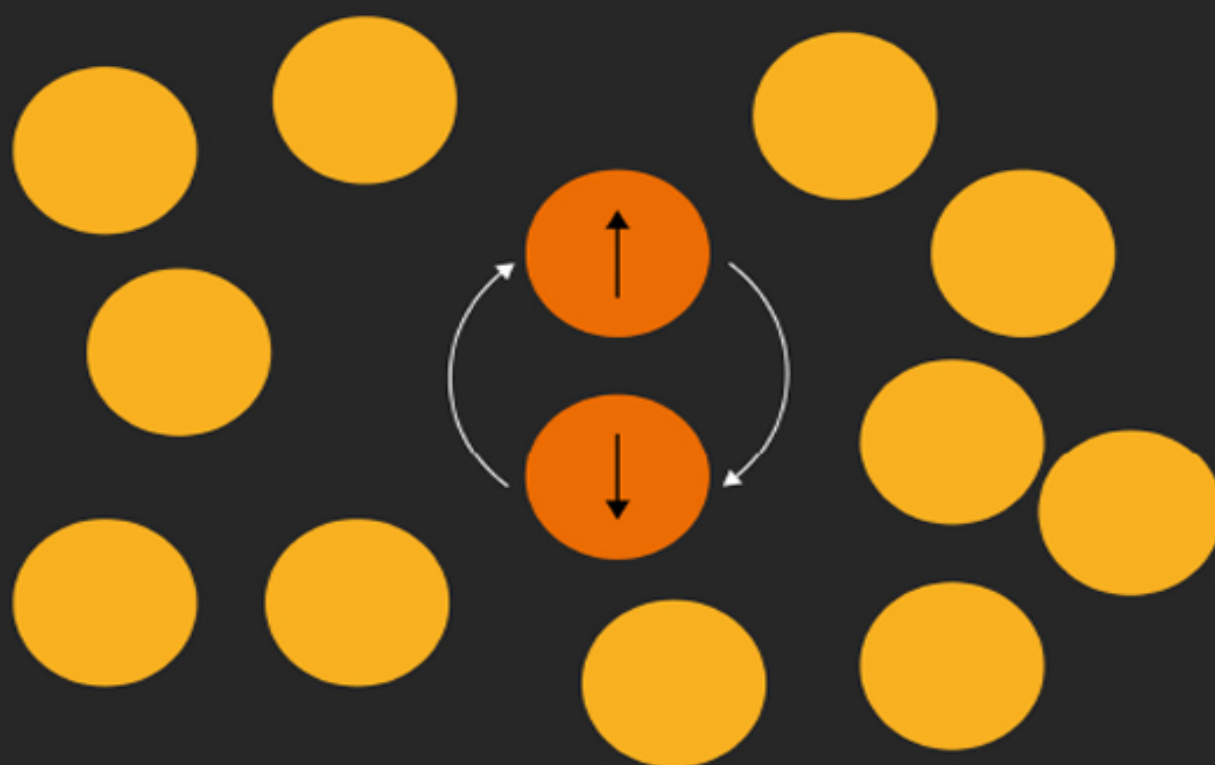# Entropy Beyond the Second Law

Thermodynamics and statistical mechanics for equilibrium, nonequilibrium, classical, and quantum systems

Phil Attard

SECOND EDITION

Entropy Beyond the Second Law (Second Edition)

Thermodynamics and statistical mechanics for equilibrium, non-equilibrium, classical, and quantum systems

Phil Attard

IOP Publishing, Bristol, UK

Contents

4   Entropy for Brownian motion, fluctuations, and stochastic trajectories

5   Equilibrium statistical mechanics and entropy in classical phase space

7   Quantum statistical mechanics as entropy collapse in classical phase space

8   Entropy in Bose–Einstein condensation and the λ-transition

10   High temperature superconductivity: electrons paired by entropy

Preface

'A theory is the more impressive the greater the simplicity of its premises, the more different kinds of things it relates, and the more extended its area of applicability. Therefore the deep impression that classical thermodynamics made upon me. It is the only physical theory of universal content which I am convinced will never be overthrown'

Albert Einstein, quoted in Schlipp P A 1979 *Autobiographical Notes. A Centennial Edition* (Open Court Publishing Company) p 31

'The law that entropy always increases holds, I think, the supreme position among the laws of Nature ...if your theory is found to be against the second law of thermodynamics I can give you no hope'

Eddington A S 1928 *The Nature of the Physical World* (Cambridge: Cambridge University Press) chapter 4

'There is no concept in the whole field of physics which is more difficult to understand than is the concept of entropy, nor is there one which is more fundamental'

Sears F 1950 *Principles of Physics I: Mechanics, Heat, and Sound* (Reading, MA: Addison-Wesley) p 459

'Few physical concepts have caused as much confusion and misunderstanding as has that of entropy'

Lovelock J 1979, *Gaia: A New Look at Life on Earth* (Oxford: Oxford University Press) p 2

'our science colleagues may...help us with 'entropy', which to me is a more difficult concept than anything economics has to offer.'

