dots \quad (7.27)$$

With this result, Eq. (7.25) takes the form<sup>9</sup>

$$w_1 = \frac{e^{-(p_x^2 + p_y^2 + p_z^2)/2mkT}}{V(2\pi mkT)^{3/2}} \quad (7.28)$$

Here, we have dropped the subscript 1 from the momentum on the right-hand side since particle 1 can be any of the  $N$  particles. This result states some very basic facts: the probability distribution of momentum along any direction is the Gaussian function,

$$f(p) = \frac{e^{-p/2mkT}}{\sqrt{2\pi mkT}} \quad (7.29)$$

Its mean is zero (i.e. the particle has equal probability to be moving to the left or to the right) and the standard deviation is  $\sqrt{mkT}$ , which implies that the distribution becomes more wide with increasing temperature and increasing particle mass. The probability distribution of position (independently of momentum) is

$$h(x, y, z) = \frac{1}{V} \quad (7.30)$$

This distribution is uniform in space, that is, the particle has equal probability to be found anywhere. Because these distributions are independent of each other, the one-particle distribution function is given by the product

$$w_1 = f(p_x)f(p_y)f(p_z)h(x, y, z) \quad (7.31)$$

Finally, we note that these distributions are independent of the intermolecular potential. Accordingly, these results are valid for any molecular system, regardless of interaction.

---

*Exercise:* In a crystalline solid the probability to find a particle in space is *not* uniform because particles can be found only in the vicinity of the crystallographic lattice sites. What part of the above derivations has to be revised to account for a crystalline phase correctly?

*Exercise:* Obtain the distribution of velocities  $v_x$  in the  $x$  direction and the probability of  $v = (v_x^2 + v_y^2 + v_z^2)^{1/2}$ .

---

## 7.6 Ideal gas

Consider a collection of point particles that experience no interaction whatsoever. This amounts to setting  $U = 0$ . The configuration integral with respect to  $dr_{jx}dr_{jy}dr_{jz}$  spans the volume available to particle  $j$ ; since we have assumed “point” particles, the result of this integration is  $V$ , the entire volume of the system. Each of the  $N$  particles gives rise to the same result, thus finally, the canonical partition function is

$$Q_N = \frac{V^N Q_{\text{int}}^N}{N! \Lambda^{3N}} = \frac{1}{N!} \left( \frac{V Q_{\text{int}}}{\Lambda^3} \right)^N \quad (7.32)$$

This situation can be visualized as the limit of any type of particles at very low density. Then, the vast majority of configurations involve particles at large distances from each other. Such particles experience no interaction; moreover, given the large interparticle distances, molecules can indeed be treated as point particles regardless of the details of their molecular structure. This represents the so-called ideal-gas limit.

First, we examine the log of the partition function. Since  $\log Q = -A/kT$ , we must show that  $\ln Q$  is an extensive property. Taking the logarithm of Eq. (7.32) we have:

$$\ln Q_N = N(\ln V + \ln Q_{\text{int}} - 3 \ln \Lambda) - \ln(N!)$$

Using Sterling's formula,  $\ln N! = N(\ln N - 1)$ , the above becomes

$$\ln Q_N = N \left( \ln \frac{V}{N} + \ln Q_{\text{int}} - 3 \ln \Lambda + 1 \right)$$

Notice that everything inside the parenthesis involves intensive variables, namely,  $V/N$  (volume per molecule) and  $T$ , which is part of  $\Lambda$  and  $Q_{\text{int}}$ . This means that  $\ln Q/N$  is an intensive property, or, that  $\ln Q$  is extensive. Therefore, the Helmholtz energy that is obtained from the ideal-gas partition function in Eq. (7.32) is indeed an extensive property:

$$A = -NkT \left( \ln \frac{V}{N} + \ln Q_{\text{int}} - 3 \ln \Lambda + 1 \right) \quad (7.33)$$

To obtain the equation of state, we apply the relationship,

$$P = - \left( \frac{\partial A}{\partial V} \right)_{T,N}$$

to Eq. (7.33). Noting that the only term that depends on  $V$  is  $\ln V/N$ , we find

$$P = \frac{NkT}{V} \quad (7.34)$$

Thus we have obtained the ideal-gas equation of state. Since the molar volume,  $v$  is  $v = VN_{\text{AV}}/N$ , where  $N_{\text{AV}}$  is Avogadro's number, the above can also be written as

$$P = \frac{N_{\text{AV}}kT}{v} \quad (7.35)$$

Comparing with the classical result,  $P = RT/v$  we conclude

$$R = N_{\text{AV}}k \quad (7.36)$$

and thus we have related Boltzmann's constant, which up to this point was merely a factor, to the ideal-gas constant.

---

*Exercise:* Obtain the constant-pressure heat capacity based on the partition function in Eq. (7.32).

---

### 7.6.1 Ideal gas treatment in the grand canonical ensemble

The ideal-gas state can also be treated within the grand canonical ensemble. To demonstrate how different ensembles lead to identical results, we present the grand canonical treatment here. First, we write the canonical partition function in Eq. (7.32) as

$$Q_N = \frac{V^N Q_{\text{int}}^N}{N! \Lambda^{3N}} = \frac{Q_1^N}{N!} \quad (7.37)$$

where  $Q_1$  is the partition function of a single molecule, defined in Eq. (7.17):

$$Q_1 = \frac{V}{\Lambda^3} Q_{\text{int}} \quad [7.17]$$

Next, expressing the grand canonical partition function as a summation of canonical partition functions:<sup>10</sup>

<sup>10</sup>See Problem 5.5.

$$\Xi = e^{PV/kT} = \sum_{N=0}^{\infty} \frac{(Q_1 \lambda)^N}{N!} \quad (7.38)$$

where

$$\lambda = e^{\mu/kT} \quad (7.39)$$

We recognize the summation on the right hand side of Eq. (7.38) as the series expansion of  $\exp(Q_1 \lambda)$ . Thus,

$$\Xi = e^{PV/kT} = e^{Q_1 \lambda} \quad (7.40)$$

from which

$$\ln \Xi = Q_1 \lambda = \frac{PV}{kT} \quad (7.41)$$

Finally, using Eq. (5.67), for the average number of particles in  $V$  we find

$$\bar{N} = \lambda \left( \frac{\partial \ln \Xi}{\partial \lambda} \right)_{V,T} = Q_1 \lambda = \frac{PV}{kT} \quad (7.42)$$

where we treated  $Q_1$  as constant with respect to  $\lambda$  at constant  $T$ ,  $V$ , based on our earlier observation that  $Q_1$  depends on  $V$ ,  $T$ , only. The above result is rearranged to read

$$PV = \bar{N} kT \quad (7.43)$$

Thus we have recovered the ideal-gas law, except that now it is the *mean* number of particles that appears. As we discussed previously, even though the number of particles in the grand canonical ensemble is allowed to vary, fluctuations are very narrowly distributed about the mean. In this sense, Eqs. (7.35) and (7.43) are the same.

**On the ideal-gas constant** The name “ideal-gas constant” is a misnomer in the sense that it is *not* specific to the ideal-gas state—it is a universal physical parameter like the Boltzmann constant or the Avogadro number. The name is used for historical reasons, as the value of  $R$  was determined experimentally as the limiting value of the ratio  $Pv/T$  at low pressures. In our development of statistical mechanics we have seen that the appearance of  $k$  (which is nothing

but the ideal-gas constant expressed per molecule instead of mole) is in no special way linked with ideal gases. In fact,  $k$  is linked to temperature: The fundamental quantity in statistical thermodynamics is the product  $kT$ , which has units of energy, and not  $k$  or  $T$  alone. Since  $T$  has been given dimensions that do not conform to any system of units,<sup>11</sup>  $k$  represents the conversion factor which, when multiplied by  $T$ , produces units of energy. In a rational system of units, entropy would be dimensionless, temperature would have units of energy,  $k$  would be equal to 1, and  $R$  would be equal to Avogadro's number. The downside though is that such rational system would deprive Boltzmann of a constant named to honor his contributions to rational thermodynamics.

<sup>11</sup>Temperature units such °C, °F, etc. are arbitrarily associated with the linear expansion of a column of mercury or other medium as recorded when this medium is transported from freezing water into boiling water. The units of kelvin or rankine are no more rational; they merely shift the zero of the measurement scale.

### 7.6.2 Other properties in the ideal-gas state

Let us apply the relationship

$$S = - \left( \frac{\partial A}{\partial T} \right)_{N,V} \quad (7.44)$$

to Eq. (7.33). Noting that  $\ln \Lambda = -\ln T^{1/2} + \ln(h^2/2\pi mk)^{1/2}$ , the result is

$$S = Nk \left( \ln \frac{V}{N} + \ln Q_{\text{int}} - 3 \ln \Lambda + 1 \right) + NkT \left( \frac{3}{2T} + \frac{d \ln Q_{\text{int}}}{dT} \right) \quad (7.45)$$

Recognizing the first term in parenthesis on the right-hand side as  $-A/T$ , also equal to  $(TS - U)/T$ , the above can be solved for the internal energy:

$$U = NkT \left( \frac{3}{2} + T \frac{d \ln Q_{\text{int}}}{dT} \right) \quad (7.46)$$

As we expect from classical thermodynamics,  $U$  depends only on  $T$ . It follows that the constant-volume heat capacity is

$$C_V = Nk \left( \frac{3}{2} + \frac{\partial}{\partial T} T^2 \frac{\partial \ln Q_{\text{int}}}{\partial T} \right) \quad (7.47)$$

For a monoatomic gas, nuclear and electronic contributions to the partition function do not depend strongly on temperature and since vibrational modes are not present,  $dQ_{\text{int}}/dT = 0$ . The heat capacity then is equal to  $3Nk$ , which reflects the three translational degrees of freedom. We cannot say much for molecules that contain two or more atoms because the partition function we have been working on includes only translational motion but no rotation, bending, etc. We expect, however, that adding degrees of freedom (rotation, bending) makes additional contributions to heat capacity.

We now return to Eq. (7.45) and make use of Eq. (7.47) to write the entropy as

$$\frac{S}{N} = k \ln \frac{V}{N} + \int_{T_0}^T \frac{C_V/N}{T} dT + \left[ \frac{5}{2} + \ln \Lambda - \frac{d(T \ln Q_{\text{int}})}{dT} \right]_{T=T_0}$$

where  $T_0$  is some arbitrary temperature.<sup>12</sup> Despite the complicated appearance,

everything past the integral in  $T$  is a constant. To bring this to a more familiar form, we calculate  $S_2 - S_1$  between states  $(T_1, V_1, N)$  and  $(T_2, V_2, N)$ , respectively:

$$\frac{S_2 - S_1}{N} = \int_{T_1}^{T_2} \frac{C_V}{T} dT + k \ln \frac{V_2/N}{V_1/N} \quad (7.48)$$

which is the familiar expression from classical thermodynamics.

### 7.6.3 Non-ideal systems

The partition function in the presence of interactions can be expressed in terms of the ideal-gas partition function:

$$Q_N = Q_N^{\text{ig}} \frac{Z_N}{V^N} \quad (7.49)$$

where

$$Q^{\text{id}} = \frac{1}{N!} \left( \frac{V}{\Lambda^3} Q_{\text{int}} \right)^N \quad (7.50)$$

Taking the logarithm in both sides of Eq. (7.49),

$$\ln Q_N = \ln Q_N^{\text{id}} + \ln(Z_N/V^N)$$

Since  $A = -kT \ln Q$  and  $A^{\text{ig}} = -kT \ln Q^{\text{ig}}$ , we set

$$A^r = -kT \ln(Z_N/V^N) \quad (7.51)$$

Here,  $A^r$  plays the role of a correction that must be added to the ideal-gas part of the Helmholtz energy for deviations from ideality. These deviations are solely due to molecular interactions and are captured in the configuration integral. In this respect,  $A^r$  is a *residual* property. Notice that the usual “residual” property is defined as a correction to the ideal-gas state at the pressure and temperature of the system.<sup>13</sup>

## 7.7 Virial expansion

<sup>14</sup>The virial equation is a series expansion of the equation of state in terms of density:

$$\frac{P}{kT} = \rho + B_2 \rho^2 + B_3 \rho^3 + \dots \quad (7.52)$$

where  $\rho = N/V$  is the particle density and  $B_i$  is the  $i^{\text{th}}$  virial coefficient. For a given substance are functions of temperature only. The first term of the expansion corresponds to the ideal gas state ( $\rho \rightarrow 0$ ). The higher coefficients represent correction factors that account for interparticle interactions. In a mathematical sense, the virial equation is a Taylor expansion of the compressibility factor ( $PV/\rho kT$ ) in powers of  $\rho$  but the significance of the expansion lies in the fact that the virial coefficients are directly related to intermolecular forces. For simplicity, we will derive the virial equation of a pure monoatomic gas.

<sup>13</sup>The usual residual property is defined as

$$A^R(T, P) = A(T, P) - A^{\text{ig}}(T, P)$$

The residual property obtained here is defined

$$A^r(T, P) = A(T, P) - A^{\text{ig}}(T, RT/V)$$

where  $V$  is the volume of the system at  $T, P$ .

<sup>14</sup>This discussion can be found in Hill on p. 261; also in Tester and Modell on p. 415.

First we write the grand partition function in the form

$$\Xi = \sum_{N=0}^{\infty} Q_N \lambda^N = 1 + \sum_{N=1}^{\infty} Q_N \lambda^N \quad (7.53)$$

To obtain this result we used the fact that  $Q_0 = 1$  since with  $N = 0$  the only possible microstate is one with zero energy. For the canonical partition function we use Eq. (7.17), which we write as

$$Q_N = \frac{Z_N}{N!} \left( \frac{Q_1}{V} \right)^N \quad (7.54)$$

For convenience we define the quantity  $\rho_0$  as follows

$$\rho_0 = \frac{Q_1 \lambda}{V} \quad (7.55)$$

and note that it has units of 1/volume. Combining Eqs. (7.53–7.55) the grand canonical partition function becomes

$$\Xi = e^{PV/kT} = 1 + \sum_{N=1}^{\infty} \frac{Z_N}{N!} \rho_0^N \quad (7.56)$$

We take the logarithm of both sides, expand the right-hand side,<sup>15</sup> divide both sides by  $V$  and collect the result in powers of  $\rho_0$ . This produces the series expansion

$$\frac{P}{kT} = \rho_0 + \underbrace{\left( \frac{Z_2 - Z_1^2}{2!V} \right)}_{b_2} \rho_0^2 + \dots = b_1 \rho_0 + b_2 \rho_0^2 + \dots \quad (7.57)$$

with

$$b_1 = 1; \quad b_2 = \frac{Z_2 - Z_1^2}{2!V}; \quad \dots \quad (7.58)$$

Thus we have an expansion in terms of the modified density  $\rho_0$ . We must now relate  $\rho_0$  to  $\rho$ . To obtain this relationship we note that

$$\bar{N} = \left( \frac{\partial \ln \Xi}{\partial \ln \lambda} \right)_{TV} = \left( \frac{\partial (PV/kT)}{\partial \ln \rho_0} \right)_{TV} \Rightarrow \rho = \rho_0 \left( \frac{\partial (P/kT)}{\partial \rho_0} \right)_{TV} \quad (7.59)$$

Here, we made use of the fact that  $(\partial \ln \lambda)_{TV} = (\partial \ln \rho_0)_{TV}$ , which follows from the definition of  $\rho_0$ . Next, we use Eq. (7.57) to evaluate the derivative in  $\rho_0$ , Eq. (7.59):

$$\rho = b_1 \rho_0 + 2b_2 \rho_0^2 + \dots \quad (7.60)$$

This can be inverted to produce a series expansion of  $\rho_0$  in terms of  $\rho$ . First, we write

$$\rho_0 = a_1 \rho + a_2 \rho^2 + \dots \quad (7.61)$$

<sup>15</sup>Use the expansion

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$$

with  $x = \sum Z_N \rho_0^N / N!$ .

whose coefficients  $a_n$  are to be determined. We insert this series into Eq. (7.60), expand the powers and collect terms in powers of  $\rho$ :

$$\rho = \underbrace{a_1}_{=1} \rho + \underbrace{(a_2 + 2b_2 a_1^2)}_{=0} \rho^2 + \dots \quad (7.62)$$

Since the left-hand side is equal to  $\rho$ , the coefficient of the linear term on the right-hand side must be 1 and all other coefficients must be 0. This establishes the unknown coefficients  $a_n$  as follows:

$$a_1 = 1; \quad a_2 = -2b_2; \quad \dots \quad (7.63)$$

The last step is to insert Eq. (7.61) into (7.57) and collect the result in powers of  $\rho$ :

$$\frac{P}{kT} = \underbrace{a_1 b_1}_{=1} \rho + \underbrace{(b_1 a_2 + b_2 a_1^2)}_{=-b_2} \rho^2 + \dots \quad (7.64)$$

or, finally,

$$\frac{P}{kT} = \rho + B_2 \rho^2 + \dots \quad (7.65)$$

with

$$B_2 = \frac{Z_1^2 - Z_2}{2V} \quad (7.66)$$

The higher order coefficients can be obtained by tracking the higher order powers in the expansions. Here, however, we will focus on the second virial coefficient only.

**Second virial coefficient** The configuration integrals of order 1 and 2 take the form,

$$Z_1 = V \quad (7.67)$$

$$Z_2 = \int \int e^{-\phi_{12}/kT} d\mathbf{r}_1 d\mathbf{r}_2 \quad (7.68)$$

where  $\phi_{12} = \phi_{12}(|\mathbf{r}_1 - \mathbf{r}_2|)$  is the interaction potential between two particles and  $\mathbf{r}_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$  is the interparticle distance.<sup>16</sup> First we write the term  $Z_1^2$  as a double integral in  $d\mathbf{r}_1 d\mathbf{r}_2$

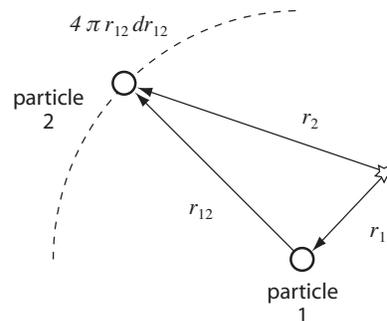
$$Z_1^2 = V^2 = \int \int d\mathbf{r}_1 d\mathbf{r}_2 \quad (7.69)$$

and then express the numerator of Eq. (7.66) in the form

$$Z_2 - Z_1^2 = \int_V \left(1 - e^{-\phi_{12}/kT}\right) d\mathbf{r}_1 d\mathbf{r}_2 \quad (7.70)$$

To evaluate this integral we change the integration variables from  $\mathbf{r}_1, \mathbf{r}_2$  to  $\mathbf{r}_1, \mathbf{r}_{12} = \mathbf{r}_2 - \mathbf{r}_1$ . The spherical symmetry of the problem provides the following relationship between the original and the transformed variables<sup>17</sup>

$$d\mathbf{r}_1 d\mathbf{r}_2 = d\mathbf{r}_1 d\mathbf{r}_{12} = (4\pi r_1^2 dr_1)(4\pi r_{12}^2 dr_{12}) \quad (7.71)$$



**Figure 7.4:** Geometry for the calculation of the second virial coefficient.

<sup>16</sup>We are treating the particles as spheres, i.e., the interaction depends only on the center-to-center distance between the spheres. For more complicated molecular structures the potential interaction also depends on the relative orientation of the molecules and their internal configuration (if molecules are not rigid objects).

<sup>17</sup>In spherical coordinates,  $d\mathbf{r} = (r^2 \sin\phi) d\phi d\theta dr$ , with  $\phi \in [0, \pi]$  and  $\theta \in [0, 2\pi]$ . For spherically symmetric systems, integration over  $\theta$  and  $\phi$  gives  $2\pi$ , therefore, integrations in  $d\mathbf{r}$  are converted into integrations in  $4\pi r^2 dr$ .

where  $dr$  refers to the magnitude of the vector  $d\mathbf{r}$ . Using this substitution, we have

$$Z_2 - Z_1^2 = \int_V (4\pi r_1^2 dr_1) \int_V 4\pi r_{12}^2 (1 - e^{-\phi_{12}/kT}) dr_{12} = 4\pi V \int_0^\infty (1 - e^{-\phi_{12}/kT}) r_{12}^2 dr_{12} \quad (7.72)$$

Substituting this result in Eq. (7.66), the second virial coefficient is obtained in the form

$$B_2 = 2\pi \int_0^\infty (1 - e^{-\phi_{12}(r)/kT}) r^2 dr \quad (7.73)$$

where for simplicity we have dropped the subscript  $_{12}$  from  $r$ . The important conclusion is that the second virial coefficient can be computed if the pair interaction between two particles is a known function of distance.

## 7.8 van der Waals equation

The formulation of the van der Waals equation is an example of how molecular concepts can be incorporated into an equation of state. Still, this can only be done approximately since the configuration integral cannot be calculated exactly. The approximation that allows this calculation to proceed is to replace the interactions among  $N$  particles with the average interaction of one particle with the rest of the system.<sup>18</sup> We will be using a pair potential that is a superposition of a repulsive hard core and an van der Waals attraction:

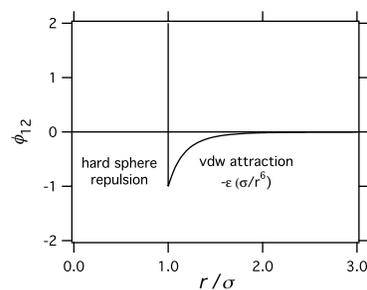
$$\phi_{12} = \begin{cases} \infty & \text{if } r < \sigma \\ -\epsilon(\sigma/r)^6 & \text{if } r > \sigma \end{cases} \quad (7.74)$$

This potential is shown in figure 7.5. The potential has a minimum at  $r = \sigma$  with depth  $\epsilon$ . The attraction decays very rapidly with distance as a result of the  $1/r^6$  dependence. For example, at  $r = 2\sigma$  the attraction is 1.5% of its maximum value and at  $r = 5\sigma$  it is less than 0.01%. Despite its small range, this interaction is sufficient to lead to a vapor-liquid transition.

The potential interaction among  $N$  molecules is calculated by adding pair interactions among all particle pairs:

$$U = \frac{1}{2} \sum_i^N \sum_{j \neq i}^N \phi_{12}(r_{ij}) \quad (7.75)$$

where  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ . The potential interaction  $U$  depends on the spatial arrangement of  $N$  particles in volume  $V$ , which precisely is the difficulty in calculating the configuration integral. We will simplify the problem by considering some average interaction<sup>19</sup> in which the spatial configurations have been averaged out. To calculate this interaction, we tag a particle and calculate its interaction with the remaining  $N - 1$  particles in the system. Consider the volume of a spherical shell of thickness  $dr$  at distance  $r$  from the tagged particle. All particles located in that shell have the same strength of interaction with the tagged particle. Assuming



**Figure 7.5:** Intermolecular potential for the van der Waals equation.

<sup>18</sup>This section is based on Hill (1986, p. 286).