Tjalling Koopmans, quoted in Faber M, Niemes H, and Stephan G 2005 *Entropy, Environment, and Resources: an Essay in Physico-economics* (Berlin: Springer) p 77

Entropy and the second law

The Second Law of Thermodynamics—entropy increases—is arguably the most fundamental law of nature, as the above quotations from Einstein, Eddington, and Sears attest. And yet for some, entropy itself is a mystery, the cause of confusion and misunderstanding. How can it be that something so fundamental and universal, as evidenced from the interests and expertise of the quoted authors, is so hard to grasp? How can it have happened that thoughtful people appreciate the significance and universal applicability of the Second Law without understanding the nature of entropy, its sole concern?

I blame Clausius, who in the middle of the 19th century initiated the field of thermodynamics by defining entropy as the integrated heat flow divided by the temperature. This is the empirical result of measurement and observation. It is undoubtedly mathematically and quantitatively correct, but it offers nothing in the way of understanding or reason or physical basis. To my mind it obscures rather than clarifies the nature of entropy, and it is little wonder that anyone approaching entropy from this viewpoint should be confused. This definition has made entropy an enigma, and the resultant mystique surrounding it has given rise to a belief in the power of the Second Law of Thermodynamics that is akin to magic. For many, the unquestioning acceptance of both is an act of faith.

It was Boltzmann, shortly after Clausius, who identified entropy with the logarithm of the number of molecular configurations, thereon founding the field of statistical mechanics. Here was not only a concrete physical picture for entropy, but also an explanation of the Second Law of

Thermodynamics. Considering the transitions between states of a system, it is intuitively obvious that these should likely be to states with more configurations rather than fewer, which is precisely the direction of entropy increase. For example, a die is more likely to land showing an even number than a six because there are three of the former and only one of the latter; the state 'even' has more entropy than the state 'six'. Of course the fact that entropy is just a way of counting explains why it and the Second Law have such universal application beyond the fields of thermodynamics and statistical mechanics.

What is a little puzzling is that Boltzmann's insight into the nature of entropy and the reason for the Second Law has not settled the matter and that the confusion over entropy persists amongst some to the present day. I suspect that part of the reason for this is that the fields of thermodynamics and statistical mechanics have remained separate from the beginning. Practitioners have always specialized in one or the other, and the concerns and approaches of each are rather different. This goes deeper than the facile observation that thermodynamics is concerned with macroscopic phenomena and statistical mechanics is focussed on microscopic and molecular behavior. To some extent the two fields appeal to different aspects of the intellect. Thermodynamics has grown as a discipline based on empirical rules and quantitative relationships that have been gleaned from observation and measurement. It can be a very practical field in which practitioners are more interested in actual outcomes and real applications than in model systems or qualitative explanations. It is often taught as a collection of expressions to be memorized without necessarily recognizing the connections between different results.

Statistical mechanics is more focussed on deducing the thermodynamic laws and relationships from the probabilistic behavior of atoms and molecules. It can produce

quantitative predictions of thermodynamic properties from the molecular interactions, although just as often the qualitative behavior of a system is what is sought. Whereas thermodynamics has discovered universal laws by reasoning inductively from empirical observation, statistical mechanics has focussed more on the reasons why those laws exist, and has sought to logically deduce them from the fundamental mechanical laws that govern the Universe.

Statistical theoretical approaches in broader fields such as biology, finance, economics, health, sociology etc often use results and procedures from thermodynamics and from statistical mechanics. However, one should be aware of a serious caveat: the laws governing the behavior of atoms and molecules, which give quantitative and predictive certainty to statistical mechanics, and reproducibility to thermodynamics, generally have no analogue in these broader fields. The statistical data that these fields utilize are empirical and historical, and because there is no underlying mechanical laws, there is no certainty that the same data will hold in the future. Despite this caveat, the general concepts of entropy and of the Second Law of Thermodynamics are frequently applied beyond the disciplines of thermodynamics and statistical mechanics.

This book takes a new and unified look at thermodynamics and statistical mechanics. It attempts to identify their essential conceptual ingredients, and to derive these two disciplines from the ground up as two ends of a spectrum linked by entropy. In the equilibrium case, chapters 1 and 2, much is already known and confirmed by laboratory measurement, computer simulation, and mathematical analysis. What is new in this case is not the final results but their interpretation. For example, in chapter 1 Boltzmann's definition of entropy is generalized as the logarithm of the weight of a state. A small change perhaps, but it accommodates the case that configurations are not equally likely, it shows that probability and entropy are

conceptually equivalent, and it goes beyond physical systems governed by molecular mechanics. These demonstrate the universality of the statistical thermodynamic formalism and its generic potential for application in diverse areas of human endeavor. In this vein the information entropy, often attributed to Shannon in his theory of communications, and also widely exploited by Jaynes in his 'MaxEnt' theory of probability and inductive reasoning, is here shown to be no more than part of the total thermodynamic entropy, as is demonstrated by results for the two-dimensional Ising model of statistical mechanics.

Similarly, what is new in chapter 2 is not so much adumbration of the conventional results of equilibrium thermodynamics, but rather that their derivation proceeds from the perspective of entropy. This shows, for example, the equivalence of free energy and total entropy, and that the minimization of the free energy is just the Second Law of Thermodynamics in a different guise. The analysis shows that the First Law of Thermodynamics is a statement about conservation laws, from which very general conclusions follow about the generic formulation of entropy for an arbitrary system. It is this idea which in part points to the way in which the entropy should be developed beyond equilibrium.

Entropy beyond the second law

'For in the sciences the authority of thousands of opinions is not worth as much as one tiny spark of reason in an individual man'

Galileo 1612 *Third Letter on Sunspots*

'If different persons, not knowing each other's work, have been pursuing different clews that led to the same result, why then it shows that there may be something in it. But if this is only the same story, filtered through two channels,

and reaching me in two ways, then that don't make it any stronger'

Abraham Lincoln 1861 (Philadelphia PA)

'A discovery must be, by definition, at variance with existing knowledge'

Szent-Györgyi A 1972 Dionysians and Apollonians *Science* **176** 966.

This book proceeds beyond the Second Law in several respects. In chapter 3 attention turns to non-equilibrium thermodynamics. For such time dependent systems, the Second Law of *Equilibrium* Thermodynamics is insufficient; it tells the direction of motion but not its speed. Nevertheless, looked at from the right perspective, and with the insight gleaned from the generic analysis of chapter 1, it is possible to develop an entropy for time varying systems that is the basis for a Second Law of *Non-Equilibrium* Thermodynamics.

The non-equilibrium entropy provides a driving force for motion in a generalized sense, as, for example, hydrodynamic fluxes, wave number selection, and pattern formation. Accordingly, steady heat flow (chapter 3), Brownian motion (chapter 4), and molecular motion in an open system (chapter 5), are analyzed from this perspective. The stochastic, dissipative equations that result finally explain and justify some known results, such as the Langevin equation, and the fluctuation-dissipation theorem, and produce some new ones besides. They provide a basis for the formulation of non-equilibrium statistical mechanics (chapter 6).

Finally, chapter 7 formulates quantum statistical mechanics from first principles by showing how the wave function of an open quantum system collapses into entropy states. This gives a type of trace for the partition function and for the statistical average of an operator, which, via a

formally exact transformation, is written as an integral over classical phase space. The leading order term in the expansion gives classical statistical mechanics. Hence the analysis shows that entropy is responsible for the fact that classical mechanics governs the motion of the world around us even though quantum mechanics governs the behavior of the underlying atomic and sub-atomic constituents.

These chapters that go beyond the usual formulation of entropy and of the Second Law lie at the cutting edge of contemporary scientific research. As such many of the results reflect my own analysis and ideas rather than a consensus of opinion. As Szent-Györgyi noted above, this is necessarily the case with any discovery. Some results, such as those for non-equilibrium thermodynamics, may be described as controversial given that they differ from the very many books and papers already published on the subject. But the point of the quotations from Galileo and Lincoln is that multiple books or authors in agreement is not as convincing as independent research (surprisingly enough, even scientists copy). Similarly, authority figures are entitled to serious consideration, but no more than that; science would not be science if authorities were exempt from being questioned or the rules of evidence. Although consensus can be consoling when it is the result of independent and critical thought, if it results from merely copying, or from taking dictation from authority, or from following the herd, then it adds nothing to the sum total of knowledge. Truth is determined by evidence alone. Evidence is not number of proponents, or status of authority, or degree of consensus. Evidence is measurement, computation, mathematical analysis, or reasoned argument.