<sup>19</sup>We use the term “average” loosely to refer to an estimate of the potential interaction. A proper mean must be calculated based on a probability distribution function.

the density of the fluid to be uniform and equal to the bulk density<sup>20</sup>  $N/V$  all around the tagged particle, the number of particles inside the spherical shell is  $(4\pi r^2 dr)(N/V)$  particles in the spherical shell therefore, the contribution to the mean potential is

$$d\tilde{\phi} = -\epsilon \left(\frac{\sigma}{r}\right)^6 (4\pi r^2) \left(\frac{N}{V}\right) \quad (7.76)$$

Integrating in  $r$  from  $\sigma$  to  $\infty$ , we obtain the mean potential

$$\tilde{\phi} = -\frac{4\pi\epsilon\sigma^3 N}{3V} = -2a\frac{N}{V} \quad (7.77)$$

Constant  $a$  introduced for convenience and is defined as

$$a = \frac{2\pi\epsilon\sigma^3}{3} = 4\epsilon V_p \quad (7.78)$$

where  $V_p = \pi\sigma^3/6$  is the volume of the particle. The mean potential interaction among  $N$  particles is

$$\tilde{U} = \frac{N\tilde{\phi}}{2} = -a\frac{N^2}{V} \quad (7.79)$$

with the factor of  $1/2$  accounting for the fact that each particle is counted twice, once as the tagged particle and once as one of the  $N-1$  untagged particles. The Corresponding approximation for the configuration integral is

$$\tilde{Z}_N = \int e^{-\tilde{U}/kT} d\mathbf{r}^{3N} = \int e^{aN^2/kTV} d\mathbf{r}^{3N} \quad (7.80)$$

More specifically, the configuration integral must be written as

$$\tilde{Z}_N = \begin{cases} \int e^{aN^2/kTV} d\mathbf{r}^{3N}, & \text{if } r_{ij} > \sigma \text{ for all } i, j \\ 0, & \text{otherwise} \end{cases}$$

This implies that integration must be performed not over all volume but over the *free* volume of the system, i.e., by excluding the hard sphere volume of the particles:

$$\tilde{Z}_N = \int_{V_F} e^{aN^2/kTV} d\mathbf{r}^{3N} = e^{aN^2/kTV} \int_{V_F} d\mathbf{r}^{3N} \quad (7.81)$$

Here,  $V_F = V - NV_E$  is the free volume, and  $V_E$  is the excluded volume per particle. The excluded volume represents volume that is inaccessible to the particles due to hard sphere repulsion. The last integral on the right-hand side gives  $V_F^N$  so that the configuration integral becomes

$$\tilde{Z}_N = e^{aN^2/kTV} (V - NV_E)^N \quad (7.82)$$

The final step is an approximation that allows us to calculate the excluded volume per particle. Consider two hard spheres in volume  $V$ . We fix the position of one first sphere and let the other move around. The volume that is spanned by the vector  $\mathbf{r}_{12}$  represents the free volume; this is shown by the shaded area in fig. 7.6. This excludes a volume equal to the the volume of a sphere with diameter

$2\sigma$ . This is the excluded volume of the two particle system. The excluded volume per particle is one half of the volume of this sphere:

$$V_E = \frac{1}{2} \left( \frac{\pi(8\sigma^3)}{6} \right) = 4V_p \quad (7.83)$$

where  $V_p = \pi\sigma^3/6$  is the volume of a single particle. For a system of  $N$  particles this calculation is not quite as simple. Thus we introduce the simplifying assumption that the excluded volume per particle in an  $N$  particle system is the same as in a two-particle system. With this, the configuration integral takes the closed form,

$$\tilde{Z}_N = (V - 4V_p)^N e^{aN^2/kTV} = (V - Nb)^N e^{aN^2/kTV} \quad (7.84)$$

where  $b$  is introduced for simplicity and is defined as

$$b = 4V_p \quad (7.85)$$

We now have the partition function in the form:

$$Q_N = \frac{Z_N}{N! \Lambda^{3N}} Q_{\text{int}}^N = \frac{(V - 4NV_p)^N e^{aN^2/kTV} Q_{\text{int}}^N}{N! \Lambda^{3N}} \quad (7.86)$$

To obtain the equation of state we calculate the pressure:

$$P = kT \left( \frac{\partial \ln Q_N}{\partial V} \right)_{N,T} \quad (7.87)$$

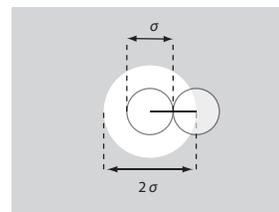
which leads to the familiar result

$$P = \frac{kT}{V/N - b} - a \left( \frac{N}{V} \right)^2 \quad (7.88)$$

The significance of this equation is that it predicts phase transition, a behavior that can be traced in the presence of a combination of attraction and repulsion in the interaction between molecules.

## 7.9 Molecular theory of corresponding states

The principle of corresponding states is a statement of the experimental observation that the  $PVT$  graph of pure substances can be scaled so as to collapse the graph of any substance onto a single, universal graph. This collapse is demonstrated by scaling  $P$ ,  $v$  and  $T$  by a set of characteristic values  $P^*$ ,  $v^*$ ,  $T^*$ , respectively. Usually we select these to be the corresponding properties at the critical point though other choices of the scaling parameters are also possible. We refer to this observed scaling to as a “principle”, it states an *approximate* in the sense that the collapse of  $PVT$  graphs is not perfect. Still, the  $PVT$  graph of certain families of molecules collapse so well that this principle must be more than a coincidence. Indeed, the basis of this behavior lies in the intermolecular potential.



**Figure 7.6:** The excluded volume of two particles is equal to the volume of a sphere with diameter  $2\sigma$ .

Suppose that the potential interaction for a family of molecules can be described by a function of the form:

$$\phi(r) = \epsilon f(r/\sigma) \quad (7.89)$$

where  $\epsilon$  (with units of energy) and  $\sigma$  (with units of length) are parameters that characterize the molecule. In other words, the functional form of the potential is the same for all molecules while the scaling parameters  $\epsilon$  and  $\sigma$  are specific to a molecule. The Lennard-Jones potential in Eq. (7.4) and the hard-sphere/van der Waals potential in Eq. (7.74) both conform to this two-parameter scaling. The total potential energy of a configuration of  $N$  particles is<sup>21</sup>

$$U = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \phi_{ij} \quad (7.90)$$

where  $\phi_{ij} = \phi_{12}(r_{ij})$  and  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ . The configuration integral is

$$Z_N = \int_V \exp\left[-\frac{U}{kT}\right] = \int_V \exp\left[-\frac{1}{2} \frac{\sum_i \sum_j \phi_{12}(r_{ij})}{kT}\right] d\mathbf{r}_1 \cdots d\mathbf{r}_N \quad (7.91)$$

Next, we make all distances dimensionless using  $\sigma$  as the characteristic length:

$$d\mathbf{r}_i = \sigma^3 d\mathbf{x}_i; \quad r_{ij} = \sigma x_{ij} \quad (7.92)$$

With these definitions and Eq. (7.89), the configuration integral becomes

$$Z_N = \sigma^{3N} \int_{V/\sigma^3} \exp\left[-\frac{\epsilon}{2kT} \sum_i \sum_j f(x_{ij})\right] d\mathbf{x}_1 \cdots d\mathbf{x}_N \quad (7.93)$$

We recognize the right-hand side is a function of  $\sigma^{3N}$ , which is factored outside the integral, of  $\epsilon/kT$ , which appears in the integrand, and of  $V/\sigma^3$ , which appears in the integration limits. We express these dependencies by writing

$$Z_N = \sigma^{3N} z_N(V/\sigma^3, kT/\epsilon) \quad (7.94)$$

where  $z$  is some function of  $V/\sigma^3$  and  $kT/\epsilon$  which must also depend on  $N$ . Since the logarithm of  $Q$  must be an extensive function, we further conclude that  $\zeta_N$  must be of the form

$$z_N = \left[ \zeta\left(\frac{V}{N\sigma^3}, \frac{kT}{\epsilon}\right) \right]^N \quad (7.95)$$

Here,  $\zeta$  is a function of *intensive* variables; accordingly, the logarithm of  $z_N$  is  $N \ln \zeta$ , which is indeed extensive. With this, the partition function takes the form:

$$Q_N = (\sigma^{3N} \zeta^N) \frac{Q_{\text{int}}^N}{N! \Lambda^{3N}} \quad (7.96)$$

The above form of the partition function has been obtained by dimensional argument based on the assumed scaling form of the intermolecular potential.

To obtain the equation of state we calculate pressure from the canonical partition function:

$$P = - \left( \frac{\partial A}{\partial V} \right)_{NT} = kT \left( \frac{\partial \ln Q_N}{\partial V} \right)_{NT} \quad (7.97)$$

<sup>21</sup>The factor 1/2 is needed because each pair in this summation is counted twice. This summation can also be written as

$$\phi(r) = \sum_{i=1}^N \sum_{j=1+1}^N \phi_{ij}$$

Here the factor 1/2 has been taken out because this summation is constructed in such a way that each pair interaction is counted only once.

We notice that the only factor on the right-hand side of Eq. (7.96) that depends on volume is  $\zeta^N$ . Therefore,

$$P = NkT \left( \frac{\partial \ln \zeta}{\partial V} \right)_{TN} = NkT \underbrace{\left( \frac{\partial \ln \zeta}{\partial (V/N\sigma^3)} \right)_{TN}}_{F(V/N\sigma^3, kT/\epsilon)} \left( \frac{1}{N\sigma^3} \right) \quad (7.98)$$

where  $F$  is a function of  $V/N\sigma^3$  and  $kT/\epsilon$ . We introduce the reduced properties,  $\tilde{P} = P/P^*$ ,  $\tilde{T} = T/T^*$ ,  $\tilde{v} = V/Nv^*$ , where:

$$P^* = \epsilon/\sigma^3; \quad T^* = \epsilon/k; \quad v^* = \sigma^3 \quad (7.99)$$

Eq. (7.98) takes the form:

$$\tilde{P} = k\tilde{T}F(\tilde{v}, \tilde{T}) \quad (7.100)$$

that is, the scaled pressure  $\tilde{P}$  is a function of the scaled temperature,  $\tilde{T}$  and scaled volume,  $\tilde{v}$ . This is precisely the type of scaling in the corresponding-states principle. The fact that the scaling does not involve the critical parameters is not a problem. It should be a simple matter to show that  $P^*$  and  $T^*$  may be replaced by *any* function of  $\epsilon/\sigma^3$  and  $\epsilon/k$ , respectively. Since<sup>22</sup>  $T_c$  is proportional to  $\epsilon/k$  and  $P_c$  proportional to  $\epsilon/\sigma^3$ , the dimensionless temperature and pressure in Eq. (7.100) may as well be defined as the usual reduced properties with respect to the critical point.

<sup>22</sup>Clearly,  $P_c$  must be a function of  $\epsilon$  and  $\sigma$ . For dimensional reasons such function must be of the form  $P_c = (\epsilon/\sigma^3)a(\epsilon, \sigma^3)$  where  $a$  is dimensionless. Since  $\epsilon$  and  $\sigma$  both have dimensions,  $a$  must be a pure number, therefore, the critical pressure is proportional to  $T^*$  in Eq. (7.99). The same logic shows that  $T_c$  is proportional to  $\epsilon/k$ .

## 7.10 Additional reading

1. This chapter is largely based on (Hill, 1986), chapters 4, 6 and 15.
2. McQuarrie (2001) follows a very similar approach to that of Hill in a somewhat more readable form (chapters 5 and 12).

## 7.11 Problems

**Problem 7.1.** a) Obtain an expression for the density fluctuation in the grand canonical ensemble and study its magnitude in the limit  $N \rightarrow \infty$ . *Hint:* Express the ratio  $\text{var}(\rho)/\langle \rho \rangle^2$  in terms of the the isothermal compressibility, defined by

$$\kappa_T = -\frac{1}{V} \left( \frac{\partial V}{\partial P} \right)_{T,N} = \frac{1}{\rho} \left( \frac{\partial \rho}{\partial P} \right)_T$$

where  $\rho = N/V$  is the density.

- b) Show that for an ideal gas the variance of  $N$  is  $N$ .
- c) Calculate  $\delta\rho/\rho$  of ambient air in a volume  $1 \text{ cm}^3$ .
- d) Determine the volume  $V$  of ambient air in which the magnitude of fluctuations is of the same order as the density itself. If this volume is a cube with linear dimension  $L$ , report the value of  $L$ .

**Problem 7.2.** a) The residual configurational Helmholtz energy is the configurational energy of  $N$  interacting particles relative to that in the ideal-gas state:

$$A^R = -kT \ln \frac{Z_N}{V^N}$$

Show that this can also be expressed in the form:

$$A^R = kT \ln \int \cdots \int \exp[\beta U(\mathbf{r}^N)] p(\mathbf{r}^N) d\mathbf{r}^N \quad (1)$$

where  $U(\mathbf{r}^N)$  is the potential energy of configuration  $\mathbf{r}^N$  and  $p(\mathbf{r}^N)$  is the probability of configuration, given by

$$p(\mathbf{r}^N) = \frac{e^{-\beta U(\mathbf{r}^N)}}{Z_N}$$

b) Another way to write Eq. (1) is

$$e^{A^R/kT} = \langle e^{U/kT} \rangle$$

i.e. the free energy is expressed as an ensemble average of the expression  $\exp(\beta U)$ .

**Problem 7.3.** a) Calculate the mean potential energy of a system of  $N$  particles in the canonical ensemble.

b) Explain how you can calculate the mean potential energy of a fluid if the equation of state is known. As a demonstration, calculate the mean potential energy using the van der Waals equation of state.

**Problem 7.4.** a) Calculate the mean velocity, the most probable velocity and the mean squared velocity of the Maxwell Boltzmann distribution.

b) In our derivation of the Maxwell Boltzmann distribution we obtained the following intermediate result

$$f(\mathbf{v}) d\mathbf{v} = \left( \frac{m}{2\pi kT} \right)^{3/2} \exp \left[ -\frac{mv^2}{2kT} \right] d\mathbf{v}$$

where  $\mathbf{v} = (v_x, v_y, v_z)$  and  $d\mathbf{v} = dv_x dv_y dv_z$ . Starting with this result obtain the distribution of velocities in the  $x$  direction. Then calculate  $\bar{v}_x$ ,  $\bar{v}_y$ ,  $\bar{v}_z$  and  $\overline{v_x^2}$ ,  $\overline{v_y^2}$ ,  $\overline{v_z^2}$ .

c) Derive the distribution of kinetic energy,  $f(\epsilon)$ , of a system of molecules at thermal equilibrium. Note:  $f(\epsilon) d\epsilon =$  fraction of particles with kinetic energy in the interval  $(\epsilon, \epsilon + d\epsilon)$  and  $\epsilon = mv^2/2$ .

d) Calculate the most probable kinetic energy and the mean energy at temperature  $T$ .

d) Make a graph of  $f(v)$  versus  $v$ ,  $f(v_x)$  versus  $v_x$ , and of  $f(\epsilon)$  versus  $\epsilon$  for liquid water at 25 °C.

**Problem 7.5.** Starting with the one-particle distribution function, obtain:

(a) The probability distribution of  $p_x$  (momentum in the  $x$  direction).

(b) The probability distribution of  $v_x$ .

(c) The probability distribution of  $v = (v_x^2 + v_y^2 + v_z^2)^{1/2}$ .

(d) The probability distribution of the kinetic energy,

$$E = \frac{v_x^2 + v_y^2 + v_z^2}{2m}$$

**Problem 7.6.** In quantum mechanics we find that the energy levels of a particle in a square box of length  $L$  is

$$E = \frac{h^2}{8mL^2} (n_x^2 + n_y^2 + n_z^2)$$

where  $n_{xyz}$  are positive integers. Calculate the number of microstates,  $\Omega$ , as a function of  $E$ .

*Hint:* Since  $n_{xyz}$  are discrete, solutions do not exist for any arbitrary value of  $E$ . If we allow the energy to vary within a small range  $dE$ , we may then calculate the corresponding

number  $d\Omega$  within this range. The total number of microstates may then be calculated by integration of  $d\Omega$  with respect to  $E$ . This approach is acceptable provided that the density of microstates is high i.e. when the distance (in energy) between neighboring microstates is small.

**Problem 7.7.** In quantum mechanics we find that the number of microstates for a single particle in a volume  $V$  and with energy  $E$  is

$$\Omega_1(E, V) = \frac{4\sqrt{2}\pi}{3} \left( \frac{Em}{h^2} \right)^{3/2} V$$

- Obtain the equation of state.
- If the energy of the particle is  $E = 10^{-20}$  Joule, what is its temperature?
- Obtain the  $c_V$  of the particle in J/mol K.

*Additional data:*  $k = 1.380 \times 10^{-23}$  J/K;  $m = 6.6 \times 10^{-26}$  kg;  $h = 6.62610^{-34}$  J s.

**Problem 7.8.** a) Recall that the chemical potential of a pure component in the canonical ensemble is

$$\mu = -kT \left( \frac{\partial \ln Q_N}{\partial N} \right)_{T, V}$$

Show that in the limit of large  $N$  the chemical potential takes the form

$$\mu = -kT \ln \frac{Q_{N+1}}{Q_N} = -kT \ln \left( \frac{Q_1}{NV} \frac{Z_{N+1}}{Z_N} \right)$$

where  $Q_N$  is the canonical partition function of  $N$  particles and is given by

$$Q_N = \frac{(2\pi mkT)^{3N/2}}{h^{3N}} \frac{Z_N}{N!}, \quad N = 0, 1, 2, \dots$$

- Use the above result to obtain an expression for the chemical potential of an ideal gas in terms of  $V$ ,  $T$  and  $N$ .
- Reconcile your result with the classical expression, according to which

$$\mu^{\text{ig}} = \mu_0 + kT \ln P/P_0$$

where  $\mu_0$  is a function of  $T$  only and  $P_0$  is constant.

- Derive the classical result for the chemical potential of the ideal-gas, given above.

**Problem 7.9.** a) Write the canonical partition function for a monoatomic substance ignoring any contributions from the internal structure of the molecule. Define all the variables that appear in your formula.

- Obtain  $C_V$  in terms of the canonical partition function and its derivatives.
- Calculate  $C_V$  for molecules that interact only with a hard-sphere potential with diameter  $\sigma$ .

**Problem 7.10.** We have  $N$  indistinguishable particles each of which can exist in microstate  $i = 1, 2, \dots$ , with probability

$$p_i = \frac{e^{-\beta E_i}}{q}$$

where  $E_i$  is the energy of microstate and  $q$  is the single-particle partition function. Obtain the probability  $p_{i,j,\dots}$  that the  $N$ -particle system is at a microstate such that particle 1 is in microstate  $i$ , particle 2 is in microstate  $j$ , etc., if the state of a particle is uncorrelated to the state of the other particles. Express the canonical partition function of the  $N$  particles in terms of  $q$  and obtain the energy of the system.

**Problem 7.11.** Suppose the factorial term is not included so that the canonical partition function of  $N$  particles is written as

$$Q_N = \frac{1}{h^{3N}} \int \int e^{-H/kT} d\mathbf{p}^N d\mathbf{r}^N$$

- a) Calculate the equation of state and entropy of  $N$  non-interacting particles.  
b) Discuss your results.

**Problem 7.12.** a) Obtain the relationship between the  $N$ -particle distribution,

$$W_N = W_N(1, 2, \dots, N),$$

and the one-particle distribution,  $w_1$  in a canonical system.

- b) Show that if there is no interaction between the particles, the  $N$ -particle function is given by

$$W_N(1, 2, \dots, N) = w_1(1)w_1(2) \cdots w_1(N)$$

where  $w_1(i)$  is the probability to find particle  $i$  at a microstate independently of the microstate of the other particles.

- c) Show that the canonical partition function in part (b) is of the form

$$Q_N = \frac{q^N}{N!}$$

where  $q$  is a function of  $T$  and  $V$  but not of  $N$ .

**Problem 7.13.** a) Show that the entropy of the canonical ensemble of  $N$  particles is

$$\frac{S}{k} = - \int (\omega_N W_N) \ln(\omega_N W_N) \frac{d\Gamma}{\omega_N}$$

where  $W_N$  is the  $N$ -particle distribution function and  $\omega_N = N!h^{3N}$  is the volume of the  $N$ -particle microstate in phase space.

- b) Define by analogy anew entropy functional,  $S_B$ , based on the one-particle distribution function,  $w_1$ :

$$\frac{S_B}{k} = -N \int (\omega_1 w_1) \ln(\omega_1 w_1) \frac{d\Gamma_1}{\omega_1}$$

where  $\Gamma_1 = (\mathbf{r}_1, \mathbf{p}_1)$  are the coordinates in phase space of particle 1 and  $\omega_1 = \omega_N^{1/N}$  is the volume of the  $N$ -particle microstate per particle. This function is referred to as the Boltzmann entropy, in contrast to  $S$ , which is also known as the Gibbs entropy. Show that

$$S_B \leq S$$

and determine the conditions for the equality to hold (see E. T. Jaynes, *Am. J. Phys.*, **33**, (1965), 391–398.)

- c) The above results apply to particles with no internal structure. How should the entropy in part (a) be modified if the internal structure of the particles is characterized by the internal partition function  $Q_{\text{int}}$ ?

**Problem 7.14.** a) Show how the equation of state can be calculated if the configuration integral is a known function of  $T$ ,  $V$ ,  $N$ ; and vice versa, i.e., the configuration integral can be calculated if the equation of state is known. Derive all necessary equations.

- b) As a demonstration, obtain the configuration integral that corresponds to the van der Waals equation of state.

## 7.11. Problems

*Hint:* Define the “canonical” residual Helmholtz energy as  $A^r(T, V, N) = A(T, V, N) - A^{\text{ig}}(T, V, N)$  and show that

$$A^r = \int_V^\infty \left( P - \frac{NkT}{V} \right) dV$$

*Note:*  $A^r$  is not the same as the regular residual Helmholtz energy,  $A^R$ , which is defined as  $A^R(T, P, N) = A(T, P, N) - A^{\text{ig}}(T, P, N)$ .

**Problem 7.15.** a) Obtain the entropy of ideal gas in the grand canonical ensemble as a function of  $T, V, \mu$ .

b) Show that the result is in agreement with that obtained from the canonical ensemble.

c) Show that the result is in agreement with that obtained in the classical development.

**Problem 7.16.** a) Critique the assumptions in the derivation of the van der Waals equation of state.

b) Write a short (up to one page) bio of van der Waals.

**Problem 7.17.** Obtain the second virial coefficient in the limit of high temperature for a gas that obeys the potential

$$\phi = -\epsilon \left( \frac{\sigma}{r} \right)^6$$

*Hint:* At high  $T$ ,  $e^{\phi/kT} \approx 1 - \phi/kT$ . Is this potential physically meaningful? Is the high- $T$  limit of the potential physically meaningful?

**Problem 7.18.** Calculate the second virial coefficient of a gas with the hard-sphere potential:

$$\phi(r) = \begin{cases} 0 & r > \sigma \\ \infty & r < \sigma \end{cases}$$

**Problem 7.19.** Consider two canonical systems of a single component. One system contains  $N_1$  particles in volume  $V_1$ , and the other  $N_2$  particles in volume  $V_2$ . Both systems have the same intensive properties, i.e.,  $T_1 = T_2$ ,  $V_1/N_1 = V_2/N_2$ . We mix the two systems isothermally to form a new system that contains  $N_1 + N_2$  particles in volume  $V_1 + V_2$ .

a) Obtain the relationship between the configuration integrals  $Z_1$ ,  $Z_2$ , and  $Z_{12}$ .

b) Use the result to show that  $\log Q$  is extensive.

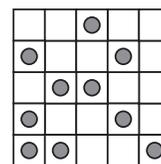
## Lattice Models

The exact calculation of the configuration integral is impossible in most situations, therefore, a number of approaches have been used to obtain approximate answers. In lattice theory, space is discretized into discrete volume elements of size  $V_0$ , thus converting the configuration integral into a summation. This approach is due to Ising who used it to model ferromagnetism by developing the partition function of a system of interacting electron spins. The basic approach, however, is applicable to the partition function of any system of interacting particles. In lattice models real space is discretized and the configuration integral becomes a summation over discrete arrangements of particles on the lattice. Yet, despite this oversimplification, lattice models are capable of predicting phase transitions and are important as tools in the understanding of the nature of phase transitions. Two common applications to chemical systems are depicted in figure 8.1. In the lattice gas model, the lattice consists of occupied and unoccupied sites; this system represents adsorption of a molecule onto the surface of an adsorbent. In the binary system, every lattice site is occupied by either species  $A$  or species  $B$ ; this system is a model for a binary solution (liquid or solid). We will focus on solutions but the basic approach should be easy to extend to other related problems.

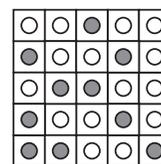
### 8.1 Elementary lattice statistics

A “lattice” is a regular arrangement of sites in space. A nearest neighbor of a given site is a site at the smallest possible distance from that site. The number of nearest neighbors,  $z$ , is the coordination number of the lattice. For a three-dimensional cubic lattice,  $z = 6$ ; for a hexagonal closed packed lattice,  $z = 12$ . Interactions are included only for particles occupying sites that nearest neighbors. The physical justification comes from the fact that the intermolecular potential decays very rapidly with distance. The assumption of nearest neighbor interactions is of course an approximation but the resulting mathematical simplification is significant and for this reason it is a standard assumption in theoretical lattice models.<sup>1</sup> When only nearest neighbor interactions are considered, the geometry of the lattice is unimportant and the coordination number alone is sufficient to describe the system.

In the basic problem we are dealing with a lattice of size  $M$  that contains  $N$  particles ( $0 \leq N \leq M$ ). The configuration integral is now replaced by a discrete



(a)



(b)

**Figure 8.1:** Schematic representation of lattice gas (left) and lattice binary solution (right).

<sup>1</sup>In simulations of lattice systems this assumption can be relaxed to include more shells of neighbors. Even then it is common practice to truncate the potential past some distance to speed up the calculation.

## 8.2. Lattice gas of non-interacting particles

summation over all possible ways to place  $N$  particles in a lattice of size  $M$ . The number of possible configuration is given by the multinomial coefficient,

$$\omega_c = \binom{M}{N} = \frac{M!}{N!(M-N)!}, \quad (8.1)$$

whose logarithm is

$$\log \omega_c = -N \log \frac{N}{M} - (M-N) \log \frac{M-N}{M}. \quad (8.2)$$

This can also be expressed as

$$\omega_c = -N \left( \log \phi + \frac{1-\phi}{\phi} \log(1-\phi) \right), \quad (8.3)$$

where  $\phi = N/M = nv_0/V$  is the fraction of space that is occupied by particles.<sup>2</sup> Written in this form it clear that  $\log \omega_c$  is as intensive property: at fixed  $N/M$ ,  $\log \omega_c$  increases in proportion to  $N$ .

The configuration integral is expressed as a summation over the discrete configuration space,

$$Z_N = \int e^{-\mathcal{U}/kT} d^N \mathbf{r} \rightarrow N! v_0^N \sum_{i=1}^{\omega_c} e^{-\mathcal{U}_i/kT}. \quad (8.4)$$

Here  $\mathcal{U}_i$  is the potential energy of configuration  $i$ , and  $v_0$  is the volume of the lattice site; the term  $v_0^N$  arises from the product of the differentials  $d\mathbf{r}_1 \cdots d\mathbf{r}_N$ , each of which is replaced by  $v_0$ . The extra  $N!$  is required because the configuration integral, as we recall, treats particles as distinguishable.<sup>3</sup> For the partition function then we have

$$Q_N = \frac{Z_N Q_{\text{int}}^N}{N! \Lambda^{3N}} = \left( \frac{v_0^N Q_{\text{int}}^N}{\Lambda^{3N}} \right) \sum_{i=1}^{\omega_c} e^{-\mathcal{U}_i/kT}. \quad (8.5)$$

The challenge is in the summation of the exponential terms that involve the energy of the lattice. The common approach is to replace the summation with its maximum term but even this problem is not trivial and can only be solved by employing suitable assumptions.

### 8.2 Lattice gas of non-interacting particles

The simplest possible lattice problem is when particles do not interact with each other. In this case  $\mathcal{U}_i = 0$  for all configurations and the solution can be obtained in closed analytic form. The configuration integral in this case is

$$\frac{Z_N}{N!} = v_0^N \sum_{i=1}^{\omega_c} (1) = v_0^N \frac{M!}{N!(M-N)!}, \quad (8.6)$$

where we have used Eq. (8.1) for the number of configurations  $\omega_c$ . Combining with Eq. (8.5), the canonical partition function is

$$Q_N = \frac{v_0^N Q_{\text{int}}^N}{\Lambda^{3N}} \frac{M!}{N!(M-N)!}. \quad (8.7)$$

<sup>2</sup>Notice that  $1/\phi = V/Nv_0 = v$ , where  $v$  is the volume per particle (or molar volume, if  $N$  is expressed in moles).

<sup>3</sup>An alternative way to write Eq. (8.4) is

$$\frac{Z_N}{N!} = v_0^N \sum_{i=1}^{\omega_c} e^{-\mathcal{U}_i/kT}.$$

Since it is  $Z_N/N!$  rather than  $Z_N$  that we need in the partition function, this is a useful form of Eq. (8.4) to remember.

## 8.2. Lattice gas of non-interacting particles

All properties may now be obtained from the partition function. Here we will calculate the equation of state and entropy and leave other properties as an exercise.

**Equation of state:** Using  $V = \nu_0 M$  we have

$$\frac{P}{kT} = \left( \frac{\partial \log Q_N}{\partial V} \right)_{TN} = \left( \frac{\partial \log Q_N}{\partial M} \right)_{TN}.$$

The only term in the partition function that depends on  $M$  is  $\omega_c$ . Using Eq. (8.2)

$$\frac{P}{kT} = \left( \frac{\partial \log \omega_c}{\partial M} \right) = \frac{1}{\nu_0} \log \frac{M}{M-N}.$$

Using again  $V = N\nu_0$  this result can be expressed in the equivalent form

$$\frac{PV}{NkT} = \frac{(M\nu_0)}{(N\nu_0)} \log \frac{1}{1 - N\nu_0/V},$$

or

$$\frac{PV}{NkT} = \frac{1}{\phi} \log \frac{1}{1-\phi}, \quad (8.8)$$

where  $\phi$  is the volume fraction of the particles. For a dilute gas,  $\phi \rightarrow 0$ ; expanding the log in powers of  $\phi$  we have

$$\log \frac{1}{1-\phi} = -\log(1-\phi) = \phi + O(\phi^2).$$

and the equation of state becomes,

$$\frac{PV}{NkT} \rightarrow \left( \frac{1}{\phi} \right) \phi = 1.$$

That is, in the dilute limit we recover the ideal-gas law.

**Equation of state:** The entropy is calculated from Eq. (5.32),

$$\frac{S}{k} = \log Q_N - \beta \left( \frac{\partial \log Q_N}{\partial \beta} \right)_{VN}. \quad (5.32)$$

Since the only terms that are function of  $\beta$  are  $\Lambda$  and  $Q_{\text{int}}$ , the derivative with respect to  $\beta$  is

$$\left( \frac{\partial \log Q_N}{\partial \beta} \right)_{VN} = \frac{\partial}{\partial \beta} (\log \Lambda^{-3N} Q_{\text{int}}^N) = -\frac{3N}{2\beta} + N \frac{\partial \log Q_{\text{int}}}{\partial \beta}.$$

Combining this with Eq. (8.7) we obtain

$$\frac{S}{k} = N \left( \log \frac{\nu_0}{\Lambda^3} + \frac{3}{2} \right) + N \left( \log Q_{\text{int}} - \beta \frac{\partial \log Q_{\text{int}}}{\partial \beta} \right) + \log \omega_c, \quad (8.9)$$

and finally, using Eq. (8.3),

$$\begin{aligned} \frac{S}{k} = N & \left( \log \frac{v_0}{\Lambda^3} + \frac{3}{2} \right) \\ & + N \left( \log Q_{\text{int}} - \beta \frac{\partial \log Q_{\text{int}}}{\partial \beta} \right) \\ & - N \left( \log \phi + \frac{1-\phi}{\phi} \log(1-\phi) \right) \end{aligned} \quad (8.10)$$

where as before,  $\phi = N/M$  is the volume fraction of space occupied by particles. In this form entropy is seen to contain three distinct contributions. The first comes from the the kinetic energy (thermal entropy), the second from internal modes of energy storage, and the third from the spatial configuration of the particle (configurational entropy).

**Comments:** Even though there is no interaction between particles, this system is *not* an ideal gas, except at low particle concentrations. Doesn't this contradict the notion of the ideal gas? No. In reality the lattice gas experiences interactions, specifically, a hard-sphere repulsion that prevents us from placing more than one particles in the same lattice site. This hard-sphere potential is built-into lattice models and is always present. Long-range interactions, if present, are in addition to the hard sphere potential. This case is discussed next.

### 8.3 Lattice gas with interactions

When particles exert interactions the summation over configurations cannot be done analytically and requires some type of approximation. Here we will use a mean-field approximation analogous to that we used in the derivation of the van der Waals equation. Given a lattice configuration, the total potential energy of the lattice is

$$\mathcal{U} = N_{11} \epsilon_{11}, \quad (8.11)$$

where  $N_{11}$  is the number of nearest neighbors and  $\epsilon_{11}$  is the strength of the intermolecular potential at a distance of one lattice size from the particle. To obtain the energy of the lattice we take the number of nearest neighbors to be equal to the coordination number times the average number of occupied sites, which is  $N/M$ . Then the number of nearest neighbors is  $N_{11} = zN/M$  and the energy of the lattice is

$$\mathcal{U} = \frac{zN^2}{M} \epsilon_{11}. \quad (8.12)$$

This result is clearly an approximation because it states that the energy of the lattice is the same for all configurations, which is not true. It represents an average interaction that is based on the average occupancy of the lattice. The configuration integral is

$$\begin{aligned} \frac{Z_N}{N!} &= v_0^N \sum_{i=1}^{\omega_c} e^{-zN^2 \epsilon_{11} / M k T} = v_0^N e^{-zN^2 \epsilon_{11} / M k T} \sum_{i=1}^{\omega_c} (1) \\ &= v_0^N \frac{M!}{N!(M-N)!} e^{-zN^2 \epsilon_{11} / M k T}, \end{aligned}$$

and with this result the partition function is

$$Q_N = \frac{v_0^N Q_{\text{int}}^N}{\Lambda^{3N}} \frac{M!}{N!(M-N)!} e^{-zN^2 \epsilon_{11}/MkT}. \quad (8.13)$$

If we remove interactions ( $\epsilon_{11} = 0$ ) this result reverts to Eq. (8.7).

This model is closely related to the van der Waals equation. If we calculate  $\epsilon_{11}$  from the van der Waals potential at distance  $r = \sigma$ , where  $\sigma$  is the diameter of the particle, also equal to the distance between neighboring lattice sites,  $\epsilon_{11} = -\sigma^6$ , we have obtained the lattice version of the van der Waals equation. The equation of state is obtained in the usual way<sup>4</sup>

<sup>4</sup>See Problem 8.9.

$$\frac{P}{kT} = \frac{1}{v_0} \log \frac{V}{V - Nv_0} + \frac{z\epsilon_{11}v_0^2}{2kT} \left(\frac{N}{V}\right)^2.$$

To explore the similarities with the van der Waals equation, we use  $v = V/N$ ,  $a = -z\epsilon_{11}v_0^2/2$  to write pressure in the form

$$P = \frac{kT}{v_0} \log \frac{v}{v - v_0} - \frac{a}{v^2},$$

For  $v \gg v_0$  the log on the right-hand side becomes

$$\log \frac{v}{v - v_0} = \log \left(1 + \frac{v_0}{v - v_0}\right) = \frac{v_0}{v - v_0} + O\left(\frac{1}{v^2}\right),$$

and the equation of state reduces to

$$P = \frac{kT}{v - v_0} - \frac{a}{v^2} + O\left(\frac{1}{v^2}\right),$$

In the dilute limit, the lattice gas with interactions goes over to the van der Waals equation. At higher pressures the two models behave differently because they treat excluded volume effects differently. In the original van der Waals derivation the excluded volume interaction was introduced by extrapolating the result from low densities. In the lattice mode the excluded volume effect is treated explicitly, though not exactly, since we have assumed configurations in space to be discrete.

## 8.4 Lattice Mixtures

The lattice treatment of solutions is based on several simplifying assumptions, most notably that the molecular size of all components are equal, therefore they fit into the same lattice, and that the lattice is completely full. The latter assumption implies that the solution is incompressible since there is no room for molecules to pack tighter. It follows that the excess volume is assumption zero and it follows that

$$G^E = A^E, \quad U^E = H^E \quad (8.14)$$

These approximations are not unrealistic since liquids are fairly incompressible but they result in a complete lack of pressure dependence in the properties calculated from lattice models. Even within these limitations, lattice models for solutions provide a good starting point for the study of complex molecular interactions.

### 8.4.1 Interaction energy

Given that molecular interactions have a short range, we will assume that only nearest neighbors contribute to the interaction between a molecule and the remaining solution. We assume that the coordination number of the lattice (number of nearest neighbors) is  $z$ . For a closed pack arrangement,  $z = 12$ . Liquids are less densely packed and  $z$  is usually taken to be closer to 10. In the development of the theory  $z$  does not need to be specified. An important property of a solution is the exchange energy,  $w$ . The situation is schematically shown in figure 8.2: the initial state is two pure liquids, one of component 1 and one of component 2. The final state is obtained by exchanging two molecule between the two liquids to create two solutions, one that contains a single molecule of component 2 surrounded by component 1, and a similar solution of component 1 in a liquid of component 2. The energy to replace a molecule in liquid 1 is

$$z\epsilon_{21} - z\epsilon_{11}$$

where  $\epsilon_{11}$  is the interaction energy between two molecules of type 1 and  $\epsilon_{21}$  is the interaction energy between components 2 and 1. Similarly, the energy to replace a molecule in the liquid of component 2 is

$$z\epsilon_{12} - z\epsilon_{22}$$

The total energy is the sum of the two and noting that  $\epsilon_{12} = \epsilon_{21}$  the result is

$$\Delta E = 2z \left[ \epsilon_{12} - \frac{\epsilon_{11} + \epsilon_{22}}{2} \right] \quad (8.15)$$

The *exchange energy* is defined as

$$w = z \left[ \epsilon_{12} - \frac{\epsilon_{11} + \epsilon_{22}}{2} \right] \quad (8.16)$$

and represents the energy change per molecule exchanged. The exchange energy depends on the difference between the cross interaction and the average self interaction between components 1 and 2. Depending on the relative magnitude of these terms,  $w$  may be positive, negative, or zero.

### 8.4.2 Lattice energy

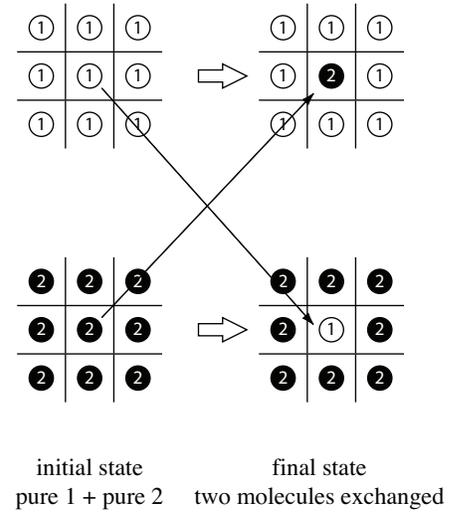
In a solution of two components at arbitrary composition, the potential interaction between a molecule and the solution depends on the composition of first neighbors. Let  $N_1$  and  $N_2$  be the number of molecules of type 1 and 2. These numbers must add up to the total number of lattice sites:

$$N_1 + N_2 = N_{\text{tot}} \quad (8.17)$$

Let  $N_{11}$ ,  $N_{22}$  and  $N_{12}$  be the number of 1-1 pairs, 2-2 pairs, and 1-2 pairs, respectively. The total interaction in the solution is

$$U = N_{11}\epsilon_{11} + N_{22}\epsilon_{22} + N_{12}\epsilon_{12} \quad (8.18)$$

The number of particles and the number of pairs are not independent but satisfy a set of conservation equations. Let us consider the number of pairs that involve



**Figure 8.2:** Calculation of interchange energy  $w$ . In general a molecule has  $z$  closest neighbors that contribute to the potential interaction.

molecules of type 1: there are two 1's in each  $N_{11}$  and one 1 in each  $N_{12}$ . If we go through all sites occupied by component 1 and count all neighbors of type 1 plus all neighbors of type 2, the result is  $2N_{11} + N_{12}$ . In doing so we have counted all neighbors of sites occupied by component 1, whose number is  $zN_1$ .<sup>5</sup> Therefore,

$$2N_{11} + N_{12} = zN_1, \quad (8.19)$$

and similarly for component 2:

$$2N_{22} + N_{12} = zN_2. \quad (8.20)$$

Using Eqs. (8.19) and (8.20) to express  $N_{11}$  and  $N_{22}$  in terms of  $N_1$ ,  $N_2$  and  $N_{12}$ , the total interaction energy then becomes

$$U = N_1 \frac{z\epsilon_{11}}{2} + N_2 \frac{z\epsilon_{22}}{2} + N_{12} \frac{\epsilon_{12}}{z} \quad (8.21)$$

This may also be written as

$$U = U_1 + U_2 + N_{12} \frac{w}{z} \quad (8.22)$$

where  $U_1 = zN_1\epsilon_{11}/2$  and  $U_2 = zN_2\epsilon_{22}/2$  are the energies of pure components before mixing.<sup>6</sup>

**Canonical partition function** The canonical partition function of a binary mixture is

$$Q_{12} = \frac{Z_{N_1, N_2}}{N_1! N_2! \Lambda_1^{3N_1} \Lambda_2^{3N_2}}$$

For the configuration integral we have<sup>7</sup>

$$\frac{Z_{N_1, N_2}}{N_1! N_2!} = \sum_{j=1}^{\omega_c} e^{-U_j/kT} = e^{-(U_1+U_2)/kT} \sum_{j=1}^{\omega_c} e^{-N_{12}|_j w/zkT} \quad (8.23)$$

The summations run over all microstates and  $U_j$  and  $N_{12}|_j$  are understood to refer to the energy and number of 1-2 pairs in microstate  $j$ . The total number of microstates is equal to the number of ways that  $N_1$  molecules of type 1 and  $N_2$  molecules of type 2 can be placed on a lattice of size  $N = N_1 + N_2$  and is given by the multinomial coefficient:

$$\omega_c = \frac{(N_1 + N_2)!}{N_1! N_2!} \quad (8.24)$$

With these definition, the partition function of the two-component system can be written as

$$Q_{12} = Q_1 Q_2 \sum_{j=1}^{\omega_c} e^{-N_{12}|_j w/zkT} \quad (8.25)$$

where  $Q_1$  and  $Q_2$  are the partition functions of the pure components. In the lattice model, the pure component is represented by a lattice of size  $N_i$  that is completely filled by particles of component  $i$ . There is only one possible configuration and the partition function becomes

$$Q_i = \frac{v_0^{N_i} e^{-N_i z\epsilon_{ii}/z}}{\Lambda^{3N_i}}, \quad (i = 1, 2). \quad (8.26)$$

<sup>5</sup>  $N_{11}$  is multiplied by 2 because the pair is visited twice, with each particle counted once as a site occupied by component 1, and once again as a neighbor of component 1.

<sup>6</sup> To calculate the energy of the pure component we fill a lattice with  $N_1$  particles of component 1. The energy of that lattice is given by Eq. (8.21) with  $N_2 = N_{12} = 0$ .

<sup>7</sup> Recall that  $Z_{N_1, N_2}$  sums over all permutations of indistinguishable particles; accordingly,  $Z_{12}/N_1! N_2!$  is a summation over distinct states only.

Combining Eqs (8.25) and (8.26), we obtain the canonical partition function of the mixture in the form

$$Q_{12} = Q_1 Q_2 Q_{\text{mix}} \quad (8.27)$$

where

$$Q_{\text{mix}} = \sum_i^{\omega_c} e^{-w N_{12}/z k T}. \quad (8.28)$$

Since all microstates with same number of 1-2 pairs have the same energy, the result can also be expressed as a summation over  $N_{12}$ ,

$$Q_{\text{mix}} = \sum_{N_{12}} \omega_c(N_1, N_2, N_{12}) e^{-w N_{12}/z k T}. \quad (8.29)$$

where  $\omega(N_1, N_2, N_{12})$  is the number of microstates with  $N_{12}$  1-2 pairs, given  $N_1$  particles of type 1 and  $N_2$  of type 2. This partition function is directly related to the free energy of mixing, as we can easily see. Taking the logarithm of (8.27) we have,

$$-\frac{A_{12}}{k T} = -\frac{A_1}{k T} - \frac{A_2}{k T} + \log Q_{\text{mix}} \quad (8.30)$$

from which we obtain

$$\frac{\Delta_{\text{mix}} A}{k T} = -\log Q_{\text{mix}}. \quad (8.31)$$

The maximum term in the summation in Eq. (8.29) corresponds to the equilibrium number of 1-2 pairs,  $\bar{N}_{12}$ . By the maximum term method we then have

$$\log Q_{\text{mix}} = \log \omega(\bar{N}_{12}) - \frac{w \bar{N}_{12}}{z k T}. \quad (8.32)$$

The mean energy of the lattice is obtained from Eq. (8.22) with  $N_{12}$  replaced by  $\bar{N}_{12}$ .

$$U = U_1 + U_2 + N_{12}^* \frac{w}{z}. \quad (8.33)$$

We recognize the last term on the right-hand side as the excess energy of mixing:

$$\Delta_{\text{mix}} U = N_{12}^* \frac{w}{z}. \quad (8.34)$$

Using Eqs. (8.34) and (8.31), Eq. (8.32) can be written as

$$\frac{\Delta_{\text{mix}} A}{k T} = -\log \omega(N_{12}^*) + \frac{\Delta_{\text{mix}} U}{k T}, \quad (8.35)$$

and this leads to the entropy of mixing,

$$\frac{\Delta_{\text{mix}} S}{k} = \log \omega(N_{12}^*). \quad (8.36)$$

### [WORK HERE]

Therefore we have obtained a direct relationship between the free energy of mixing on the one hand and molecular interactions on the other. interactions enter through the exchange energy, a constant, and the number of 1-2 pairs

in the ensemble. All properties of mixing can be obtained from the partition function of mixing, subject to the limitations of the lattice model, which we review here again: (1) Both components are of the same size; and (2) the lattice is incompressible, i.e., the effect of pressure is absent. The last assumption is equivalent to dropping the term  $PV$  from all equations and leads to the identities  $U = H$  and  $A = G$ , which for liquids represent approximations that are commonly accepted.