It is for this reason that this book is structured a little differently to most books on thermodynamics and statistical mechanics. Since the content goes beyond existing knowledge of entropy and of the Second Law of Thermodynamics, I feel a greater than usual obligation to

justify the results with concrete evidence. I have tried to avoid simply asserting a result as commonly held, or of copying a result and citing some other source as its authority. Instead I derive almost everything from first principles, making the book largely self-contained and the presentation uniformly coherent. This does sometimes pose a problem in that general axioms can be introduced or quantities defined before it is obvious that they are useful, but it is the price to be paid for a bottom up approach. Similarly, the novelty of the forms of entropy developed here to go beyond the Second Law means that at this stage the number of concrete examples that illustrate the formalism is rather limited. Nevertheless I make many connections with known theorems and results for which there is a convincing consensus, and hopefully these should provide some sort of guide to the utility of the general approach taken here.

Entropy has given me understanding and insight into the nature of the world, and it has provided a basis for many of my contributions to the scientific literature. I hope in this book to pass on some of this knowledge to the benefit of readers in their own endeavors.

<div align="right">
Phil Attard
Sydney
July, 2017
</div>

Preface to the second edition

In this second edition I have added chapters on Bose–Einstein condensation, superfluidity, and superconductivity. I am particularly pleased to be able to include these because in many ways they provide concrete vindication of the viewpoint taken in the book, namely that entropy is central to the physical Universe, and that formulating it in a generic statistical sense not only provides a basis for novel mathematical treatments, but also gives new insight into diverse physical phenomena. These new applications illustrate just how far the Second Law of Thermodynamics can be extended beyond its traditional purview.

One of the limitations in the first edition is the lack of examples and applications for the classical phase space formulation of quantum statistical mechanics, chapter 7. Apart from the ideal gas and the numerical results for the Lennard-Jones fluid, there is little concrete evidence provided to support the claim that this new approach is practically, not just potentially, useful in real world applications. In this second edition the analysis of Bose–Einstein condensation, chapter 8, which focusses upon the permutation loop expansion, shows that the classical phase space formulation gives a practical computer simulation algorithm for the λ-transition in ^{4}He, which is tested for a Lennard-Jones model. It interprets the λ-transition as the transition from position to momentum permutation loop dominance, and it gives the new physical interpretation of condensation as signifying multiple highly-occupied low-lying momentum states, rather than the more conventional ground energy state.

This new understanding extends to the phenomenon of superfluidity, chapter 9, where theoretical prediction and

experimental measurement are compared for the critical velocity and for the fountain pressure. Equations of motion are given, which explain how condensed bosons collide with uncondensed bosons without dissipating momentum. And in chapter 10, the phenomenon of Cooper pairing of electrons, which is central to the understanding of superconductivity, is formulated again on the basis of the permutation loop expansion. The difference from the conventional Bardeen–Cooper–Schrieffer (BCS) theory that invokes lattice vibrations is that here it is the potential of mean force that binds the electron pair. This is a temperature-dependent potential, and it is therefore of entropic origin, which fits nicely in with the central premise of this book. A detailed exploration of the oscillatory transition in the pair potential of mean force for the one component plasma, using the classical equilibrium techniques of chapters 2 and 5, shows that the predicted superconducting transition occurs within the range measured for high temperature superconductors for realistic parameter values.

The treatment of Bose–Einstein condensation, superfluidity, and superconductivity depends upon the loop expansion for the permutation entropy due to wave function symmetrization. I have therefore added section 7.7 which derives this in detail for single-particle states, and illustrates it with the well-known example of a system of independent quantum oscillators, section 7.7.4, albeit from the new perspective of the loop expansion.

I've also taken the opportunity of this second edition to correct a number of typographical errors in the original. The titles of the original chapters have changed, but their content remains largely the same. I hope that readers find the new chapters and sections a worthwhile extension of the book and an informative and perhaps surprising illustration of the way in which novel types of entropy arise in important applications.

Phil Attard
Sydney
June, 2022

Author biography

Phil Attard

Dr Phil Attard has researched broadly in statistical mechanics, thermodynamics, and colloid and surface science. He has held academic positions at various universities in Australia, Europe, and North America, and he was a Professorial Research Fellow of the Australian Research Council. He has authored some 150 papers, 10 review articles, and 5 books.

Attard has made seminal contributions to the theory of electrolytes and the electric double layer, to measurement techniques for atomic force microscopy, and to computer simulation and integral equation algorithms for condensed matter. He is perhaps best known for his discovery of nanobubbles and for his role in establishing their nature.