## 8.5 Random solution

A special case arises when  $w = 0$ . Then, the summand in Eq. (8.29) is equal to 1 and the summation is equal to the number of configurations:

$$Q_{\text{mix}} = \sum_{i=1}^{\omega_c} (1) = \omega_c = \frac{(N_1 + N_2)!}{N_1!N_2!}.$$

The free energy of mixing is

$$\frac{\Delta_{\text{mix}}A}{kT} = -\log Q_{\text{mix}} = N(x_1 \log x_1 + x_2 \log x_2). \quad (8.37)$$

We have obtained ideal solution behavior under the condition that the exchange energy is zero, or

$$\epsilon_{12} = \frac{\epsilon_{11} + \epsilon_{22}}{2}$$

This is a weaker condition than the requirement that cross and self interactions be equal and requires instead the self interaction to be equal to the average self interaction. This weaker condition of ideality is a result of two special properties of the lattice model we have used: first, the symmetry of the lattice, which is such that all sites have exactly the same number of neighbors; and second, the fact that the lattice is completely full. In continuous space as well as in compressible lattice models<sup>8</sup> ideality requires the stricter condition  $\epsilon_{11} = \epsilon_{22} = \epsilon_{12}$ .

All other properties of mixing follow from Eq. (8.37). The entropy of mixing is

$$\Delta_{\text{mix}}S = -\left(\frac{\partial \Delta_{\text{mix}}A}{\partial T}\right)_{V, N_1, N_2} = Nk(x_1 \log x_1 + x_2 \log x_2), \quad (8.38)$$

and the energy of mixing is

$$\Delta_{\text{mix}}U = \Delta_{\text{mix}}A + T\Delta_{\text{mix}}S = 0. \quad (8.39)$$

We also have

$$\Delta_{\text{mix}}V = 0, \quad \Delta_{\text{mix}}G = \Delta_{\text{mix}}A, \quad \Delta_{\text{mix}}H = \Delta_{\text{mix}}U, \quad (8.40)$$

by virtue of incompressibility. Therefore, all properties of the ideal solution are recovered.

It is a useful exercise to calculate the mean number of 1–2 pairs in the random solution. With  $w = 0$ , all lattice configurations are equally probable. On average, then, the neighbors of a site are random, as if chosen randomly from a large pool of  $N_1$  particles of component 1 and  $N_2$  particles of component 2. There are

<sup>8</sup>It is possible to construct lattice models that include unoccupied sites in order to incorporate the effect of pressure.

$N_1$  sites of component 1 surrounded by a total  $zN_1$  neighbors. In the random solution the composition of these neighbors is the same as in the bulk, i.e., they contain a fraction  $N_1/(N_1 + N_2)$  of component 1 and a fraction  $N_2/(N_1 + N_2)$  of component 2. The number of 1–2 pairs then is

$$N_{12}^* = \frac{zN_1N_2}{N_1 + N_2}. \quad (8.41)$$

The corresponding numbers of 1–1 and 2–2 pairs are obtained by substituting this into Eqs. (8.19) and (8.20):

$$N_{11}^* = \frac{z}{2} \frac{N_1^2}{N_1 + N_2}, \quad N_{22}^* = \frac{z}{2} \frac{N_2^2}{N_1 + N_2}. \quad (8.42)$$

Using  $x_i = N_i/(N_1 + N_2)$  for the mol fraction of component  $i$ , these results can be expressed in much simpler form as

$$\frac{N_{12}^*}{zN} = x_1x_2; \quad \frac{N_{11}^*}{zN} = \frac{x_1^2}{2}; \quad \frac{N_{22}^*}{zN} = \frac{x_2^2}{2}. \quad (8.43)$$

where  $N = N_1 + N_2$  is the total number of particles. Since  $zN$  is the total number of ordered pairs on the lattice, these results have a simple probabilistic interpretation: the probability to find a pair  $i$ – $j$  is equal to the probability  $x_i$  to find an  $i$ , times the probability  $x_j$  to find a  $j$ ; if  $i = j$  the result must be divided by 2 to correct for overcounting.<sup>9</sup>

## 8.6 Regular solution

When  $w \neq 0$  the solution deviates from ideal behavior and an immediate consequence is that number of 1–2 pairs is no longer given by Eq. (8.41). With  $w > 0$  cross interactions are more repulsive than self interactions and we expect  $N_{12}$  to be less than  $N_{12}^*$ ; by contrast,  $w < 0$  indicates preferential attraction between components and the number of 1–2 pairs is larger than that in ideal solution. Both conditions result in a solution entropy that is *lower* than that of the random solution, i.e., the excess entropy is negative.<sup>10</sup> The magnitude of the excess entropy depends on the interaction and increases as  $|w|$  increases. The *regular* solution is a solution characterized by interactions whose magnitude is strong enough to exhibit deviations from ideal-solution behavior, yet weak enough that their effect on the entropy of mixing can be neglected. Under the assumption of regular solution, the mean number of 1–2 pairs is given by Eq. (8.41). We then replace all  $N_{12}$  in the summation in Eq. (8.29) with the mean number  $N_{12}^*$ . This produces the following result for the partition function of mixing:

$$Q_{\text{mix}} = w_c e^{-N_{12}^*/z kT}. \quad (8.44)$$

Using Eqs. (8.41) and (8.1) the final result is

$$Q_{\text{mix}} = \frac{(N_1 + N_2)!}{N_1!N_2!} \exp\left(-\frac{N_1N_2}{N_1 + N_2} \frac{w}{kT}\right). \quad (8.45)$$

<sup>9</sup>By “ordered” pairs we mean that a pair of dissimilar particles is counted twice, once as 1–2 and once as 2–1, but a pair of similar particles is counted only once.

<sup>10</sup>The random solution has the maximum entropy at fixed  $N_1, N_2$  because all configurations are equally probable. The result of interactions is to favor certain configurations at the expense of others. The result is a *decrease* of the entropy of the solution, regardless of whether it is due to preferential attraction or repulsion between dissimilar components.

The free energy of mixing is

$$\begin{aligned} \frac{\Delta_{\text{mix}}A}{kT} &= -\log Q_{\text{mix}} \\ &= N_1 \ln \frac{N_1}{N_1 + N_2} + N_2 \ln \frac{N_2}{N_1 + N_2} + \frac{w}{kT} \frac{N_1 N_2}{N_1 + N_2}, \end{aligned} \quad (8.46)$$

and can be expressed more conveniently in terms of the mol fractions,  $x_i = N_i/(N_1 + N_2)$  in the form

$$\frac{\Delta_{\text{mix}}A}{(N_1 + N_2)kT} = x_1 \ln x_1 + x_2 \ln x_2 + \frac{w}{kT} x_1 x_2. \quad (8.47)$$

The term  $x_1 \ln x_1 + x_2 \ln x_2$  on the right-hand side is the free energy of mixing of ideal solution; accordingly, the extra term on the far right is the excess free energy of the solution:

$$A^E = G^E = \frac{w}{kT} x_1 x_2, \quad (8.48)$$

where we have used  $G^E = A^E$  due to the incompressible property of the lattice. It is an easy exercise to show that the activity coefficients of the two components are

$$\log \gamma_1 = \frac{w}{kT} x_2^2, \quad \log \gamma_2 = \frac{w}{kT} x_1^2. \quad (8.49)$$

We recognize the result as the simplified Margules equation. The lattice derivation indicates that this model is appropriate for components that are similar in size and whose interaction is not too strong.<sup>11</sup> Other properties of the regular solution can be obtained easily and are summarized in Table 8.1.

If we neglect for a moment the restriction that this model requires  $w/kT$  to be small in magnitude and allow  $w/kT$  to span the entire range from  $-\infty$  to  $+\infty$ , we obtain phase behavior that includes phase separation. The Gibbs energy of mixing is

$$\frac{\Delta_{\text{mix}}G}{NkT} = x_1 \log x_1 + x_2 \log x_2 + \frac{w}{kT} x_1 x_2. \quad (8.50)$$

Figure 8.3 shows the Gibbs energy from this above equation for various values of the ratio  $w/kT$ . For all negative values of  $w/kT$  and for positive values up to 2, the Gibbs free energy is a convex function of  $x_1$ , which means that the system forms stable solutions at all compositions. For  $w/kT > 2$ , the Gibbs energy begins to form concave regions. These are unstable and must be removed using the double-tangent method: we locate the line that is tangent to the two stable branches such that the concave portion lies entirely above the tangent line. We then erase the concave portion and connect the stable branches by a straight tie line. This corrected graph is shown in Fig. 8.4 with tie lines indicated by dashed lines and the phase boundaries by open circles. As we see, by increasing the ratio  $w/kT$  the mutual solubility decreases as the phase boundaries approach the pure components at  $x_1 = 1$  and  $x_1 = 0$ .

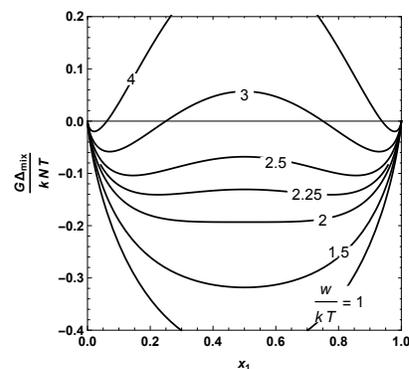
The conditions for phase splitting can be obtained analytically for this model. Stability requires the Gibbs energy to be a convex function of  $x_1$ ; mathematically,

$$\frac{d^2}{dx_1^2} \left( \frac{\Delta_{\text{mix}}G}{NkT} \right) \geq 0. \quad (8.51)$$

<sup>11</sup>More precisely, it is the ratio  $|w/kT|$  that must be small.

**Table 8.1:** Properties of regular solution

$V^E = 0$
$S^E = 0$
$U^E = H^E = \frac{w}{kT} x_1 x_2$
$A^E = G^E = \frac{w}{kT} x_1 x_2$
$\log \gamma_1 = \frac{w}{kT} x_2^2; \quad \log \gamma_2 = \frac{w}{kT} x_1^2$



**Figure 8.3:** The Gibbs energy of regular solution according to Eq. (8.50) for various values of the ratio  $w/kT$ . For sufficiently large values of  $w/kT$  the system exhibits unstable behavior, indicated by the concave regions of the graph.

## 8.7. Quasichemical approximation (QCA)

Using (8.50) this leads to the condition,

$$\frac{w}{kT} \leq \frac{1}{2x_1(1-x_1)}. \quad (8.52)$$

For the system to form a stable solution at all compositions, this inequality must hold for any  $x_1$ . The quantity on the right-hand side is convex and achieves its minimum value at  $x_1 = x_2 = 1/2$ , and its minimum is 2. Therefore, the convex condition is satisfied for all  $x_1$  if  $w/kT \leq 2$  but is violated for some  $x_1$  if  $w/kT > 2$ . The phase behavior of the system can then be summarized as follows:

- $w/kT \leq 2$ : The two components are mutually soluble at all compositions.
- $w/kT > 2$ : The system has limited mutual solubility and splits into two liquid phases when this solubility is exceeded. The phase boundaries are calculated by the double tangent method. This calculation can be simplified by noting that the symmetry of the system is such that the phase boundaries are at the two minimums of the Gibbs energy of mixing. The two phase boundaries,  $x_1^a$  and  $x_1^b$ , correspond to the two roots of the equation that defines the minimum in the Gibbs energy of mixing:

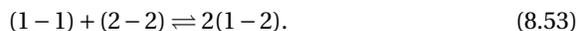
$$\frac{d}{dx_1} \left( \frac{\Delta_{\text{mix}}G}{NkT} \right) = 0 = \frac{w}{kT} (1 - 2x_1) + \log \frac{x_1}{1-x_1}.$$

For  $w/kT > 2$  this equation has two roots in  $x_1$ , located symmetrically around  $x_1 = 0.5$ , which correspond to the component-1 rich phase (larger root) and the component-2 rich phase (smaller root).

The phase behavior of the lattice model of the regular solution is an instructive exercise in producing the phase diagram of a solution using lattice concepts. It is an approximate model, however, as it assumes the excess entropy to be zero. An improvement over this assumption is provided by the *quasichemical approximation*, which is discussed next.

### 8.7 Quasichemical approximation (QCA)

The quasichemical theory attempts to incorporate the effect of interactions on the number of 1-2 pairs by treating particle pairs as the product of a chemical reaction. If we exchange the position of two particles between two pairs, one of type 1-1 and one of type of 2-2, the result is two pairs of type 1-2 (Fig. (8.5)). This can be written schematically as a reversible chemical reaction,

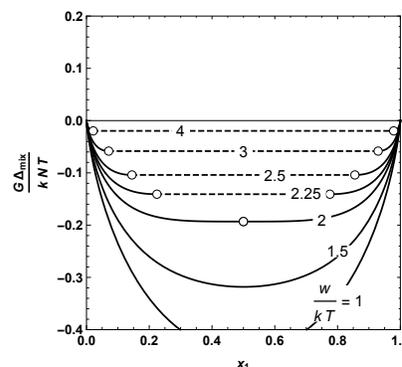


with equilibrium constant

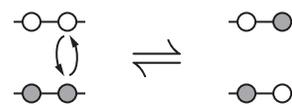
$$K = \frac{N_{12}^2}{N_{11} N_{22}}. \quad (8.54)$$

The energy difference between products and reactants is

$$\Delta U = 2\epsilon_{12} - \epsilon_{11} - \epsilon_{22} = \frac{2w}{z}, \quad (8.55)$$



**Figure 8.4:** The stable behavior of regular solution is obtained by removing the concave portions and joining the stable segments with a tie line (dashed lines). The composition of the equilibrium phases is indicated by the open circles.



**Figure 8.5:** The formation of 1-2 pairs can be viewed as the result of an exchange of two particles between a 1-1 pair and a 2-2 pair.

and since the lattice is incompressible, this is also equal to the enthalpy of the reaction. The equilibrium constant then can be written as

$$K = K^* e^{-2w/zkT}, \quad (8.56)$$

where  $K^*$  is constant to be determined. In the quasichemical approximation (QCA), the relative abundance of pairs is determined by this equilibrium constant. If we set  $w/zkT = 0$ , we obtain  $K = K^*$ , thus we recognize  $K^*$  to be the equilibrium constant in the random solution. Using their values from Eq. (8.43) we find  $K^* = 4$ . Now that we have the equilibrium constant as a function of the ratio  $w/zkT$ , we can calculate the number of 1–2 pairs by solving Eq. (8.54) along with (8.19) and (8.20). This requires somewhat tedious algebra<sup>12</sup> but the final result can be expressed in the form,

$$N'_{12} = zx_1x_2(N_1 + N_2) \left( \frac{2}{y+1} \right), \quad (8.57)$$

where  $y$  is

$$y = \sqrt{1 - 4x_1x_2(e^{2w/zkT} - 1)}. \quad (8.58)$$

When  $w/zkT \rightarrow 0$  we find  $y \rightarrow 1$ , and Eq. (8.57) reverts to Eq. (8.41) for random solution, as it should. The energy of the solution is obtained from Eq. (8.22),

$$U = U_1 + U_2 + N'_{12} \frac{w}{z}, \quad (8.59)$$

and we recognize the last term as the excess energy of the solution:

$$\frac{U^E}{N_1 + N_2} = \frac{2wx_1x_2}{y+1}. \quad (8.60)$$

The last step is to obtain the excess free energy from its relationship to the excess energy:

$$A^E = \frac{z}{2w} \int_0^\beta U^E d\left(\frac{2w}{zkT}\right). \quad (8.61)$$

To calculate the integral we change the integration variable to  $y$  and note that when  $2w/zkT \rightarrow 0$  we also have  $y \rightarrow 0$ . Using

$$d\left(\frac{2w}{zkT}\right) = \frac{2ydy}{y^2 - (1 - 2x_1)^2} \quad (8.62)$$

along with Eq. (8.60), the integral in Eq. (8.61) becomes

$$\frac{A^E}{N_1 + N_2} = \frac{zx_1x_2}{2} \int_0^{2w/zkT} \frac{2ydy}{(y+1)(y^2 - (1 - 2x_1)^2)}, \quad (8.63)$$

and the final result is

$$\frac{A^E}{kT(N_1 + N_2)} = \frac{z}{2} \left( x_1 \log \frac{y+1-2x_2}{(y+1)x_1} + x_2 \log \frac{y+1-2x_1}{(y+1)x_2} \right). \quad (8.64)$$

This is also the same as the excess Gibbs energy since the excess volume is zero. The activity coefficients can now be calculated by partial molar differentiation

<sup>12</sup>First write the equilibrium condition as

$$N'_{12} = e^{2w/zkT} \left( \frac{zN_1 - N'_{12}}{2} \right) \left( \frac{zN_2 - N'_{12}}{2} \right),$$

where we have used the balance equations to express  $N_{11}$  and  $N_{22}$  in terms of  $N'_{12}$ . Next, write  $N'_{12}$  as

$$N'_{12} = zx_1x_2(N_1 + N_2) \left( \frac{2}{y+1} \right),$$

where  $y$  is to be determined. Finally, eliminate  $N'_{12}$  between these two equations and solve for  $y$  to obtain Eq. (8.58). The details of this derivation can be found in Guggenheim (1952, pp. 38–39)

of the above result, but the the symmetry of the right-hand side allows us to recognize the logarithm of activity coefficients as the the factors that are multiplied by  $x_1$  and  $x_2$  in the two additive terms of the right-hand side:

$$\log \gamma_1 = \frac{z}{2} \log \frac{y+1-2x_2}{(y+1)x_1}; \quad \log \gamma_2 = \frac{z}{2} \log \frac{y+1-2x_1}{(y+1)x_2}. \quad (8.65)$$

In the limit  $2w/zkT \rightarrow 0$  ( $y \rightarrow 0$ ) we recover the ideal solutions as both  $\gamma_1$  and  $\gamma_2$  go to 1.

### 8.7.1 Combinatorial interpretation of QCA

In the previous section we derived the free energy of a lattice solution by treating the three types of nearest neighbors pair, 1-1, 2-2 and 1-2 as reactants. It is possible to obtain the same final results using a combinatorial approach that does not explicitly invoke the quasichemical assumption. The starting point is Eq. (8.29) for the partition function of the mixture,

$$Q_{\text{mix}} = \sum_{N_{12}} g(N_1, N_2, N_{12}) e^{-wN_{12}/zkT}. \quad (8.29)$$

The goal is to determine the number  $g(N_1, N_2, N_{12})$  of configurations that contain  $N_{12}$  pairs of type 1-2. In all, the lattice contains four types of ordered pairs: 1-1, 2-2, 1-2 and 2-1 (types 1-2 and 2-1 are distinct from each other). Their numbers are summarized in Table 8.2; the the total number of pairs is

$$N_{11} + N_{22} + \frac{N_{12}}{2} + \frac{N_{12}}{2} = \frac{z}{2}(N_1 + N_2). \quad (8.66)$$

If these pairs were independent, the number of ways to place them on the lattice would be given by the multinomial coefficient

$$g'(N_1, N_2, N_{12}) = \frac{(N_{11} + N_{22} + N_{12} + N_{21})!}{N_{11}!N_{22}!N_{12}!N_{21}!} = \frac{\{\frac{z}{2}(N_1 + N_2)\}!}{\{\frac{z}{2}(N_1 - N_{12}/z)\}! \{\frac{z}{2}(N_2 - N_{12}/z)\}! \{\frac{z}{2}N_{12}\}! \{\frac{z}{2}N_{12}\}!}.$$

The pairs, however, are not independent<sup>13</sup> and one consequence is that the summation the above multinomial over all  $N_{12}$  is not equal to the total number of configurations  $\omega_c = (N_1 + N_2)!/N_1!N_2!$ . Thus we introduce a correction factor to the above result to ensure that the summation has the correct value:

$$g(N_1, N_2, N_{12}) = h(N_1, N_2) \times \frac{\{\frac{z}{2}(N_1 + N_2)\}!}{\{\frac{z}{2}(N_1 - N_{12}/z)\}! \{\frac{z}{2}(N_2 - N_{12}/z)\}! \{\frac{z}{2}N_{12}\}! \{\frac{z}{2}N_{12}\}!}. \quad (8.67)$$

Here,  $h(N_1, N_2)$  is a correction factor that is independent of  $N_{12}$  and ensures that the summation of  $g(N_1, N_2, N_{12})$  over all  $N_{12}$  produces the correct total number of configurations:

$$\sum_{N_{12}} g(N_1, N_2, N_{12}) = \frac{(N_1 + N_2)!}{N_1!N_2!}. \quad (8.68)$$

**Table 8.2:** Number of different pairs on the lattice. The balance Eqs. (8.19) and (8.20) were used to express  $N_{11}$  and  $N_{22}$  in terms of  $N_1$ ,  $N_2$  and  $N_{12}$ . Of the  $N_{12}$  order pairs between 1 and 2, half are of type 1-2 and half of type 2-1.

type of pair	number of pairs
1-1	$\frac{z}{2}(N_1 - \frac{1}{z}N_{12})$
1-2	$\frac{z}{2}(N_2 - \frac{1}{z}N_{12})$
1-2	$\frac{1}{2}N_{12}$
2-1	$\frac{1}{2}N_{12}$

<sup>13</sup>For example, if we place a 1-2 pair in front of an 1-1 pair we form a new 2-1 pair, but if we reverse the order we form a new 1-1 pair. That is to say, the four types of pairs are not independent of each other.

## 8.8. Flory-Huggins theory for polymer solutions

To obtain  $h(N_1, N_2)$ , we replace the summation with its maximum term, which we determine by taking the derivative of  $\log g$  with respect to  $N_{12}$  and setting it equal to zero. The value of  $N_{12}$  that maximizes  $g$  is<sup>14</sup>

$$N_{12}^* = \frac{zN_1N_2}{N_1 + N_2}, \quad (8.69)$$

and is the same as the number of 1–2 pairs in a random lattice. Using this result in Eq. (8.67) we solve for  $h$  and find

$$h(N_1, N_2) = \frac{1}{g'(N_1, N_2, N_{12}^*)} \frac{(N_1 + N_2)!}{N_1!N_2!}, \quad (8.70)$$

with  $g'$  from Eq. (8.66) and  $N_{12}^*$  from (8.69). For  $g$  we finally obtain

$$g(N_1, N_2, N_{12}) = \frac{(N_1 + N_2)!}{N_1!N_2!} \times \frac{\left\{\frac{z}{2}(N_1 - N_{12}^*/z)\right\}! \cdot \left\{\frac{z}{2}(N_2 - N_{12}^*/z)\right\}! \cdot \left(\left\{\frac{z}{2}N_{12}^*\right\}!\right)^2}{\left\{\frac{z}{2}(N_1 - N_{12}/z)\right\}! \cdot \left\{\frac{z}{2}(N_2 - N_{12}/z)\right\}! \cdot \left(\left\{\frac{z}{2}N_{12}\right\}!\right)^2}. \quad (8.71)$$

Now that we have the number of configurations, we return to the partition function of mixing and take its logarithm. We use the maximum term method the log to replace the summation is replaced with its maximum term,

$$\begin{aligned} \log Q_{\text{mix}} &= \log \left( \sum_{N_{12}} g(N_1, N_2, N_{12}) e^{-N_{12}w/zkT} \right) \\ &= \log g(N_1, N_2, \bar{N}_{12}) - \frac{\bar{N}_{12}w}{zkT}, \end{aligned} \quad (8.72)$$

where  $\bar{N}_{12}$  is the equilibrium number of 1–2 and is determined by the maximization condition

$$\frac{d \log (g(N_1, N_2, N_{12}) e^{-N_{12}w/zkT})}{d\bar{N}_{12}} = 0. \quad (8.73)$$

The result for  $\bar{N}_{12}$  is<sup>15</sup>

$$\frac{\bar{N}_{12}^2}{\bar{N}_{11}\bar{N}_{22}} = 4e^{w/zkT} \quad (8.74)$$

where  $\bar{N}_{11} = \frac{z}{2}(N_1 - \bar{N}_{12})$ ,  $\bar{N}_{22} = \frac{z}{2}(N_2 - \bar{N}_{12})$ , are the corresponding numbers of 1–1 and 2–2 pairs when  $N_{12} = \bar{N}_{12}$ . Equation (8.74) is the same as (8.54).

The combinatorial treatment is equivalent to the quasichemical approximation, as it leads to the same relationship between pairs 1–1, 2–2 and 1–2. The derivation of the free energy of mixing is mathematically simpler in the quasichemical approximation, but the combinatorial treatment offers an advantage in that it produces with an expression for the number of configurations on the lattice. This will be useful when we extend the theory to molecules of unequal size.

[SHOW GRAPHS WITH QCA AND DISCUSS]

<sup>15</sup>We start by writing  $\log g(N_1, N_2, N_{12})$  as

$$\log g = \log \frac{(N_1 + N_2)!}{N_1!N_2!} + \log g'^* - \log g'$$

where  $g'^* = g'(N_1, N_2, N_{12}^*)$  and  $g' = g'(N_1, N_2, N_{12})$ . Only  $\log g'$  is a function of  $N_{12}$ . The maximization of

$$\log g - N_{12}w/zkT$$

in Eq. (8.73) is then equivalent to the maximization of

$$\begin{aligned} B &= \log g' - N_{12} \frac{w}{zkT} \\ &= -N_{11} \log \frac{N_{11}}{A} - N_{22} \log \frac{N_{22}}{A} \\ &\quad - N_{12} \log \frac{\frac{1}{2}N_{12}}{A} - N_{12} \frac{w}{zkT}, \end{aligned}$$

with

$$\begin{aligned} N_{11} &= \frac{z}{2}(N_1 - N_{12}/z), \\ N_{22} &= \frac{z}{2}(N_2 - N_{12}/z), \\ A &= \frac{z}{2}(N_1 + N_2). \end{aligned}$$

Setting the derivative of  $B$  with respect to  $N_{12}$  equal to zero we find

$$\begin{aligned} \log(N_1 - zN_{12}) + \log(N_2 - zN_{12}) \\ - 2 \log N_{12} - \frac{2w}{zkT} = 0, \end{aligned}$$

which is equivalent to Eq. (8.74).

## 8.8 Flory-Huggins theory for polymer solutions

The Flory-Huggins theory is a lattice theory for polymer-solvent systems.<sup>16</sup> In its basic form it deals with athermal solutions, namely, solutions for which  $\Delta_{\text{mix}}H$  is zero. This presumes that the monomeric unit of the polymer and the solvent are similar in nature. The goal of the theory is to calculate the entropy of mixing that arises from the different size between polymer and solvent. The result is often used outside the polymer field to quantify entropic effects in solution due to size differences between two molecules, not necessarily polymers.

The assumptions of the theory are: (i) the monomeric unit and the solvent are of the same size, therefore, a lattice site is occupied either by a solvent molecule or by a monomeric unit. We will refer to monomeric units as “polymer segments” or just “segments.” (2) all polymer chains are identical and consist of  $r$  segments. That is,  $r$  is the degree of polymerization. (3) there is no interaction between polymer and solvent, or more precisely, the interchange energy between solvent and a monomer unit is 0. If the lattice contains  $N_s$  solvent molecules and  $N_p$  polymer chains, the total number of sites,  $N$  is

$$N = N_s + rN_p$$

Now we calculate the number of ways we can place  $N_p$  polymer chains on a lattice of size  $N$ . We do not need to consider the placement of the solvent because it occupies those sites not taken by the polymer. First we place the free segment of all  $r$  chains on the lattice. For the segment of the first chain we have  $N$  possible placements, for the first segment of the second chain we have  $N - 1$  placements and so on, therefore, the total number of placements of the free ends is

# of ways to place 1st segment of all chains =

$$\frac{N(N-1)\cdots(N-N_p+1)}{N_p!} = \frac{N!}{N_p!(N-N_p)!}$$

Here the factorial  $N_p!$  has been introduced to correct the count for the fact that the  $N_p$  chains are indistinguishable.

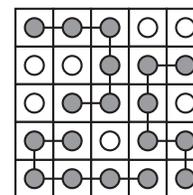
We then place the remaining  $(r-1)N_p$  segments one at a time. When it is time to place segment  $i$ , it must be placed on one of the  $z-1$  neighbors of the segment to which it is attached. At that point the lattice contains  $N_p + i - 1$  segments and the probability that a site is unoccupied is  $(N - N_p - i + 1)/N$ . Then, the number of ways to place segment  $i$  on the lattice is

# of ways to place  $i$ th segment =

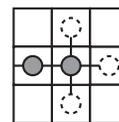
$$(z-1) \frac{N - N_p - i + 1}{N}, \quad i = 1, \dots, (r-1)N_p.$$

The total number of ways to place all  $N_p$  chains on the lattice is given by the product of these terms:

$$\Omega(N_s, N_p) = \frac{N!}{N_p!(N-N_p)!} \prod_{i=1}^{(r-1)N_p} (z-1) \frac{N - N_p - i + 1}{N}.$$



**Figure 8.6:** Flory-Huggins lattice for polymer/solvent solution with  $r = 6$ ,  $N_p = 3$ ,  $N_s = 7$ .



**Figure 8.7:** The  $i$ th segment (dashed circle) may occupy any one of the  $z-1$  neighbors of the segment to which it is attached, provided it is unoccupied. The probability that a neighbor is unoccupied is  $(N - N_p - i + 1)/N$ .

Working out the product, the final result is

$$\Omega(N_s, N_p) = \left( \frac{z-1}{N} \right)^{(r-1)N_p} \frac{N!}{N_p! N_s!}, \quad (8.75)$$

with  $N = N_s + rN_p$ . As a check, if we set  $r = 1$ , effectively reducing the polymer chain to a monomer, the result reverts to that of equally sized molecules.

### 8.8.1 Entropy of mixing

In the absence of interactions the ensemble of configurations constitutes a microcanonical ensemble with partition function  $\Omega$ . Therefore, the entropy of the polymer mixture is

$$S(N_s, N_p) = k \log \Omega(N_s, N_p), \quad (8.76)$$

and the entropy of mixing is

$$\Delta_{\text{mix}} S = S(N_s, N_p) - S(N_s, 0) - S(0, N_p) = \log \frac{\Omega(N_s, N_p)}{\Omega(N_s, 0) \Omega(0, N_p)}.$$

The number of lattice configurations of the pure solvent is 1 (there is only one way to place  $N_s$  molecules on  $N_s$  sites);<sup>17</sup> the corresponding result for the pure polymer is obtained by setting  $N_s = 0$  in Eq. (8.75):

$$\Omega(N_p, 0) = \left( \frac{z-1}{rN_p} \right)^{(r-1)N_p} \frac{(rN_p)!}{N_p!}.$$

<sup>17</sup>Verify this by setting  $N_p = 0$ ,  $N = N_s$  in Eq. (8.75).

We then have

$$\frac{\Omega(N_s, N_p)}{\Omega(N_s, 0) \Omega(0, N_p)} = \left( \frac{rN_p}{N} \right)^{(r-1)N_p} \frac{N!}{N_s! (rN_p)!}.$$

Taking the log of this expression and using the Stirling formula for the factorial, the entropy of mixing is found to be

$$\Delta_{\text{mix}} S = -k \left( N_s \log \frac{N_s}{N_s + rN_p} + N_p \log \frac{rN_p}{N_s + rN_p} \right) \quad (8.77)$$

The ratios  $N_s/N$  and  $rN_p/N$  represent the volume fraction occupied by the solvent and the polymer respectively.<sup>18</sup> With this in mind the final result can be expressed in more compact form as

$$\frac{\Delta_{\text{mix}} S}{k(N_s + N_p)} = -x_s \log \phi_s - x_p \log \phi_p \quad (8.78)$$

where

$$\phi_s = \frac{N_s}{N_s + rN_p}, \quad \phi_p = \frac{rN_p}{N_s + rN_p}, \quad (8.79)$$

are the volume fractions, and

$$x_s = \frac{N_s}{N_s + N_p}, \quad x_p = \frac{N_p}{N_s + N_p}, \quad (8.80)$$

are the mol fractions of the solvent and polymer.

<sup>18</sup>Ratios over the total number of lattice sites  $N_s + rN_p$  represent volume fractions; ratios over the total number of moles  $N_s + N_p$  represent mol fractions. If the system is viewed as a solution of *monomers* in the solvent, the volume fractions are the mol fractions of this solution.

### 8.8.2 Polymer-solvent interactions

To extend the theory to interacting systems we adopt the random-solution approximation. We treat the system as a solution of monomers with mol fraction  $\phi_p$  (the mol fraction of the solvent is  $\phi_s = 1 - \phi_p$ ). We write the interaction parameter between monomer and solvent as  $\chi = w/kT$  and with this, the energy of the solution is expressed in the form,

$$E(N_s, N_p) = E_s + E_p + N(\chi kT)\phi_s\phi_p. \quad (8.81)$$

This result essentially treats the system as a solution of monomers in solvent with respective mol fractions  $\phi_p$  and  $\phi_s$ . As in the random solution theory, we ignore the effect of interactions on the entropy and take Eq. (8.78) to be the entropy of mixing. This approximation is acceptable when  $\chi$  is small but here we adopt it anyway regardless of the magnitude of  $\chi$ . Putting these results together, the free energy of mixing is<sup>19</sup>

$$\frac{\Delta_{\text{mix}}A}{kT} = N\chi\phi_s\phi_p + N_s \log \phi_s + N_p \log \phi_p. \quad (8.82)$$

The theory itself does not provide a way for the calculation of the interaction parameter  $\chi$ . This parameter is normally deduced experimentally.

<sup>19</sup>Recall that for lattice models with fixed volume, the Helmholtz and Gibbs free energies are the same.

## 8.9 Quasichemical treatment for molecules of dissimilar size

The Flory-Huggins model provides a simple analytical result that is often used to describe mixtures of molecules of unequal size, not necessarily polymer chains. An alternative treatment is based on the quasichemical approximation (QCA) (Guggenheim, 1944a,c). The physical model is more detailed than that of the Flory-Huggins theory and as a result, both the derivations and the final expressions of the theory are more involved, but also more general. The QCA treatment forms the basis of UNIQUAC, an industrial strength model for activity coefficients that is discussed in section ??.

The physical model for the molecule is a chain that contains  $r$  connected particles (think of  $r$  as the “size” of the molecule). The chain could be linear or branched, but cannot contain cycles.<sup>20</sup> The constituent particles fit on a lattice with coordination number  $z$  and can be connected to one or more of their  $z$  immediate neighbors. Given a molecule that contains  $r$  particles, the number of non bonded nearest neighbors is

$$f = (z - 2)r + 2. \quad (8.83)$$

This can be easily confirmed in the case of a linear chain (the ends of the chain have  $z - 1$  free neighbors and the remaining  $r - 2$  particles have  $z - 2$  neighbors each) but it should then be obvious that when an end particle is moved and bonded to some other site as to produce a branched chain the number of non bonded neighbors remains the same, as long as no cycles are formed (Fig. 8.8). Equation (8.83) gives the number of available bonds around the chain, but as figure 8.8 shows, some of these bonds may point to the same lattice site. We ignore this complication and take  $f$  to be the number of nearest neighbors of the

<sup>20</sup>A cycle is a closed loop, such as an aromatic ring or cyclohexane, for example.

## 8.9. Quasichemical treatment for molecules of dissimilar size

molecule. For convenience we define  $q = f/z$  so that the number of neighbors around a molecule of size  $r$  is  $qz$ , i.e.,

$$q = \frac{(z-2)r+2}{z}. \quad (8.84)$$

In the simplest case we are dealing with a mixture of molecules of size  $r$  (component 1) and molecules of size  $r = 1$  (component 2). The lattice is completely full, therefore the total number of lattice is  $rN_1 + N_2$ . The volume fractions  $\phi_i$  of the components are

$$\phi_1 = \frac{rN_1}{rN_1 + N_2}, \quad \phi_2 = \frac{N_2}{rN_1 + N_2}, \quad (8.85)$$

and have a simple probabilistic interpretation:  $\phi_i$  is the probability that a randomly chosen site is occupied by component  $i$ . We will develop an expression for the number of configurations in the absence of interactions (athermal solution) first. We will then extend the result to a system of interacting particles.

### 8.9.1 Athermal mixture

In the absence of interactions we write the configurational partition function of the mixture by analogy to Eq. (8.27) as<sup>21</sup>

$$Q_{12} = Q_{c1}Q_{c2}\omega_c(N_1, N_2) = q_1^{N_1}q_2^{N_2}\omega_c(N_1, N_2) \quad (8.86)$$

where  $q_i^{N_i} = Q_{ci}$  is the configurational partition function of individual molecules of type  $i$ ,  $Q_{ci}$  is the configurational partition function of a lattice that contains pure component  $i$ , and  $\omega_c(N_1, N_2)$  is the number of ways to place  $N_1$  molecules of type 1 and  $N_2$  molecules of type 2 on the lattice. The free energy of the lattice is  $A_{12}/kT = -\log Q_{12}$ , and since the lattice is incompressible,

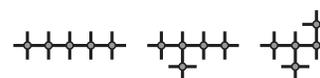
$$\frac{A}{kT} = \frac{G}{kT} = -N_1 \log q_1 - N_2 \log q_2 - \log \omega_c(N_1, N_2). \quad (8.87)$$

Consider now the equilibrium exchange of molecules between the lattice and an ideal-gas mixture of the two components at the same temperature and chemical potential as in the lattice. There are two possible exchanges:

- 1→2 One molecule of type 1 on the lattice is replaced with  $r$  molecules of type 2 from the gas phase.
- 2→1  $r$  molecules of type 2 on the lattice are replaced by one molecule of type 1 from the gas phase. In this case of course the  $r$  molecules that are removed must be in such proximity to each other that a molecule of type 1 can fit in the sites that are freed.

The exchange 1→2 amounts to evaporation of one molecule of type 1 and the condensation of  $r$  molecules of type 2. Its rate is proportional to the probability to find a molecule of type 1 on the lattice, times the pressure of component 2 raised to power  $r$ :<sup>22</sup>

$$R_{1\rightarrow 2} \propto \frac{rN_1}{rN_1 + N_2} P_2^r. \quad (8.88)$$



**Figure 8.8:** Three possible structures of a pentamer ( $r = 5$ ) in a 2-dimensional lattice with  $z = 4$ . All  $r$ -mers have 12 free bonds, in agreement with Eq. (8.83).

<sup>21</sup>The configurational partition function  $Q_{ci}$  of component  $i$  is

$$Q_{ci} = Q_i \frac{\Lambda_i^{3N_i}}{\nu_0^{N_i}},$$

where  $Q_i$  is the partition function of pure  $i$ .

<sup>22</sup>The probability to find a molecule of type 1 on the lattice is equal to the probability that a randomly selected site on the lattice belongs to a molecule of type 1. This is equal to the volume fraction of the component,

$$\phi_1 = \frac{rN_1}{rN_1 + N_2}.$$

## 8.9. Quasichemical treatment for molecules of dissimilar size

The rate of exchange 2→1 is equal to the probability to find  $r$  molecules of type 2 in connected sites, times the pressure of component 1:<sup>23</sup>

$$R_{2\rightarrow 1} \propto \frac{N_2}{rN_1 + N_2} \left( \frac{N_2}{qN_1 + N_2} \right)^{r-1} P_1. \quad (8.89)$$

In an equilibrated system these rates are equal. We must have then,

$$C \frac{P_1}{P_2^r} = \frac{rN_1}{N_2} \left( \frac{qN_1 + N_2}{N_2} \right)^{r-1}, \quad (8.90)$$

where  $C$  is a proportionality constant that depends on the structure of the molecules and is a function of  $q$ ,  $r$  and  $z$ . In the ideal-gas state pressure is related to chemical potential via

$$\mu_i = \mu_i^0 + kT \log \frac{P_i}{P_0} \Rightarrow P_i \propto e^{\mu_i/kT} \equiv \lambda_i.$$

Replacing the pressures in Eq. (8.89) with the chemical potentials we have

$$C' \frac{\lambda_1}{\lambda_2^r} = \frac{N_1}{N_2} \left( \frac{qN_1 + N_2}{N_2} \right)^{r-1}, \quad (8.91)$$

where  $C'$  is yet another constant that absorbs all the constant factors, including that  $r$  on the right-hand side of Eq. (8.90). It is more convenient to express the last result in the equivalent form

$$\rho \frac{\lambda_1 q_1}{(\lambda_2 q_2)^r} = \frac{N_1}{N_2} \left( \frac{qN_1 + N_2}{N_2} \right)^{r-1}, \quad (8.92)$$

where  $q_1$  and  $q_2$  are the configurational partition functions of the individual molecules and  $\rho = C' f_2^r / f_1$  is yet another constant that depends on  $z$ ,  $q$  and  $r$  but not on  $N_1$  or  $N_2$ . This equation is a relationship between the chemical potentials of the two components that is imposed by the condition that both components must fit on the lattice. Another relationship between the chemical potentials is given by the differential of the Gibbs energy at fixed temperature:

$$\frac{dG}{kT} = \frac{\mu_1}{kT} dN_1 + \frac{\mu_2}{kT} dN_2 = \lambda_1 dN_1 + \lambda_2 dN_2. \quad (8.93)$$

We integrate this differential, in conjunction with Eq. (8.92), from initial condition<sup>24</sup>  $N_1 = 0$ ,  $G/kT = -N_2 \log q_2$ , to  $N_1$ ,  $G/kT$ . We leave the mathematical details as an exercise but the final result is

$$\begin{aligned} \frac{G}{kT} = \frac{A}{kT} &= \log N_1! + \log N_2! - \frac{r-1}{r-q} \log(qN_1 + N_2)! \\ &+ \frac{q-1}{r-q} \log(rN_1 + N_2)! - N_1 \log \rho - N_1 \log q_1 - N_2 \log q_2. \end{aligned} \quad (8.94)$$

Combining with Eq. (8.87) and solving for  $\omega_c(N_1, N_2)$ , we find

$$\begin{aligned} \log \omega_c(N_1, N_2) &= \\ & - \log N_1! - \log N_2! \\ & + \frac{r-1}{r-q} \log(qN_1 + N_2)! \\ & - \frac{q-1}{r-q} \log(rN_1 + N_2)! + N_1 \log \rho. \end{aligned} \quad (8.95)$$

<sup>24</sup>To derive the initial condition, start with Eq. (8.87) and set  $N_1 = 0$ . Noting that  $\omega_c(N_1 = 0, N_2) = 1$  (there is only one way to place  $N_2$  monomers on the lattice) we obtain  $G/kT = -N_2 \log q_2$ , which gives the initial condition.

## 8.9. Quasichemical treatment for molecules of dissimilar size

If we use Eq. (8.84) that relates  $q$ ,  $r$  and  $z$ , we obtain the relationships,

$$\frac{r-1}{r-q} = \frac{z}{2}, \quad \frac{q-1}{r-1} = \frac{z}{2} - 1, \quad (8.96)$$

and with these substitutions Eq. (8.95) simplifies to

$$\begin{aligned} \log \omega_c(N_1, N_2) = & \\ & -\log N_1! - \log N_2! \\ & + \frac{z}{2} \log(qN_1 + N_2)! \\ & - \left(\frac{z}{2} - 1\right) \log(rN_1 + N_2)! + N_1 \log \rho. \end{aligned} \quad (8.97)$$

This result gives the number of lattice configurations in a mixture of  $N_1$   $r$ -mers and  $N_2$  monomers. The only remaining question before we can use this results for calculations is the parameter  $\rho$ . To identify the meaning of this parameter we examine Eq. (8.97) in the dilute limit for component 1,  $N_1 \ll N_2$ . In this limit Eq. (8.97)  $\omega_c$  asymptotically behaves as

$$\omega_c(N_1, N_2) \rightarrow \frac{N_{\text{tot}}^{N_1}}{N_1!} \rho^{N_1} \quad (8.98)$$

where  $N_{\text{tot}} = rN_1 + N_2$  is the total number of lattice sites.<sup>25</sup> With  $\rho = 1$  the right-hand side is the number of ways to place  $N_1$  monomers on a lattice of size  $N_{\text{tot}}$  in the ideal-gas limit. For dimers of larger particles there are more than one ways to place the particle on the lattice; we may interpret then  $\rho$  to refer to the number of ways to place an  $r$ -mer on the lattice. This number depends on the structure of the molecule and the coordination number of the lattice. For a dimer composed of identical particles the number of placements is  $z/2$ , but if the constituent particles are different, the number is  $2z$  (see Fig. 8.9). For a rigid straight molecule of any size  $r$  the number of placements is  $z/\sigma$ , where  $\sigma$  is the symmetry number of the molecule. It is not possible, however, to write a general expression for arbitrary shapes.

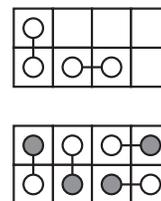
**Extension to multiple components** Recalling that for monomers ( $r = 1$ ) we have  $q = 1$ , Eq. (8.97) suggests a simple extension to arbitrary mixtures that contain molecules of several sizes  $r_i$  with corresponding parameters  $q_i$ . The result is

$$\begin{aligned} \omega_c(\{N_i\}) = & -\sum_i \log N_i! + \frac{z}{2} \log \left( \sum_i q_i N_i \right)! \\ & - \left( \frac{z}{2} - 1 \right) \log \left( \sum_i r_i N_i \right)! + \sum_i N_i \log \rho_i, \end{aligned} \quad (8.99)$$

where  $q_i$  gives the neighbors of component  $i$  and is defined as  $q_i = (zr - 2r + 2)/z$ . We offer no justification for this result other than the fact that it reverts to (8.97) for two components with  $r_1 = r$ ,  $q_1 = 1$ ,  $r_2 = q_2 = 1$ .