Recent research has focused on non-equilibrium thermodynamics and statistical mechanics, and on the formulation of quantum statistical mechanics. He has identified a new type of entropy—the second entropy—as

the variational principle for non-equilibrium thermodynamics, and he has derived the general form of the non-equilibrium probability distribution for statistical mechanics. The theory provides a coherent approach to non-equilibrium systems and to irreversible processes, and has led to the development of stochastic molecular dynamics and non-equilibrium Monte Carlo computer simulation algorithms. His formulation of quantum statistical mechanics in classical phase space has led to the permutation loop expansion for wave function symmetrization, and has provided new insight into Bose–Einstein condensation, superfluidity, and high temperature superconductivity.

Previous books:

Attard P 2002 *Thermodynamics and Statistical Mechanics: Equilibrium by Entropy Maximisation* (London: Academic)

Attard P 2012 *Non-Equilibrium Thermodynamics and Statistical Mechanics: Foundations and Applications* (Oxford: Oxford University Press)

Attard P 2015 *Quantum Statistical Mechanics: Equilibrium and Non-Equilibrium Theory from First Principles* (Bristol: IOP Publishing)

Attard P 2018 *Entropy Beyond the Second Law. Thermodynamics and Statistical Mechanics for Equilibrium, Non-Equilibrium, Classical, and Quantum Systems* (Bristol: IOP Publishing)

Attard P 2021 *Quantum Statistical Mechanics in Classical Phase Space* (Bristol: IOP Publishing)

Entropy Beyond the Second Law (Second Edition)

Thermodynamics and statistical mechanics for equilibrium, non-equilibrium, classical, and quantum systems

Phil Attard

Chapter 1

Entropy counts: universal nature of the statistical entropy

'How do I love thee, let me count the ways'

(Browning 1850)

'You should call it entropy, because nobody knows what entropy really is'

(von Neumann to Shannon circa 1939)

'Shannon entropy and Boltzmann entropy are completely unrelated'

(Kline 1999)

The main aim of this chapter is to define entropy in its generic form. Several applications and illustrations are also given as an aid to understanding and applying entropy.

One of the issues that will be addressed in this chapter is the meaning of the Gibbs–Shannon information entropy,

$$S = -k_\mathrm{B} \sum_\alpha \wp_\alpha \ln \wp_\alpha.$$

(1.1)

Here k_B is Boltzmann's constant and $\wp_\alpha$ is the probability of the state α of the system. This equation appears in almost all books on statistical mechanics or statistical thermodynamics, and it would be fair to call this the conventional expression for the entropy. It was given explicitly by Boltzmann (1866) and by Gibbs (1902). Upon it is founded Shannon's (1948) information theory and Jaynes' (1957, 2003) maximum entropy formulation of statistical mechanics and probability theory.

In my opinion, there is a great deal of misunderstanding about this expression. Most workers believe that it is 'the' entropy, and that it should be maximized. And yet there are others who believe that it is unrelated to the thermodynamic entropy. I myself believe neither, and in section 1.1.3 I will give the derivation of the full expression for the entropy and explain the proper use of this particular expression for the so-called information entropy.

1.1 Entropy, weight, probability, and information

One might think that any book on entropy, particularly any book that focuses on its role in thermodynamics and statistical mechanics, would begin with the Second Law of Thermodynamics,

All systems evolve in the direction of increasing entropy.

(1.2a)

Alternatively,

The entropy increases during spontaneous changes of a system.

(1.2b)

The very simplicity of the law underscores its universality and profundity.

But the problem with beginning with the Second Law is that entropy is undefined and unexplained by it. And in consequence the reason for the law is not evident. To accept the law as merely an empirical fact confirmed by countless measurement is to sell reason short. From that limited point of view the Second Law of Thermodynamics is nothing but superstition, a magic spell to mesmerize the credulous, an incantation to be memorized and repeated with mind closed to further possibilities and new applications. Much better instead to understand the nature of entropy, for then one might hope to know when the law does not apply (in fact, systems where entropy has spontaneously decreased are quite common), and to develop an appropriate version of the law for new situations beyond those that form its empirical basis.

Another reason to eschew the Second Law of Thermodynamics as a starting point is that it is the *second* law. If ordinal numbers mean anything, then surely the normal precedent should be followed by beginning with the *First* Law of Thermodynamics, which is sometimes stated as

$$\text{The change in energy is the work done plus the heat flow.} \tag{1.3a}$$

But in many ways it is preferable to give it in the form

$$\text{Energy cannot be created or destroyed,}$$
$$\text{but it can be converted from one form to another.} \tag{1.3b}$$

In short,

$$\text{Energy is conserved.} \tag{1.3c}$$

An orderly account would obviously discuss the First Law of Thermodynamics before the Second Law. But this raises the question of how to relate the First Law to entropy.

1.1.1 Conservation, counting, and linear additivity

The First Law is based upon the concept of conservation. Things that are conserved can be counted. And counting is what defines entropy.

The ability to count is a fundamental human ability that occurs in a core part of the brain, which indicates that it is very old in evolutionary terms. Our ancestors no doubt counted things like tribe members, animals, and possessions. These are all concrete objects that we would say in the present context share the property of being conserved. There would be little point in, say, counting the coins in your purse if the coins appeared or disappeared at random. One can shake the purse, or move it from one place to another without changing the number of coins.

This is not to say that counting is restricted to conserved objects. Immaterial things such as kisses can be counted without invoking any conservation law. But certainly in the present context it is the counting of conserved objects that is the prime focus.

Closely connected to the notion of conservation is the notion of linear additivity. For example, one can transfer some coins from one purse to another. The total number of coins is conserved, which is to say that the sum of the coins in the two purses is the same before and after the transfer.

The notion of linear additivity is not identical to that of conservation. For example, the square root of the number of coins in a purse is conserved. But the sum of the square roots of the coins in two purses is not the same after a transfer as before. Obviously linear additive conserved variables are the easiest to deal with. And for all practical purposes we can restrict attention to the conservation of linear additive quantities. Of course it is no

accident that the quantities that form the basis for scientific theories are precisely those quantities that are both conserved and linear additive. Examples include energy, momentum, mass, and charge. More familiar to non-scientists, quantities like length, area, volume, and number are also conserved and linear additive.

Some quantities of value in the physical sciences are linear additive without necessarily being conserved. As we shall see, entropy is one such linear additive, non-conserved quantity. As we shall also see, it is most useful to formulate entropy as dependent on linear additive quantities, which mostly but not always are also conserved quantities. Part of the challenge in developing scientific theories from first principles is in identifying and formulating the linear additive quantities that will form its basis.

One of the direct consequences of linear additivity is the thermodynamic concept of extensivity. Essentially linear additive conserved variables scale with the size of the system, other things (intensive variables) being equal. Intensive variables are neither linear additive nor conserved; examples include pressure, temperature, voltage and surface tension. It will be argued that entropy is best formulated as dependent solely on extensive variables. This very simple idea of identifying the extensive variables and ensuring the proper scaling with system size will be used profitably in several chapters below to guide the development of quite sophisticated theory.

1.1.2 Probability, weight, and entropy

From counting it is but a short step to probability, which leads directly to entropy. Probability theory is best formalized in terms of set theory. With the exception of parts of chapter 7, classical probability theory is invoked throughout.

We suppose that the system exists in a variety of states. These states can be microstates or macrostates. Microstates are the smallest indivisible states of the system. For example the specified positions and momenta of all the particles. A macrostate is a set of microstates with the same value of some physical observable. A collective of macrostates correspond to the same physical observable, for example the energy collective.

The set of all microstates is complete (the system at any one time must be in a particular microstate), and the microstates are disjoint (the system cannot be in more than one microstate at a time).

The set of all macrostates of a given collective is complete, and the macrostates of a given collective are disjoint. Degenerate macrostates are forbidden in the formalism, which means that there is a one-to-one relationship between the macrostate labels and the values of the corresponding physical observable. The system can be simultaneously in macrostates of different collectives.

We are interested in the probability of the system being in a particular state. We label microstates by a Roman letter and macrostates by a Greek letter. We suppose that each microstate j has a non-negative weight w_j. The weights are usually defined up to a positive scale factor. Weight is the generalization of number when the objects or states being counted are not identical.

For the general formulation of the theory it does not matter how these weights come about or are calculated. For the application of the theory to a specific problem it does. The most convenient situation is when the microstates can be chosen such that they have equal weight, which may then be set to unity. In this case weight is the same as number. This is not necessary for either the formal development of the theory or for specific applications.

The weight of a macrostate α is the total weight of the microstates that it contains,

$$W_\alpha = \sum_j w_j \delta(A_j - A_\alpha) = \sum_{j \in \alpha} w_j.$$

The first equality contains a Kronecker-δ function ($\delta(0) = 1$, $\delta(x) = 0$, $x \neq 0$). This follows because the microstate j is in the macrostate α if the value of the relevant observable for the microstate is equal to the value that defines the macrostate, $A_j = A_\alpha$. In the event that the microstates have equal weight and that this is set to unity, then the macrostate weight is just the number of microstates that it contains.