<sup>25</sup>To confirm this, take the log of both sides of Eq. (8.98), using (8.97) for the left-hand side, set  $N_1 = xN_2$  and expand in power series of  $x$  around  $x = 0$ . Then both sides give

$$N_2 \left\{ x(1 - \log x) + O(x^2) \right\}.$$



**Figure 8.9:** There are two ways to place a dimer composed of two identical particles on a lattice with coordination number 4. If the constituent particles are different, that number is 4.

### 8.9.2 Mixture with interactions

The number of  $(ij)$  nearest neighbors is  $N_{ij}$ . A site occupied by component  $i$  has neighbors that are either type  $i$ , or type  $k \neq i$ . The quantity

$$2N_{ii} + \sum_{k \neq i} N_{ik}$$

counts the nearest neighbors of all sites occupied by component  $i$ ; it must be equal then to the total number of nearest neighbors around all  $r_i$ -mers, which in turn is equal to  $f_i N_i = z q_i N_i$ :

$$2N_{ii} + \sum_{k \neq i} N_{ik} = z q_i N_i. \quad (8.100)$$

The notation  $\sum_{k \neq i}^K \dots$  is shortcut for the sum of all unordered pairs between dissimilar particles and can be written more explicitly as

$$\sum_{k \neq i}^K (\dots) = \sum_{k=1}^{i-1} (\dots) + \sum_{k=i+1}^K (\dots). \quad (8.101)$$

For example, for  $K = 3$ ,  $i = 2$ , Eq. (8.100) becomes

$$2N_{22} + N_{21} + N_{23} = z q_2 N_2.$$

In a perfectly random lattice, the number of  $(ij)$  pairs is

$$N_{ij} = \begin{cases} z \frac{(q_i N_i)(q_j N_j)}{\sum_i q_i N_i} & \text{if } i \neq j \\ z \frac{(q_i N_i)^2}{2 \sum_i q_i N_i} & \text{if } i = j \end{cases}$$

These results are obtained by analogy to the all-monomer case ( $q_i = 1$ ) by making the substitution  $N_i \rightarrow q_i N_i$ .

**Energetics** The total energy of the lattice is

$$U = \sum_{i=1}^{K-1} \sum_{j=i}^K N_{ij} \epsilon_{ij} \quad (8.102)$$

where  $\epsilon_{ij}$  is the interaction between an  $(ij)$  pair.<sup>26</sup> Using the balance equations, Eq. (8.100), this is also equal to

$$U = \frac{1}{2} \sum_{i=1}^K q_i z N_i \epsilon_{ii} + \frac{1}{2} \sum_{i=1}^{K-1} \sum_{j=i+1}^K N_{ij} \Delta \epsilon_{ij}, \quad (8.103)$$

where  $\Delta \epsilon_{ij}$  is

$$\Delta \epsilon_{ij} = 2\epsilon_{ii} - \epsilon_{ii} - \epsilon_{jj} = \frac{2w_{ij}}{z} \quad (8.104)$$

<sup>26</sup>This double summation is constructed so that every  $(ij)$  pair appears only once regardless of the order in  $i$  and  $j$ . For example, for three components the energy is

$$U = u_{11} + u_{12} + u_{13} + u_{22} + u_{23} + u_{33}$$

with  $u_{ij} = N_{ij} \epsilon_{ij}$ .

## 8.9. Quasichemical treatment for molecules of dissimilar size

and  $w_{ij}$  is the exchange energy,

$$w_{ij} = z \left( \epsilon_{ii} - \frac{\epsilon_{ii} + \epsilon_{jj}}{2} \right). \quad (8.105)$$

Notice that  $\Delta\epsilon_{ij}$  can be viewed as the energy change of the reaction



that converts self pairs into cross pairs. Returning to Eq. (8.103), the first term on the right-hand side is the sum of the energies of pure components before mixing. The second term then is the excess energy:

$$U^E = \sum_{i=1}^{K-1} \sum_{j=i+1}^K \frac{N_{ij}}{z} w_{ij}. \quad (8.107)$$

This energy depends only on the number of pairs between dissimilar components.

**Combinatorics** The number of ways to place all the components on the lattice is<sup>27</sup>

$$g(\{N_i\}) = \frac{z}{2} \log \left( \sum_i q_i N_i \right)! - \left( \frac{z}{2} - 1 \right) \log \left( \sum_i r_i N_i \right)! - \sum_i \log N_i! + \sum_i N_i \log \rho_i, \quad (8.108)$$

<sup>27</sup>This result is derived in Guggenheim (1944a) and Guggenheim (1944b); we use it without proof.

where  $\rho_i$  is the number of distinct ways to place molecule  $i$  on the lattice once we have specified the position of one of its particles. We now want to obtain the number of configurations with fixed energy. This means that we need to evaluate the number of configurations that have a specified number of  $(ij)$  pairs. This number, which we denote as  $g(\{N_1, N_{ij}\})$ , will be obtained by a heuristic argument. This number must satisfy the following conditions:

1. Its logarithm  $g(\{N_1, N_{ij}\})$  must be extensive in  $N_i$ , i.e., homogeneous with degree 1.
2. It must reduce to  $g(\{N_i\})$  when  $w_{ij}/zkT = 0$ .
3. The  $N_{ij}$  that maximize the quantity

$$g(\{N_1, N_{ij}\}) e^{-U(\{N_1, N_{ij}\})/kT}$$

must satisfy the QCA equilibrium,

$$\frac{N_{ij}}{N_{ii}N_{jj}} = 4e^{-2w_{ij}/zkT} \quad (8.109)$$

A mathematical expression that satisfies these conditions is the following:

$$\log g(\{N_1, N_{ij}\}) = \log g(\{N_i\}) + \sum_i \log \frac{N_{ii}^*!}{N_{ii}!} + \sum_{ij} \log \frac{N_{ij}^*! 2^{N_{ij}}}{N_{ij}! 2^{N_{ij}^*}}. \quad (8.110)$$

with  $g(\{N_1, N_{ij}\})$  from Eq. (8.99) and  $N_{ij}^*$  given by the random solution equations. The free energy of mixing is

$$-\frac{\Delta_{\text{mix}}A}{kT} = \log g(\{N_1, N_{ij}\}) - \sum_{i=1}^{K-1} \sum_{j=i+1}^K N_{ij} \frac{w_{ij}}{zkT}. \quad (8.111)$$

The equilibrium number of  $(ij)$  pairs,  $\bar{N}_{12}$  is determined by the maximization of the free energy,

$$\frac{\partial}{\partial N_{ij}} \left\{ \log g(\{N_1, N_{ij}\}) - \sum_{i=1}^{K-1} \sum_{j=i+1}^K N_{ij} \frac{w_{ij}}{zkT} \right\}_{\bar{N}_{ij}} = 0. \quad (8.112)$$

Using Eq. (8.110) for  $G(\{N_i, N_{ij}\})$  and the balance equations in (8.100) to express the  $N_{ii}$  in terms of  $N_{ij}$ , the above maximization condition reduces to

$$\underbrace{\log 2 + \frac{1}{2} \log \frac{1}{2} \left( zq_i \bar{N}_i - \sum_{k \neq i} \bar{N}_{ik} \right)}_{\bar{N}_{11}} + \underbrace{\frac{1}{2} \log \frac{1}{2} \left( zq_j \bar{N}_j - \sum_{k \neq i} \bar{N}_{jk} \right)}_{\bar{N}_{11}} - \log \bar{N}_{ij} - \frac{w_{ij}}{zkT} = 0,$$

or

$$\frac{\bar{N}_{12}^2}{\bar{N}_{11} \bar{N}_{22}} = 4e^{w_{ij}/zkT}. \quad [8.109]$$

This conforms that the equilibrium numbers of  $(ij)$  pairs satisfy the quasichemical equilibrium condition, condition 3 is therefore satisfied. With  $w_{ij}/zkT = 0$  we obtain

$$\frac{\bar{N}_{12}^2}{\bar{N}_{11} \bar{N}_{22}} = 4 = \frac{N_{12}^{*2}}{N_{11}^* N_{22}^*}$$

and in this case Eq. (8.110) reverts to  $g(\{N_1\})$ , which satisfies condition 2. It is left as an exercise to show that condition 1 is indeed satisfied. Notice that we have arrived at the result for  $g(\{N_i, N_{ij}\})$  not by rigorous combinatorial arguments but by a heuristic extrapolation from  $g(\{N_i\})$ , while satisfying the three conditions that  $g(\{N_i, N_{ij}\})$  is known to obey. Equation (8.110) is not the only function that satisfies the given conditions. It is nonetheless a cleverly constructed function with no adjustable parameters that works well in the vicinity of  $w_{ij}/zkT = 0$ .

## 8.10 Additional reading

Lattice statistics are discussed in (Hill, 1986) but the material is scattered over several chapters (7, 14, 21) and the topic is introduced in the context of magnetization, which is not a very relevant case study for chemical engineers.

A concise to lattice statistics is given in (Prausnitz et al., 1999) (chapter 7) with an emphasis that is more appropriate for chemical engineers, but the coverage is brief.

The quasichemical approximation is discussed in detail in (Guggenheim, 1952). The exposition is rigorous yet accessible to readers with standard background in statistical thermodynamics.

The UNIQUAC method was developed by Prausnitz in the 1970's. The original paper (Abrams and Prausnitz, 1975) is a highly recommended read, especially for graduate students in chemical engineering, as an example that demonstrates how theory informs the development of practical engineering tools, in this case the development of a successful model for the calculation of activity coefficients.

## 8.11 Problems

**Problem 8.1.** Consider a lattice gas without interactions.

- Show that the entropy of the lattice gas goes over to that of the ideal gas when  $\phi \rightarrow \infty$ .
- Obtain the second virial coefficient.
- Show that this lattice gas does not undergo a vapor-liquid phase transition.

**Problem 8.2.** a) Calculate the total number of microstates for the 10 particles on  $5 \times 5$  lattice. If all microstates have the same energy, what is the probability to find all particles occupying just the two top rows of the box?

b) Suppose a lattice with  $N$  sites contains a single molecule with energy  $E$ . Calculate the number of microstates and the entropy of the box. Discuss the result in terms of classical thermodynamics.

c) Suppose a lattice is fully occupied by two different types of molecules,  $A$  and  $B$ . There are  $N_A$  molecules of type  $A$ ,  $N_B$  of type  $B$ . The total number of lattice sites is  $N_A + N_B = N$ . All molecules have the same energy. Calculate the total number of microstates in terms of  $N_A$  and  $N_B$ .

d) Define  $x_A = N_A/N$ , and  $x_B = N_B/N$ . Calculate the molar entropy of the box,  $S/N$  in terms of  $x_A$  and  $x_B$ , assuming that  $N_A$  and  $N_B$  are very large numbers. Discuss the results in terms of statistical thermodynamics.

e) Calculate the probability to find all the molecules occupying the top row of the box.

**Problem 8.3.** Consider the ensemble of configurations that are produced when  $N$  identical molecules are placed on a lattice with  $M$  sites. The volume of a lattice site is  $V_0$ . There is no interaction between the molecules beyond excluded volume.

- Identify the type of the ensemble.
- Calculate the entropy of the system.
- Obtain the equation of state of this system. Does it have the proper behavior at the ideal-gas limit?

Assume the kinetic energy of the molecules is independent of their arrangement on the lattice.

**Problem 8.4.** In surface adsorption, molecules bind weakly to sites on a the surface of solid. If desorption is neglected and if the adsorbed molecules have sufficient surface mobility to move from site to site, this system is equivalent to a two-dimensional molecular fluid. As a model for this system we consider  $N$  particles adsorbed on a two-dimensional lattice that contains  $M$  sites ( $M \geq N$ ), as shown schematically in Fig. 8.10a. Adsorbed particles interact with the substrate with energy  $\epsilon$  but not with each another so that the energy of the system is  $U = N\epsilon$ , independently of how the particles are arranged on the lattice. In this treatment the only contribution to the energy of the system is due to the interaction with the surface, i.e. these particles possess no kinetic energy.

- Calculate the number of distinct arrangements of  $N$  particles in  $M$  lattice sites.
- Obtain the canonical partition function.
- Calculate the constant-volume heat capacity,  $C_V$ .

- d) Calculate the Gibbs energy of the system.  
 e) We define the two-dimensional pressure,  $P$ , of the adsorbed particles as

$$P = - \left( \frac{\partial A}{\partial V} \right)_{TN}$$

where  $A$  is the Helmholtz energy,  $V = Ma_0$  is the two-dimensional “volume” (i.e. area) of the lattice and  $a_0$  is the area of a lattice site. Obtain the equation of state in the form of a relationship between  $P$ ,  $T$ ,  $M$ ,  $N$  and  $a_0$ .

f) The lattice of Fig. 8.10a is placed inside a much larger lattice that acts as a bath. The temperature of the bath,  $T$ , and the chemical potential  $\mu$  of the particles are both fixed (Fig. 8.10b). The boundaries of smaller lattice are permeable to particles, i.e. adsorbed particles can move in and out of lattice  $M$ . Obtain the grand canonical partition function in closed form as a function of  $M$ ,  $\mu$ ,  $T$  and  $\epsilon$ . g) Calculate the mean surface coverage,  $\bar{\theta} = \bar{N}/M$ , inside the smaller lattice as a function of  $\mu$ ,  $T$  and  $\epsilon$ .

h) Calculate the variance  $\text{var}(\theta) = (\overline{\theta^2} - \bar{\theta}^2)$  of the surface coverage  $\theta = N/M$ , and express the ratio  $\sqrt{\text{var}(\theta)}/\bar{\theta}$  in terms of  $\mu$ ,  $\epsilon$ ,  $T$  and  $M$ .

**Problem 8.5.** Consider the following lattice model for a one-component fluid of interacting particles. A lattice of size  $M$  and coordination number  $z$  contains  $N < M$  particles (at most one particle per site). The volume of the lattice is  $V = \nu_0 M$ , where  $\nu_0$  is the volume of the lattice site. Particles interact with their nearest neighbors with a potential  $\phi_{11}$  (the interaction is zero, if a neighbour lattice site is unoccupied). Let  $N_{11}$  be the number of nearest neighbor 1–1 pairs,  $N_{10}$  the number of 1–0 pairs, and  $N_{00}$  the number of 0–0 pairs, where 1 stands for an occupied site and 0 for an empty one.

- a) Express the energy of the system in terms of the number  $N_{11}$  of pairs of occupied sites and write an expression for the configuration integral.  
 b) Calculate the mean number  $\bar{N}_{11}$  assuming that there is no correlation between a particle and its neighbors (i.e. by invoking the assumption that the fluid is random).  
 c) Obtain the partition function assuming that  $N_{11} = \bar{N}_{11}$  for all configurations, i.e., all configurations have the same number of 1–1 pairs, equal to the mean number of pairs in a random fluid.  
 d) Obtain the equation of state and express the result in terms of  $P$ ,  $T$ ,  $\nu = V/N$ ,  $\nu_0$ ,  $\phi_{11}$ , and  $z$ .  
 e) Determine the conditions under which this model predicts vapor-liquid equilibrium.

**Problem 8.6.** A lattice of size  $M$  contains  $N$  identical particles. The volume of a lattice site is  $\nu_0$  and the internal partition function of the particles is  $Q_{\text{int}} = 1$ . There is no interaction between the particles except for the fact that no two particles can occupy the same lattice site.

- a) Write the canonical partition function in terms of  $\nu_0$ ,  $M$ ,  $N$  and  $\Lambda = h/\sqrt{2\pi mkT}$ .  
 b) Suppose this lattice is embedded inside a larger lattice with temperature  $T$  and fixed chemical potential  $\mu$ . The lattice may now contain any number of particles, from  $N = 0$  (totally empty) to  $N = M$  (totally full). Obtain an expression for the grand canonical partition function in terms of  $M$ ,  $\mu$  and  $\beta = 1/kT$ .  
 c) Express the pressure of the grand canonical ensemble in terms of  $\mu$  and  $\beta$ .

Some useful mathematical identities:

$$\sum_{i=0}^M \frac{M!}{N!(M-N)!} x^N y^{M-N} = (x+y)^M$$

$$\lim_{x \rightarrow \infty} \left( 1 + \frac{a}{x} \right)^x = e^a$$

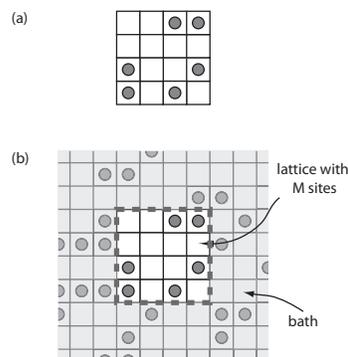


Figure 8.10: See Problem 8.6.

**Problem 8.7.** A lattice with  $M$  sites (the volume of the site is  $v_0$ ) contains  $N_1$  particles of type  $A$  and  $N_2$  particles of type  $B$  with  $N_1 + N_2 < M$ . There is no interaction between particles except for excluded volume, i.e., no two particles can occupy the same site. Both types of particles are structureless, which is to say that their internal partition function is 1.

- Express the number of configurations  $\Omega_c$  in terms of  $M$ ,  $N_1$ ,  $N_2$ .
- Obtain the configuration integral.
- Obtain the canonical partition function.
- Obtain the equation of state.
- Explain why the equation of state is not given by the ideal gas law even though there is no interaction energy.
- Under what conditions does the equation of state revert to the ideal-gas law? Provide a mathematical proof.
- A lattice of size  $M_1$  with  $N_1$  particles of type  $A$  is combined with a lattice of size  $M_2$  with  $N_2$  particles of type  $B$  to form a lattice that of size  $M = M_1 + M_2$  that contains  $N_1$  particles of type  $A$  and  $N_2$  particles of type  $B$  (there is no interaction between particles). This process is conducted by maintaining temperature and pressure the same in all three lattices. Determine the change in entropy as a function of  $N_1$  and  $N_2$ . How does it compare to the ideal entropy of mixing?

**Problem 8.8.** A lattice with  $M$  sites (the volume of the site is  $v_0$ ) contains  $N_1$  particles of type  $A$  and  $N_2$  particles of type  $B$  with  $N_1 + N_2 < M$ . There is no interaction between particles except for excluded volume, i.e., no two particles can occupy the same site. Both types of particles are structureless, which is to say that their internal partition function is 1.

- Express the number of configurations  $\Omega_c$  in terms of  $M$ ,  $N_1$ ,  $N_2$ .
- Obtain the configuration integral.
- Obtain the canonical partition function.
- Obtain the equation of state.
- Explain why the equation of state is not given by the ideal gas law even though there is no interaction energy.
- Under what conditions does the equation of state revert to the ideal-gas law? Provide a mathematical proof.
- A lattice of size  $M_1$  with  $N_1$  particles of type  $A$  is combined with a lattice of size  $M_2$  with  $N_2$  particles of type  $B$  to form a lattice that of size  $M = M_1 + M_2$  that contains  $N_1$  particles of type  $A$  and  $N_2$  particles of type  $B$  (there is no interaction between particles). This process is conducted by maintaining temperature and pressure the same in all three lattices. Determine the change in entropy as a function of  $N_1$  and  $N_2$ . How does it compare to the ideal entropy of mixing?

**Problem 8.9.** a) Show that the equation of state of a lattice gas with nearest neighbor interactions is

$$\frac{P}{kT} = \frac{1}{v_0} \log \frac{V}{V - Nv_0} + \frac{z\phi_{11}v_0^2}{2kT} \left(\frac{N}{V}\right)^2.$$

- Obtain the critical pressure, critical temperature and critical volume in terms of the parameters  $v_0$  and  $a = z\phi_{11}v_0^2/2$ .
- Show that the equation of state can be expressed in the reduced form

$$P_r = \frac{2T_r \log\left(\frac{2v_r}{2v_r - 1}\right) - \frac{1}{v_r^2}}{\log(4) - 1}$$

with  $P_r = P/P_c$ ,  $v_r = v/v_c$ ,  $T_r = T/T_c$ .

## 8.11. Problems

**Problem 8.10.** Consider  $N$  particles with internal partition function  $Q_i$  and the thermal De Broglie length  $\Lambda$ . The particles are placed in a three dimensional lattice of total volume  $V$  that contains  $M$  sites with coordination number  $z$ . A particle interacts only with its  $z$  neighbors. The strength of the interaction per pair is  $\epsilon$  (there is no interaction with unoccupied sites).

- a) Write an expression for the energy of the lattice in terms of the number of particles,  $N$ , and the number of nearest neighbor pairs,  $N_{11}$ .
- b) Obtain the average number of 1–1 pairs,  $\bar{N}_{11}$ , assuming that the fluid is random, i.e., the probability to find a 1–1 pair is the same as when there is no interaction between particles.
- c) Obtain the canonical partition function assuming that the number of 1–1 pairs is equal to  $\bar{N}_{11}$  in all configurations.
- d) Obtain the equation of state.
- e) Under what conditions you expect this model to predict vapor/liquid equilibrium?

**Problem 8.11.** Explain whether the following statements are true or false and why:

- a) If the exchange energy,  $w$ , of two components is zero, the resulting solution is ideal.
- b) The excess entropy of solution is always negative.
- c) Regular solutions exhibit only positive deviations from ideality.
- d) Components that form athermal solutions are fully miscible.

**Problem 8.12.** In this problem we perform additional calculations on the lattice model discussed in lecture 19.

Consider a lattice with  $L$  sites that contains  $N$  particles such that at most one particle may occupy a site. The volume occupied by the lattice is  $V = \nu_* L$ , where  $\nu_*$  is the volume of each lattice site. The energy of this system is the sum of the energies of the individual particles,

$$E = \sum_{i=1}^N E_i,$$

i.e., particles do not interact with each other, and  $E_i$  takes discrete values from the set

$$\{E_*, 2E_*, 3E_*, \dots\}$$

Particles are indistinguishable (they all belong to the same component) but lattice sites are distinguishable (they can be identified by their  $(x, y, z)$  coordinates).

- a) Obtain the microcanonical partition function as a function of  $\mathcal{E}$ ,  $L$  and  $N$ , where  $\mathcal{E} = E/E_*$  and  $E$  is the total energy of the lattice. Simplify your result using the Stirling approximation for  $\log x!$  and that  $\mathcal{E}$ ,  $V$  and  $N$  are all large numbers.

*Hint:* The number of ways to partition an integer number  $\mathcal{E}$  into a sum of  $N$  non zero terms:

$$\binom{\mathcal{E} - 1}{N - 1}$$

- b) Obtain the equation of state. Does this system undergo a vapor-liquid phase transition?

c) Under what conditions does this equation go over to the ideal gas state?