The total system weight is the sum of the weights of the microstates. Because a collective is complete and disjoint, this is also equal to the sum of the macrostate weights,

$$W = \sum_j w_j = \sum_\alpha W_\alpha.$$

The probability of the system being in the microstate j is just

$$\wp(j) = \frac{w_j}{W}.$$

This is obviously normalized, $\sum_j \wp(j) = 1$. Similarly the probability of the system being in the macrostate α is just

$$\wp(\alpha) = \frac{W_\alpha}{W}.$$

Consider now two different macrostate collectives, for example energy and number of particles (without enforcing any conservation laws at this time). Label the macrostates of these by α and β. Because macrostates of different collectives are not disjoint, the probability that the system simultaneously has value α for the first observable and β for the second is

$$\wp(\alpha, \beta) = \frac{1}{W} \sum_j \delta(A_j - A_\alpha)\delta(B_j - B_\beta)w_j$$
$$= \frac{1}{W} \sum_{j \in A_\alpha \cap B_\beta} w_j.$$

This is proportional to the total weight of the set formed by the intersection of the two macrostates, figure 1.1.
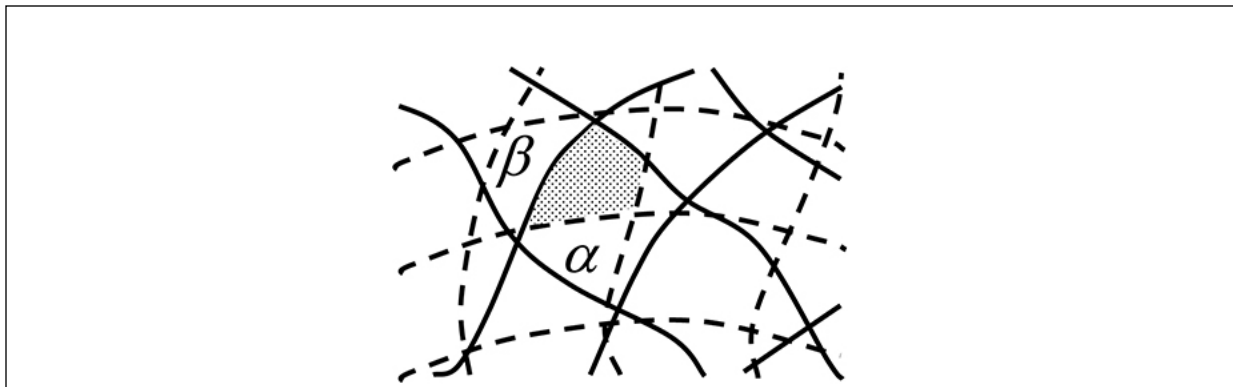
Figure 1.1. Two macrostate collectives. The α macrostate of one (solid lines) and the β macrostate of the other (dashed lines) are labeled. The joint state $\alpha\beta$, their intersection, is shaded.

Given that the system is in the macrostate α, then the probability that it is in the macrostate β is

$$\wp(\beta \mid \alpha) = \frac{1}{W_\alpha} \sum_{j \in A_\alpha \cap B_\beta} w_j$$
$$= \frac{\wp(\alpha, \beta)}{\wp(\alpha)}.$$

(1.9)

This is called the conditional probability, and it is usually written in the form of Bayes' theorem

$$\wp(\alpha, \beta) = \wp(\beta \mid \alpha)\wp(\alpha).$$

(1.10)

By the usual rules of set theory, the probability that the system is in the macrostate α, or in the macrostate β is

$$\wp(A_\alpha \cup B_\beta) = \wp(\alpha) + \wp(\beta) - \wp(\alpha, \beta).$$

(1.11)

The final term corrects for double counting of the microstates in the intersection of the two macrostates.

Now consider the case that the total system is made of two independent, non-interacting sub-systems. The macrostates of interest for the first sub-system may be labeled α and that for the second β. It does not matter in this case whether or not these correspond to the same physical observable because the sub-systems are independent. The weight of the macrostate $\alpha\beta$ of the total system is the product of the weights of the individual systems,

$$W_{\text{tot}}(\alpha\beta) = W_1(\alpha)W_2(\beta).$$

(1.12)

(This is readily checked by taking the weight equal to the number of microstates.) The total weight of the total system is also just the product of the total weights,

$$W_{\text{tot}} = \sum_{\alpha,\beta} W_{\text{tot}}(\alpha\beta)$$
$$= \sum_\alpha W_1(\alpha) \sum_\beta W_2(\beta)$$
$$= W_1 W_2.$$

(1.13)