- d) Obtain the  $c_V$ . Make a graph of  $c_V/NE_*$  as a function of  $kT$  from  $kT = 0$  to 2. Discuss the limiting behavior of the  $C_V$  in the two limits  $T \rightarrow 0$  and  $T \rightarrow \infty$ .

e) A lattice with  $L = 5 \times 5 \times 5$  contains  $N = 8$  particles with total energy  $\mathcal{E} = 80$ . Calculate the following probabilities: the probability that each particle occupies one of the eight vertices of the box regardless of their energy; the probability that all eight particles have the same energy regardless of where they are located; the probability that the four particles have the same energy *and* they are located at the eight vertices of the box.

f) According to the hint in part (a) the number of ways to assign a total energy  $\mathcal{E} = 6$  to  $N = 3$  particles is

$$\binom{\mathcal{E}-1}{N-1} = \binom{5}{2} = \frac{5!}{2!3!} = 10$$

This corresponds to the following 10 ways to partition the energy:

$$\begin{aligned} &1+2+3, \quad 1+3+2, \quad 2+1+3, \quad 2+3+1, \quad 3+1+2, \quad 3+2+1, \\ &1+1+4, \quad 1+4+1, \quad 4+1+1, \\ &2+2+2. \end{aligned}$$

Explain why we include all permutations of the particle energies even though particles are *indistinguishable*?

g) Each line on the table of energies in the previous part represents a possible frequency distribution of the energy in all possible permutations that this distribution can be manifested. For example, the second line represents the frequency distribution  $n_1 = 2, n_2 = 0, n_3 = 0, n_4 = 1$ , which states that two particles are at energy level 1 and one particle is at energy level 4. It is clear from this enumeration that some distributions are more probable than others because they are represented by more permutations. Determine the most probable energy distribution given  $\mathcal{E}$ ,  $L$  and  $N$  assuming that  $\mathcal{E} \gg 1$  (you may approximate summations over energy levels with integrals).

**Problem 8.13.** In this problem you will use a Mathematica notebook to sample the microcanonical phase of the lattice gas in problem 8.12 using Monte Carlo (the notebook will be provided). The simulation assigns energy to  $N$  lattice points from the discrete spectrum  $\{1, 2, 3, \dots\}$  such that the total energy is fixed at the desired value. At each step the simulation applies two moves: first, it swaps two randomly chosen sites. This is equivalent to moving particles in space. In the second move we chose two particles at random and redistribute their energy between them. All moves are accepted with probability 1 because all microcanonical configurations are equally probable. For simplicity the simulation is conducted in two dimensions.

- Run the simulation with on a lattice  $L = 30 \times 30 \times 10$  with  $N = 250$  and with total energy  $\mathcal{E} = 6000$ .
- Calculate the energy distribution of the particles, fit an exponential function and obtain the temperature of the lattice from this fit.
- Make a plot of the lattice energy as a function of temperature in the vicinity of  $\mathcal{E} = 6000$  and use it to calculate the  $c_V$  of the lattice at  $\mathcal{E} = 6000, L = 900, N = 250$ .

## Distribution Functions

[This chapter is work in progress] The basic problem in statistical mechanics is the distribution of  $N$  interacting particles in space. The quantity that embodies this distribution is the configuration integral, which we may view as the normalization constant for the probability to obtain a particular arrangement of particles in space. In this chapter we develop connections between the spatial distribution of particles and thermodynamics.

### 9.1 Probability of Configurational Microstate

The fundamental stochastic variable in statistical mechanics is the configurational microstate, namely, the probability to find the particles in a specified region of space. We consider  $N$  identical particles in volume  $V$  that interact via a potential interaction. To write the probability of microstate we first treat the particles as indistinguishable, then we will apply the necessary corrections for indistinguishability. The configuration of this system is described by the multidimensional vector  $\mathbf{r}^N$ ,

$$\mathbf{r}^N = \{\mathbf{r}_1, \mathbf{r}_2 \cdots \mathbf{r}_N\},$$

that specifies the positions of particles  $i = 1, 2, \dots, N$ . With  $N$ ,  $V$  and  $T$  fixed, the probability of configuration  $\mathbf{r}^N$  is<sup>1</sup>

$$\mathcal{P}^{(N)}(\mathbf{r}_1, \dots, \mathbf{r}_N) = \frac{e^{-\mathcal{U}/kT}}{\int_V e^{-\mathcal{U}/kT} d\mathbf{r}_1 \cdots d\mathbf{r}_N} = \frac{e^{-\mathcal{U}/kT}}{Z_N}, \quad (9.1)$$

where  $\mathcal{U} = \mathcal{U}(\mathbf{r}^N)$  is the energy of the configuration. This probability satisfies the normalization

$$\int_V \mathcal{P}^{(N)}(\mathbf{r}_1, \dots, \mathbf{r}_N) d\mathbf{r}_1 \cdots d\mathbf{r}_N = 1, \quad (9.2)$$

in which all vector positions are integrated within the volume of the system. Equation (9.1) gives the probability to find particle 1 in position  $\mathbf{r}_1$ , particle 2 in  $\mathbf{r}_2$  and so on. When particles are *indistinguishable* we are interested in the probability that the positions specified by  $\mathbf{r}^N$  are occupied regardless of the identity of the particle that occupies them. If we denote this probability by  $\rho^{(N)}$ , we must have

$$\rho^{(N)} = N! \mathcal{P}^{(N)}, \quad (9.3)$$

since all  $N!$  permutations among  $N$  distinguishable particles represent the same microstate when the particles are treated as indistinguishable. We integrate  $\rho^{(N)}$  over all  $\mathbf{r}_i$  within the volume of the system and using Eq. (9.2) we find

$$\int_V \rho^{(N)}(\mathbf{r}_1, \dots, \mathbf{r}_N) d\mathbf{r}_1 \cdots d\mathbf{r}_N = N!. \quad (9.4)$$

<sup>1</sup>Since  $\mathbf{r}^N$  is a continuous variable,  $\mathcal{P}^{(N)}$  represents a probability *density* function. For simplicity we will refer to it as “probability.”

## 9.1. Probability of Configurational Microstate

It may seem curious that the result is  $N!$  and not 1, as we expect for a properly normalized probability density. The reason is that the integral in Eq. (9.4) spans the phase space of indistinguishable particles not once, but  $N!$  times. Here is why. When we write integrals of the form

$$\int_V d\mathbf{r}_1 \cdots d\mathbf{r}_N$$

we implicitly assume that all  $\mathbf{r}_i$  span the same volume. If the particles are distinguishable, this integration visits each configuration once, and as Eq. (9.2) shows the probability density in this case has the correct normalization. If the particles are treated as indistinguishable and the integration is conducted as before by letting all  $\mathbf{r}_i$  span the same range, the integral visits every microstate  $N!$  times. It is possible to reformulate the integration limits to ensure that each state is visited once, but this forces a set of integration limits that is different for each variable. It is easier to treat all vector positions in the same way and apply corrections for overcounting afterwards. We will call the phase space before the application of the correction as augmented phase space to emphasize that it contains multiple replicas of all states.

**Indistinguishability and the augmented phase space** Indistinguishability will cause confusion until we recognize that it refers to the *observation* and not to the particles. The particles may always be viewed to be intrinsically indistinguishable; it is the measurement that may or may not be sensitive to the identity of the particles in the same sense that a color-blind observer would not distinguish between identically sized balls of different colors, while a color-abled observer would. Both observers observe the same system but register different results. The general approach to dealing with identical particles is to first treat them as distinguishable, then apply corrections due to indistinguishability. As it turns out, the degree of overcounting is the same for all states, which makes the correction a simple multiplication by the multiplicity factor.

As an example, consider the discrete space of two particles that assume the discrete states  $s = 1, 2, 3, 4, 5$  but cannot occupy the same state simultaneously.<sup>2</sup> We construct the phase space of this system below:

	1	2	3	4	5
1	–	•	•	•	•
2	•	–	•	•	•
3	•	•	–	•	•
4	•	•	•	–	•
5	•	•	•	•	–

This is the phase space of distinguishable particles and consists of 20 states of ordered pairs  $(i, j)$  where the row position ( $i$ ) indicates the state of particle #1 and the column position ( $j$ ) the state of particle #2. To an observer who is indifferent to the identity of the particles the 10 states above the diagonal are the same as those below it. The phase space of indistinguishable particles then consists of 10 states, those above the diagonal, or those below it. To calculate probabilities (i.e., *ratios* of frequencies) with indistinguishable particles we can work with upper half of the table, with the lower half, or with the entire table. We chose to work with the full table<sup>3</sup> with the understanding that it represents an augmented phase space that contains every state twice. If the probability of state is summed over half of the table the result will be 1, as it should, but if it is summed over the augmented table the result is 2.

### 9.1.1 $k$ -particle distribution

Now we consider the following case: we focus on  $k$  positions in space and seek the probability to find particles sitting on them regardless of the position of the remaining

<sup>2</sup>The diagonal have a multiplicity of 1 whereas all other states have multiplicity 2. By not permitting particles to occupy the same state simultaneously we remove the diagonal elements from consideration and simplify the counting of microstates. This simplification does has no effect on the generality of our argument. Moreover, the diagonal elements represent a small fraction of the total states in the table, especially when the number of states per particle is large. Whether the diagonal states are permissible or not, they have a negligible effect at the end.

<sup>3</sup>The upper half of the table is defined by the conditions

$$i = 1 \cdots 4; \quad j = i + 1, \cdots 5,$$

and the lower half by

$$i = 2 \cdots 5; \quad j = 1, \cdots i - 1.$$

These specifications treat the two state variables asymmetrically, i.e.,  $i$  and  $j$  do not span the same range. The complete table, on the other hand, is described by the symmetric conditions

$$i = 1, \cdots 5, \quad j = 1, \cdots 5, \quad i \neq j.$$

We prefer to work with the full table because then all variables are treated symmetrically. The price we must pay for this convenience is that we introduce an augmented phase space on which the normalization constant of probabilities is not 1 but equal to the degree of overcounting.

## 9.1. Probability of Configurational Microstate

$N - k$  particles. Once again we first answer this question for distinguishable particles. The probability to find  $k$  particles, say particles 1 through  $k$ , at positions  $\mathbf{r}_1, \dots, \mathbf{r}_k$  is obtained by integrating the probability density  $\mathcal{P}^{(N)}$  with respect to the positions of the  $N - k$  particles:

$$\mathcal{P}^{(k)}(\mathbf{r}_1, \dots, \mathbf{r}_k) = \frac{1}{Z_N} \int_V e^{-\mathcal{U}/kT} d\mathbf{r}_{k+1} \dots d\mathbf{r}_N, \quad (9.5)$$

and satisfies the normalization

$$\int_V \mathcal{P}^{(k)}(\mathbf{r}_1, \dots, \mathbf{r}_k) d\mathbf{r}_1 \dots d\mathbf{r}_k = 1. \quad (9.6)$$

If the identity of the particles is unimportant, the probability to find positions  $\mathbf{r}_1 \dots \mathbf{r}_k$  occupied is given by  $\mathcal{P}^{(k)}$  multiplied by the number of ways that we can choose  $k$  distinguishable particles out of  $N$  total. There are  $N$  choices for the first particle,  $N - 1$  for the second one, and so on, so the answer is<sup>4</sup>

$$N(N-1) \dots (N-k+1) = \frac{N!}{(N-k)!}.$$

<sup>4</sup>This can also be written as

$$k! \binom{N}{k}.$$

The probability  $\rho^{(k)}$  to find  $k$  indistinguishable particles at the specified positions  $\mathbf{r}_1 \dots \mathbf{r}_k$  then is

$$\rho^{(k)}(\mathbf{r}_1, \dots, \mathbf{r}_k) = \frac{N!}{(N-k)!} \mathcal{P}^{(k)}(\mathbf{r}_1, \dots, \mathbf{r}_k) = \frac{N!}{(N-k)!} \int_V \frac{e^{-\mathcal{U}/kT}}{Z_N} d\mathbf{r}_{k+1} \dots d\mathbf{r}_N. \quad (9.7)$$

The binomial coefficient is the number of ways to choose  $k$  *indistinguishable* particles out of  $N$ . Because the particles are distinguishable we multiply with the number of permutations  $k!$ .

Inserting this result in Eq. (9.6) we find the normalization condition for  $\rho^{(k)}$ :

$$\int_V \rho^{(k)}(\mathbf{r}_1, \dots, \mathbf{r}_k) d\mathbf{r}_1 \dots d\mathbf{r}_k = \frac{N!}{(N-k)!}. \quad (9.8)$$

As with Eq. (9.4), this integral is not 1 because it visits every state  $N!/(N-k)!$  times.

Both probability densities  $\mathcal{P}^{(k)}$  and  $\rho^{(k)}$  are called *distribution functions*. They are functions of  $k$  vector positions and have dimensions of  $V^{-k}$ .  $\mathcal{P}^{(k)}$  treats particles as distinguishable, has a straightforward connection to the configuration integral and is easier to manipulate mathematically;  $\rho^{(k)}$  is physically more relevant because it treats particles as indistinguishable but its integral over space is not unity but equal to the degree of overcounting.

### 9.1.2 Particle density

The correlation function  $\rho^{(1)}$  is the probability density to find a particle at a specific position in space. For an isotropic fluid this probability is uniform because a particle is equally likely to be anywhere in space, i.e.,

$$\rho^{(1)} = \text{const.}$$

From Eq. (9.8) then we obtain

$$\rho^{(1)} \int_V d\mathbf{r}_1 = \rho^{(1)} V = N,$$

or

$$\rho^{(1)} = \frac{N}{V} = \rho = \text{particle density.} \quad (9.9)$$

## 9.2. Correlation functions

For an isotropic system  $\rho^{(1)}$  is simply equal to the particle density. From Eq. (??) we also obtain

$$\mathcal{P}^{(1)} = \frac{1}{V}, \quad (9.10)$$

which states the probability to find a particle at any given location is uniform anywhere inside the system. This result is true in the bulk of an isotropic fluid and does not include edge effects near walls.

### 9.1.3 Random fluid

$\mathcal{P}^k$  and  $\rho^{(k)}$  refer to the probability that  $k$  particles are *simultaneously* found to occupy  $k$  specified points in space, that is, they are *joint* probabilities. If the probability to find a particle at a given position is independent of all other particles, then the joint probability is the product of one-particle distributions:

$$\mathcal{P}^{(k)}(\mathbf{r}_1, \dots, \mathbf{r}_k) = \mathcal{P}^{(1)}(\mathbf{r}_1)\mathcal{P}^{(1)}(\mathbf{r}_2) \dots \mathcal{P}^{(1)}(\mathbf{r}_k) = \frac{1}{V^k}, \quad (9.11)$$

where we have used Eq. (9.10). Using Eq. (9.7), for  $\rho^{(k)}$  we find<sup>5,6</sup>

$$\rho^{(k)} = \frac{1}{V^k} \frac{N!}{(N-k)!} = \rho^k \frac{N!}{N^k(N-k)!}; \quad (\text{random fluid}). \quad (9.12)$$

In the random fluid all distributions functions can be expressed in terms of the density of the fluid. In real fluids particle positions are correlated via molecular interactions. If particles repel strongly, we will find that in the immediate vicinity of a particle the density is lower than that in the bulk fluid. The opposite is true if particles attract. The random fluid represents a system without molecular interactions –an ideal gas. It is an idealization but also a convenient reference state to quantify the effect of interactions on the distribution functions.

<sup>5</sup>Do not confuse the  $k$ -particle probability density,  $\rho^{(k)}$ , with  $\rho^k$ , which is the particle density raised to power  $k$ .

<sup>6</sup>With  $k = 1$  Eq. (9.12) gives  $\rho^{(1)} = \rho$ , which is always true, whether the system is randomly distributed or not. For all other  $k$ , Eq. (9.12) is specific to random distribution.

## 9.2 Correlation functions

We may describe the distribution of real fluids starting with the random distribution and applying corrections for the effect of intermolecular interactions. We extend Eq. (9.12) to real fluid by writing

$$\rho^{(k)} = \rho^k g^{(k)}. \quad (9.13)$$

where  $g^{(k)} = g^{(k)}(\mathbf{r}_1, \dots, \mathbf{r}_k)$  is the  $k$ -particle correlation function. To understand the role of the correlation function, we apply it to a random fluid. Combining with Eq. (9.12) we find

$$g^{(k)} = \frac{N!}{N^k(N-k)!}; \quad (\text{random fluid}) \quad (9.14)$$

and for  $k \ll N$  this becomes<sup>7</sup>

$$g^{(k)} \rightarrow 1 + O(k/N); \quad (\text{random fluid}).$$

<sup>7</sup>See problem 9.1.

**(Inconsistency: problem 9.1)** says  $g^{(k)} \rightarrow \rho^k + O(k/N)$ . We will be concerned with correlations between a small number of particles relative to the total number, in which case the condition  $k \ll N$  is true. In the absence of interactions, the correlation function is asymptotically equal to 1; interactions cause its value to deviate from 1 and this makes the correlation function a useful tool to study the effect of interactions.

The most useful correlation function is the pair correlation function ( $k = 2$ ), which describes correlations between a pair of particles as a function of their position in space.

## 9.2. Correlation functions

Using Eq. (9.13) in combination with (9.7) and (9.5), the pair correlation function is

$$\begin{aligned}
 g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) &= \frac{\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2)}{\rho^2} = \\
 &= \frac{N(N-1)}{\rho^2} P^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \\
 &= \frac{N(N-1)}{\rho^2} \frac{\int_V e^{-\mathcal{U}/kT} d\mathbf{r}_3 \cdots d\mathbf{r}_N}{Z_N}, \quad (9.15)
 \end{aligned}$$

and satisfies the normalization condition

$$\begin{aligned}
 \int_V g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 &= \\
 &= \frac{N(N-1)}{\rho^2 Z_N} \int_V d\mathbf{r}_1 d\mathbf{r}_2 \int_V e^{-\mathcal{U}/kT} d\mathbf{r}_3 \cdots d\mathbf{r}_N = \\
 &= \frac{N(N-1)}{\rho^2}. \quad (9.16)
 \end{aligned}$$

In isotropic fluids the pair correlation function depends on the relative distance  $\mathbf{r}_{12} = |\mathbf{r}_2 - \mathbf{r}_1|$  between the particles and not on the absolute positions  $\mathbf{r}_1$  and  $\mathbf{r}_2$ . Thus we write

$$g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = g^{(2)}(\mathbf{r}_{12}). \quad (9.17)$$

In integrations over  $\mathbf{r}_1, \mathbf{r}_2$ , we may make the variable change

$$(\mathbf{r}_1, \mathbf{r}_2) \rightarrow (\mathbf{r}_1, \mathbf{r}_{12} = |\mathbf{r}_2 - \mathbf{r}_1|)$$

which amounts to a double integration that first runs with respect to the distance of particle from particle 1, followed by integration over all positions of the first particle. The Jacobian of this transformation is 1 and thus we have

$$g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 = g^{(2)}(\mathbf{r}_{12}) d\mathbf{r}_1 d\mathbf{r}. \quad (9.18)$$

It follows that

$$\int_V g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 = V \int_V g^{(2)}(\mathbf{r}_{12}) d\mathbf{r}_1, \quad (9.19)$$

and using Eq. (9.16) we finally have <sup>8</sup>

$$\int_V g^{(2)}(\mathbf{r}_{12}) d\mathbf{r}_{12} = V \left(1 - \frac{1}{N}\right) \rightarrow V, \quad (9.20)$$

The pair correlation function has a simple physical meaning: the product

$$\rho g^{(2)}(\mathbf{r}) d\mathbf{r}$$

is the mean number of particles over all configurations within volume element  $d\mathbf{r}$  at distance  $\mathbf{r}$  from a particle and represents the influence of the central particle at distance  $\mathbf{r}$ . Accordingly,  $\rho g^{(1)}$  is the average density of the fluid at distance  $\mathbf{r}$  from a central particle, and this could be higher or lower than the bulk density  $\rho$ . A value of  $g^{(2)}$  higher than 1 indicates preferential concentration of particles due to attraction, while a value less than 1 indicates a local depletion due to repulsion. For a completely random fluid,  $g^{(2)}(\mathbf{r}) = 1$  at all  $\mathbf{r}$ . We may deduce some general characteristics of the pair correlation function based on its relationship to intermolecular interactions. We use the van der Waals interaction as a guide. At distances smaller than the van der Waals diameter  $\sigma$  the pair correlation function is zero due to very strong repulsion between the electron clouds of the particles.

<sup>8</sup>Another way to write this result is

$$\rho \int_V g^{(2)}(\mathbf{r}_{12}) d\mathbf{r}_{12} = N - 1,$$

which states that the total number of particles in the space outside the central particle is, as it should,  $N - 1$ .

At distances approximately equal to the van der Waals minimum the pair correlation has a maximum that reflects preferential attraction. At very large distances relative to molecular dimensions the pair correlation function is equal to 1: at these distances the influence of the central molecule is very weak and the density of the fluid is equal to the bulk density  $\rho$ . These arguments apply to liquid states. In dilute systems (gas states) the typical distance between particles exceeds the range of interaction and as a result the pair correlation function has much of a discernible structure.

### 9.3 Radial distribution

The radial distribution,  $g(r)$ , gives the mean particle density in a thin spherical shell at distance  $r$  from another particle, and is normalized by the bulk density. It is a mean density that is calculated by taking every particle to be the central particle, then averaging over all of them. The radial distribution function is closely related to the pair correlation function. Its formal definition of  $g(r)$  is

$$g(|\mathbf{r}_2 - \mathbf{r}_1|) = \frac{\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2)}{\rho} \quad (9.21)$$

From Eq. (9.15) then we see that in an isotropic fluid the two functions are related through a simple proportionality,

$$g(\mathbf{r}) = \rho g^{(2)}(\mathbf{r}). \quad (9.22)$$

Functions  $g^{(2)}$  and  $g$  are closely related but not the same.  $g$  is a function of a single variable, the distance between two particles in the fluid.  $g^{(2)}$  is a function of two variables,  $\mathbf{r}_1$  and  $\mathbf{r}_2$ , which refer to the actual positions of the two particles relative to a common reference. The dimensions of the pair correlation function are inverse volume (particle density) while the pair correlation function is dimensionless. The pair correlation function satisfies the normalization (compare with Eq. (9.20))

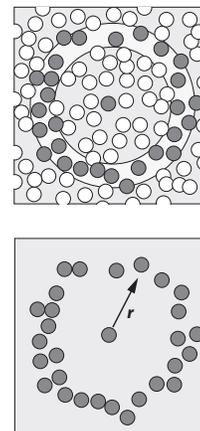
$$\int_0^V g(r) 4\pi r^2 dr = N - 1.$$

The result states that the total number of particles excluding the central particle is  $N - 1$ . Even though the integral is taken from  $r = 0$  to  $\infty$ , it is understood that it does not include the central particle at  $r = 0$ . For large  $N$  this point is inconsequential and we may write  $N - 1 \approx N$ .

The radial distribution, and by extension the pair correlation function as well, is an important tool in the study of condensed phases and can be determined experimentally by scattering experiments. When an electromagnetic wave is scattered by a collection of particles, the interference pattern that develops is directly related to the radial distribution (Fig. 9.2). The radial distribution function is also closely related to several thermodynamic properties, as we see in the next section. This makes it a very important tool in the study of condensed phases.

### 9.4 Mechanical properties

Since the potential energy is the summation of their pairwise interactions, the mean potential energy—and by extension all other properties related to it—can be expressed in terms of the pair correlation function.



**Figure 9.1:** The radial distribution is the average density at distance  $r$  from another particle, normalized by the bulk density.

**Mean potential energy** Since intermolecular interactions are pairwise additive, we may calculate them by the following procedure: given particle at position, all particles at distance  $\mathbf{r}$  have potential energy,  $\phi(\mathbf{r}_{12})$ ; there are  $\rho g(\mathbf{r})(4\pi r^2 dr)$  particles in the spherical shell of radius  $r_{12}$  and thickness  $dr$ . The total potential interaction for the given particle is given by the integral

$$\int \rho g(r)\phi(r)(4\pi r^2)dr.$$

The total potential energy of the system is obtained by multiplying this result by  $N/2$ ,<sup>9</sup>

$$\overline{\mathcal{U}} = 2\pi\rho N \int_0^\infty g(r)\phi(r)r^2 dr. \quad (9.23)$$

This is the mean energy of the system because it is calculated based on the average arrangement of molecules in space. It is the correct ensemble average because the pair correlation function weighs configurations with the proper Boltzmann factor, as we see from Eq. (9.15).

**Pressure** We will calculate pressure from its relationship to the canonical partition function,

$$\frac{P}{kT} = \left( \frac{\partial \log Q}{\partial V} \right)_{TN} = \left( \frac{\partial \log Z_N}{\partial V} \right)_{TN}. \quad (9.24)$$

We will take the volume  $V$  of the container to be a cube; we do this for simplicity knowing that pressure is independent of the shape of the container. The configuration integral then is

$$Z_N = \int_0^{V^{1/3}} \dots \int_0^{V^{1/3}} e^{-\mathcal{U}/kT} dx_1 dy_1 dz_1 \dots dx_N dy_N dz_N \quad (9.25)$$

with all components integrated from 0 to  $V^{1/3}$ . We introduce a new set of coordinates,  $(a_i, b_i, c_i)$  defined as

$$a_i = x_i/V^{1/3}, \quad b_i = y_i/V^{1/3}, \quad c_i = z_i/V^{1/3}$$

and write the configuration integral as

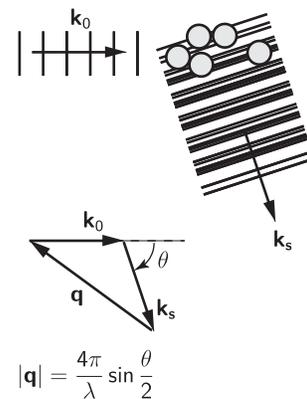
$$Z_N = V^N \int_0^1 \dots \int_0^1 e^{-\mathcal{U}/kT} da_1 db_1 dc_1 \dots da_N db_N dc_N. \quad (9.26)$$

Its derivative with respect to volume is

$$\begin{aligned} \left( \frac{\partial Z}{\partial V} \right)_{TN} &= NV^{N-1} \int_0^1 \dots \int_0^1 e^{-\mathcal{U}/kT} da_1 db_1 dc_1 \dots da_N db_N dc_N \\ &\quad - \frac{V^N}{kT} \int_0^1 \dots \int_0^1 \left( \frac{d\mathcal{U}}{dV} \right) e^{-\mathcal{U}/kT} da_1 db_1 dc_1 \dots da_N db_N dc_N. \end{aligned}$$

To calculate the derivative of  $\mathcal{U}$  with respect to volume, we recall that the potential energy is the sum of pairwise interactions,

$$\mathcal{U} = \sum_{1 \leq i < j \leq N} \phi(|\mathbf{r}_{ij}|).$$



**Figure 9.2:** The electric component of electromagnetic radiation induces a dipole on irradiated molecules that oscillates with the same frequency as the incident radiation. This turns molecules into secondary point sources that emit radiation in all directions (scattering). When radiation hits a collection of particles the scattered waves form an interference pattern that depends on the phase of the scattered waves, which in turn depend on the relative arrangement of the particles (in this figure the interference pattern is shown as a band of black and white lines.)

In this schematic the interference is produced by the superposition of five waves with the same wavelength but whose phases are different depending on the position of each particle relative to the incident wave. The scattered intensity at angle  $\theta$  relative to the incident beam is proportional to the structure factor,

$$S(q) = 1 + \rho \int g(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r},$$

where  $\mathbf{q}$  is the scattering wave vector, defined as the difference between the incident and the scattered vectors, and whose magnitude is  $(4\pi/\lambda) \sin \theta/2$  ( $\lambda$  is the wavelength of the radiation and  $\theta$  is the angle where we observe the scattered radiation). In a scattering experiment we measure the scattered intensity at various angles from which we obtain  $S(q)$ . The pair correlation function is obtained by inverse Fourier transform of the structure factor.

<sup>9</sup>The division by 2 is because the interaction between two particles is counted twice, once with each particle as the central particle in the calculation.

## 9.4. Mechanical properties

This summation contains  $N(N-1)/2$  terms, one for each pair of particles. The distance  $|\mathbf{r}_{ij}| = r_{ij}$  between particles  $i$  and  $j$  is

$$r_{ij} = \left( (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2 \right)^{1/2} = V^{1/3} \left( (a_i - a_j)^2 + (b_i - b_j)^2 + (c_i - c_j)^2 \right)^{1/2}.$$

For the derivative  $d\mathcal{U}/dV$  we now have

$$\frac{d\mathcal{U}}{dV} = \sum \frac{d\phi(r_{12})}{dr_{12}} \frac{dr_{12}}{dV} = \sum \frac{r_{ij}}{3V} \frac{d\phi(r_{12})}{dr_{12}}. \quad (9.27)$$

We now combine these results to obtain the derivative of  $\log Z_N$  with respect to  $V$ :

$$\left( \frac{\partial \log Z_N}{\partial V} \right)_{TN} = \frac{1}{Z_N} \left( \frac{\partial Z_N}{\partial V} \right)_{TN} = \frac{N}{V} - \frac{1}{kT} \left( \frac{1}{Z_N} \int_V \left( \frac{d\mathcal{U}}{dV} \right) e^{-\mathcal{U}/kT} d\mathbf{r}_1 \cdots d\mathbf{r}_N \right), \quad (9.28)$$

with the integrals expressed in terms of the original variables. From Eq. (9.27), the derivative  $d\mathcal{U}/dV$  is a summation of  $N(N-1)/2$  terms. Correspondingly, the integral in the above result consists of  $N(N-1)/2$  identical terms that are integrated over the probability to find two particles at distance  $r_{12}$ :

$$\begin{aligned} \frac{1}{Z_N} \int_V \left( \frac{d\mathcal{U}}{dV} \right) e^{-\mathcal{U}/kT} d\mathbf{r}_1 \cdots d\mathbf{r}_N &= \\ \frac{1}{2} \int_V \frac{r_{12}}{3V} \frac{d\phi(r_{12})}{dr_{12}} \left( \underbrace{\int_V \frac{N(N-1)}{Z_N} e^{-\mathcal{U}/kT} d\mathbf{r}_3 \cdots d\mathbf{r}_N}_{\rho^{(2)} = \rho g} \right) d\mathbf{r}_1 d\mathbf{r}_2 &= \\ \frac{1}{6V} \int_V r_{12} \frac{d\phi(r_{12})}{dr_{12}} \rho^{(2)}(\mathbf{r}_{12}) d\mathbf{r}_1 d\mathbf{r}_2 &= \\ \frac{\rho}{6V} \int_V r_{12} \frac{d\phi(r_{12})}{dr_{12}} g(\mathbf{r}_{12}) d\mathbf{r}_1 d\mathbf{r}_2. \end{aligned}$$

The last step is to change the integration variables from  $(\mathbf{r}_1, \mathbf{r}_2)$  to  $(\mathbf{r}_1, \mathbf{r}_2 - \mathbf{r}_1)$ . The integration with respect to  $\mathbf{r}_1$  gives  $V$  and the above integral becomes

$$\frac{1}{Z_N} \int_V \left( \frac{d\mathcal{U}}{dV} \right) e^{-\mathcal{U}/kT} d\mathbf{r}_1 \cdots d\mathbf{r}_N = \frac{\rho}{6} \int_V r_{12} \frac{d\phi(r_{12})}{dr_{12}} g(\mathbf{r}_{12}) d\mathbf{r}_{12}.$$

Combining with Eq. (9.28), the final expression for the pressure is

$$\frac{P}{kT} = \frac{N}{V} - \frac{N}{6kTV} \int_V r_{12} \frac{d\phi(r_{12})}{dr_{12}} g(\mathbf{r}_{12}) d\mathbf{r}_{12}. \quad (9.29)$$

In spherical coordinates,  $d\mathbf{r}_{12} = 4\pi r_{12}^2$  and the final result can be expressed in the equivalent form,

$$\frac{PV}{NkT} = 1 - \frac{2\pi}{3kT} \int_0^\infty \frac{d\phi(r)}{dr} g(r) r^3 dr, \quad (9.30)$$

where for simplicity we write  $r$  instead of  $r_{12}$ . Notice that the left-hand side is the compressibility factor; in the absence of interactions,  $\phi = 0$  and thus we recover the ideal-gas result,  $PV/NkT = 1$ .

## 9.5 Free energy and the coupling parameter

In the previous section we examined two mechanical properties, mean energy and pressure, and obtained their relationship to the pair correlation function. With mechanical properties in general, this connection is established by writing the mechanical property as an ensemble average over canonical configurations. With non-mechanical properties such as entropy, free energy or chemical potential, a different approach must be used. It is enough to relate one non-mechanical property to the pair correlation function, all others can be obtained through their mutual relationships.

One way is to obtain non mechanical properties in terms of the radial distribution is based on the thermodynamic relationship

$$\bar{E} = \left( \frac{\partial A / kT}{\partial kT} \right)_{VN},$$

which we may integrate to obtain  $A$  at  $T$ . This integration requires a known initial state and this is provided by the condition  $T = \infty$ , which corresponds to the ideal-gas state.

An alternative methodology is based on the so-called coupling parameter. We introduce a set of parameters  $\xi_i$  such that the energy of  $N$  particles is written as

$$U = \sum_{1 \leq i < j \leq N} \xi_i \xi_j \phi(r_{ij}).$$

With  $\xi_i = 1$  for all  $i$ , we obtain the true energy of the system –we say that the particles are fully coupled; with  $\xi = 0$  we turn all interactions off and thus bring the system to the ideal-gas state (particles are fully uncoupled). We may take the more expanded view that these parameters can be changed continuously between 0 and 1, and independently between different particles. Suppose that  $\xi_1$  of particle 1 is set to 0 (for simplicity we will use  $\xi$  to refer to the parameter of particle 1) while all other  $\xi_i = 1$ . This amounts to removing particle 1 from the system. With  $\xi_1 = 1$  (which from now on we will call  $\xi$  for simplicity) particle 1 is fully coupled. By varying  $\xi$  continuously from 0 to 1 we introduce the particle into the system in a continuous manner. The variation of the coupling parameter  $\xi$  corresponds to a thermodynamic path along which we may perform thermodynamic integrations. This path is “hypothetical” in the sense that it represents a mathematical construction, not a physical path that can be implemented experimentally. In fact this is exactly what we do when we calculate residual properties with the reference set at the “hypothetical” ideal-gas state, even though we did not use the term “coupling parameter.”

We will now use the coupling parameter to obtain the chemical potential in a one-component system in terms of the radial distribution. We begin with the fundamental relationship,

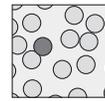
$$\frac{\mu}{kT} = - \left( \frac{\partial \log Q_N}{\partial N} \right)_{TV} = \log \Lambda^3 + \log N - \left( \frac{\partial \log Z_N}{\partial N} \right)_{TV}, \quad (9.31)$$

and express the derivative of the configuration integral as a finite difference in  $N$ ,

$$\left( \frac{\partial \log Z_N}{\partial N} \right)_{TV} = \log \frac{Z_N}{Z_{N-1}}. \quad (9.32)$$

We will now express the configuration integral in terms of the coupling parameter. First we write the energy of the system in the form,

$$\mathcal{Q} = \sum_{j=2}^N \xi \phi(r_{1j}) + \sum_{2 \leq i < j \leq N} \phi(r_{ij}). \quad (9.33)$$


 $Z(\xi = 1) = Z_N$ 

 $Z_1 = V$ 
 $Z_{N-1}$ 
 $Z_{N-1} = VZ(\xi = 0)$ 

**Figure 9.3:** Schematic representation of coupling. The uncoupled state consists of two subsystem, one that contains  $N - 1$  coupled particles plus one that contains the uncoupled particle in the ideal gas state.

## 9.5. Free energy and the coupling parameter

The first summation in this expression contains the contribution of all pairs that contain particle 1 and the second summation contains the contribution of all the other pairs. We define  $Z(\xi)$  as

$$Z(\xi) = \int_V e^{-\mathcal{U}(\xi)/kT} d\mathbf{r}_1 \cdots d\mathbf{r}_N, \quad (9.34)$$

With  $\xi = 1$  we obtain a system of fully coupled particles:

$$Z(\xi = 1) = \int_V e^{-\mathcal{U}_N/kT} d\mathbf{r}_1 \cdots d\mathbf{r}_N = Z_N.$$

With  $\xi = 0$  particle 1 uncouples from the system and we obtain

$$Z(\xi = 0) = \int_V d\mathbf{r}_1 \int_V e^{-\mathcal{U}_N} d\mathbf{r}_2 \cdots d\mathbf{r}_N = V Z_{N-1}.$$

Combining these results we obtain

$$\log \frac{Z_N}{Z_{N-1}} = \log V + \log \frac{Z(\xi = 1)}{Z(\xi = 0)} = \log V + \int_0^1 \frac{d \log Z(\xi)}{d\xi} d\xi. \quad (9.35)$$

This allows us to calculate the derivative of  $\log Z_N$  with respect to  $N$  as an integral over  $\xi$ , from the uncoupled state to the fully coupled one. The physical meaning of this integration is illustrated in Fig. 9.3. The uncoupled state consists of two subsystems, one that contains the  $N - 1$  coupled particles, and one that contains the uncoupled particle. This should now make it clear that uncoupling does *not delete* a particle but it transports it to the ideal-gas state, i.e., in a state of zero interactions. The presence of  $V$  in the final result is precisely due to uncoupling and represents the configuration integral of the uncoupled particle.

The last step is to relate the integral in Eq. (9.35) to the radial distribution. We start with the derivative  $d \log Z(\xi)/d\xi$ , for which we have

$$\begin{aligned} \frac{d \log Z(\xi)}{d\xi} &= -\frac{1}{kT Z(\xi)} \int_V e^{-U(\xi)/kT} \left( \frac{d\mathcal{U}(\xi)}{d\xi} \right) d\mathbf{r}_1 \cdots d\mathbf{r}_N = \\ &= -\frac{1}{kT Z(\xi)} \int_V e^{-U(\xi)/kT} \left( \sum_{j=2}^N \phi(r_{1j}) \right) d\mathbf{r}_1 \cdots d\mathbf{r}_N. \end{aligned} \quad (9.36)$$

The summation inside the last integral contains  $N - 1$  terms and each term produces the same result, an integral that involves the interaction between particle 1 and each one of the other  $N - 1$  particles. Therefore,

$$\begin{aligned} \frac{d \log Z(\xi)}{d\xi} &= -\frac{1}{kT} \int_V \frac{(N-1)}{Z(\xi)} e^{-U(\xi)/kT} \phi(r_{12}) d\mathbf{r}_1 \cdots d\mathbf{r}_N = \\ &= -\frac{1}{NkT} \int_V \phi(r_{12}) \underbrace{\left( \int_V \frac{N(N-1)}{Z(\xi)} e^{-U(\xi)/kT} d\mathbf{r}_3 \cdots d\mathbf{r}_N \right)}_{\rho^{(2)}(r_{12}, \xi) = \rho g(r_{12}, \xi)} d\mathbf{r}_1 d\mathbf{r}_2 = \\ &= -\frac{\rho}{NkT} \int_V \phi(r_{12}) g(r_{12}) d\mathbf{r}_1 d\mathbf{r}_2. \end{aligned} \quad (9.37)$$

We apply the usual change of variables,

$$(\mathbf{r}_1, \mathbf{r}_2) \rightarrow (\mathbf{r}_1, \mathbf{r}_2 - \mathbf{r}_1).$$

## 9.6. Additional reading

The integration with respect to  $\mathbf{r}_1$  gives  $V$ , for the integration in  $\mathbf{r}_{12}$  we use spherical coordinates, and the result is

$$\frac{d \log Z(\xi)}{d \xi} = -\frac{1}{kT} \int_0^\infty \phi(r, \xi) g(r, \xi) 4\pi r^2 dr, \quad (9.38)$$

where for simplicity we have replaced  $r_{12}$  with  $r$ . The final result is obtained by combining Eqs. (9.31), (9.35) and (9.35):

$$\frac{\mu}{kT} = \log \rho \Lambda^3 + \frac{\rho}{kT} \int_0^\xi \phi(r) g(r, \xi) 4\pi r^2 dr. \quad (9.39)$$

The final result involves the integration of a quantity that involves the coupled radial distribution,  $g(r, \xi)$ . Here is how to understand this result: given a value of  $\xi$  that partially couples particle 1 to the rest, we compute the pair correlation function  $g(r, \xi)$ . By repeating with various values of  $\xi$  from 0 to 1 we obtain a family of radial distributions that gradually shift from the radial distribution of the  $N - 1$  particles to that of the  $N$  fully coupled particles. The chemical potential is calculated by integrating over this family of radial distributions. Only the initial and final states of this path represent states of the physical system. Nonetheless, the entire path can be realized by numerical simulation and thus Eq. (9.39) can be used to calculate the chemical potential in a one-component system from such simulations. Once chemical potential is known, all other non-mechanical properties can be calculated from their relationship to  $\mu$ .

## 9.6 Additional reading

1. This chapter follows closely Hill (1986), chapter 17.
2. This material is also discussed in McQuarrie (2001) in chapter 15.

## 9.7 Problems

**Problem 9.1.** a) Show that

$$g^{(k)} = \rho^k + O(1/N)$$

when  $k \ll N$  ( $\rho = N/V$  is the particle density).

b) Obtain the relationship between  $g^{(k)}$  and the configuration integral for  $k \ll 1$ .

c) Calculate the integral

$$\int_V g^{(k)}(\mathbf{r}_1, \dots) d\mathbf{r}_1 d\mathbf{r}_2 \dots d\mathbf{r}_k.$$

and give the result in terms of  $N$ ,  $k$  and  $V$ . Simplify this result if  $k \ll N$ .

**Problem 9.2.** Obtain the relationship between the mean potential energy and the pair correlation function starting with

$$\bar{\mathcal{U}} = \frac{1}{Z_N} \mathcal{U} e^{-\mathcal{U}/kT} d\mathbf{r}_1 \dots d\mathbf{r}_N$$

and show that the final result is the same as Eq. (9.23).

## Bibliography

- D. S. Abrams and J. M. Prausnitz. Statistical thermodynamics of liquid mixtures: A new expression for the excess gibbs energy of partly or completely miscible systems. *AIChE Journal*, 21(1):116–128, 1975. ISSN 1547-5905. doi: 10.1002/aic.690210115. URL <http://dx.doi.org/10.1002/aic.690210115>. 8.10
- O. Bühler. *A Brief Introduction to Classical, Statistical, and Quantum Mechanics*. Courant Lecture Notes. Courant Institute of Mathematical Sciences, New York University, 2006. ISBN 9780821842324. URL <http://books.google.com/books?id=mJ0ongEACAAJ>. 6.10
- P. J. Flory. Thermodynamics of high polymer solutions. *The Journal of Chemical Physics*, 10(1):51–61, 1942. doi: 10.1063/1.1723621. URL <https://doi.org/10.1063/1.1723621>.
- J. W. Gibbs. *Elementary Principles in Statistical Mechanics*. University Press, John Wilson and Son, Cambridge, USA, 1902. (available online from <https://archive.org/details/elementaryprinci00gibbrich>). 5.9, 11, 20, 6.5, 6.6, 26
- J. W. Gibbs. *Elementary Principles in Statistical Mechanics*. Ox Bow Press, Woodbridge, CT, 1981. (reprint of the 1902 edition). 6.10
- A. E. Guggenheim. Statistical thermodynamics of mixtures with zero energies of mixing. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 183(993):203–212, 1944a. ISSN 0080-4630. doi: 10.1098/rspa.1944.0032. URL <http://rspa.royalsocietypublishing.org/content/183/993/203>. 8.9, 27
- E. A. Guggenheim. Number of configurations of molecules occupying several sites. *Nature*, 153:406–407, 1944b. doi: 10.1038/153406b0. 27
- E. A. Guggenheim. Statistical thermodynamics of mixtures with non-zero energies of mixing. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 183(993):213–227, 1944c. ISSN 0080-4630. doi: 10.1098/rspa.1944.0033. URL <http://rspa.royalsocietypublishing.org/content/183/993/213>. 8.9
- E. A. Guggenheim. *Mixtures: The Theory of the Equilibrium Properties of Some Simple Classes of Mixtures Solutions and Alloys*. Oxford University Press, 1952. 12, 8.10
- T. L. Hill. *An Introduction to Statistical Thermodynamics*. Dover, New York, 1986. 1, 18, 1, 8.10, 1
- K. Huang. *Statistical Mechanics*. Wiley, New York, NY, second edition, 1963. 6.9, 6.9
- M. L. Huggins. Solutions of long chain compounds. *The Journal of Chemical Physics*, 9(5): 440–440, 2017/10/28 1941. doi: 10.1063/1.1750930. URL <https://doi.org/10.1063/1.1750930>.

## Bibliography

---

- E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106(4):620–630, 1957. doi: 10.1103/PhysRev.106.620. 5.9, 2, 13
- E. T. Jaynes. Gibbs vs Boltzmann entropies. *American Journal of Physics*, 33(5):391–398, 1965. doi: 10.1119/1.1971557. URL <http://link.aip.org/link/?AJP/33/391/1>. 6.10
- J. N. Kapur. *Maximum Entropy Methods in Science and Engineering*. Wiley Eastern Limited, Brisbane, Australia, 1989. 3
- T. Matsoukas. *Fundamentals of Chemical Engineering Thermodynamics*. Pearson Education, Inc., 2013. 1.9, 2.5, 2
- D. A. McQuarrie. *Statistical Mechanics*. University Science, Sausalito, CA, first edition, 2001. 2.5, 1, 2, 2
- R. K. Pathria and P. D. Beale. *Statistical Mechanics*. Elsevier, third edition, 2011. 6.10
- J. M. Prausnitz, R. N. Lichtenhaler, and E. G. de Azevedo. *Molecular Thermodynamics of Fluid-Phase Equilibria*. Prentice Hall PTR, New Jersey, 3rd edition edition, 1999. 8.10
- Y. A. Rozanov. *Probability Theory: A concise course*. Dover Publications, New York, NY, 1977. 4.8
- J. P. Sethna. *Entropy, Order Parameters, and Complexity*. Oxford University Press, New York, 2006. 6.10
- C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948. 5.9, 2, 3
- J. W. Tester and M. Modell. *Thermodynamics and Its Applications*. Prentice Hall, third edition, 1996. 2.5, 1
- J. Uffink. Compendium of the foundations of classical statistical physics. <https://philpapers.org/rec/UFFCOT>, 2006. 6.10
- N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North Holland Personal Librry, Amsterdam, 1990. 4.8
- L. Wasserman. *All of Statistics*. Springer, New York, NY, 2004. doi: 10.1007/978-0-387-21736-9. 4.8