Correspondingly, the probability for the total macrostate is the product of the individual probabilities,

$$\wp_{\text{tot}}(\alpha\beta) = \wp_1(\alpha)\, \wp_2(\beta).$$

As mentioned in the preceding subsection, the most convenient physical variables are those that are linear additive. The above example of two independent sub-systems shows that weight and probability are multiplicative rather than additive. But this deficiency is easy to overcome by taking logarithms. One simply defines the entropy as the logarithm of the weight. The total entropy is

$$S = k_\mathrm{B} \ln W.$$

Here $k_\mathrm{B} = 1.38 \times 10^{-23} \mathrm{J} \ \mathrm{K}^{-1}$ is Boltzmann's constant. In the event that the microstates have equal unit weight, W is just the total number of microstates and this corresponds to Boltzmann's original definition of entropy. Similarly, the entropy of a microstate is

$$S_j = k_\mathrm{B} \ln w_j,$$

and the entropy of a macrostate is

$$S_\alpha = k_\mathrm{B} \ln W_\alpha.$$

Because weights are usually defined up to a multiplicative constant, entropy is also defined up to an additive constant.

It follows from these that the probability of a state is proportional to the exponential of its entropy,

$$\wp(\alpha) = \frac{w_\alpha}{W} = \frac{\mathrm{e}^{S_\alpha/k_\mathrm{B}}}{W}.$$

Of course the same functional form holds also for microstates. It is important to note that there is a one-to-one relationship between entropy and probability (apart from the additive constant, which anyway can be absorbed into the total weight). This means, for example, it would make no sense to attempt to maximize the entropy with respect to the probability.

Entropy is commonly identified with disorder and unpredictability. This is a reasonable picture because entropy increases with weight, and weight can be thought of as number. Hence at the simplest level, the greater the entropy of a macrostate, the more microstates that it contains. Each of these microstates is a possible configuration of the system, and it is harder to predict which one the system is actually in when there are many possibilities than when there are just a few possibilities. Furthermore there are many more disordered arrangements of objects than there are ordered arrangements. This is illustrated in figure 1.2. One can see that there is just one microstate in the ordered 'phobic' macrostate (the first line, no like circles touching) and just two microstates in the ordered 'philic' macrostate (the second line, all like circles clustered together). With such a definition of order, one can see that in general the majority of arrangements are disordered. Hence again there is a consonance between increasing entropy and increasing disorder.
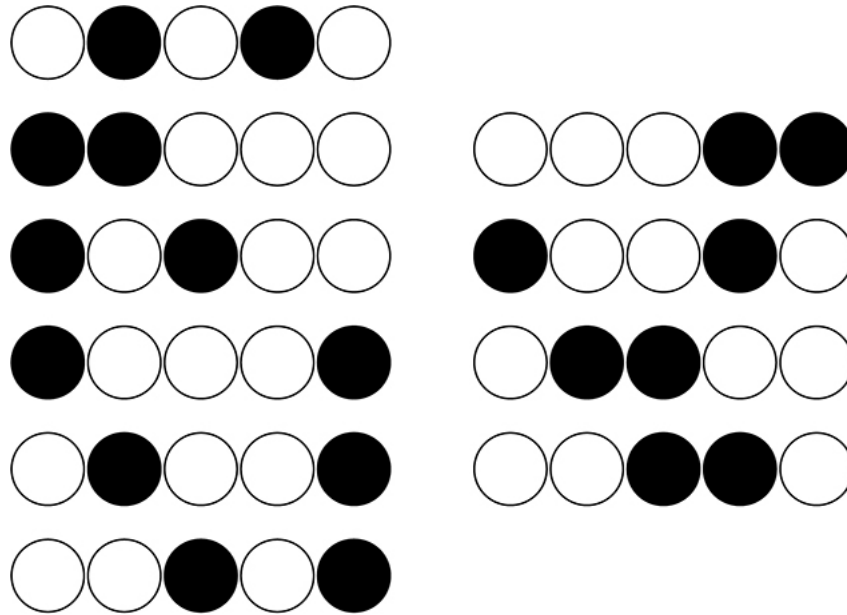
Figure 1.2. All 10 arrangements of two filled and three empty circles in a line. The first line shows the 'phobic' and the second shows the 'philic' macrostate.

1.1.3 Information entropy

In view of the above definitions it is straightforward to write the total entropy as a functional of the macrostate probability (Attard 2000, 2012),

$$
\begin{aligned}
S &= k_{\mathrm{B}} \ln W \\
&= k_{\mathrm{B}} \sum_{\alpha} \wp_{\alpha} \ln W \\
&= k_{\mathrm{B}} \sum_{\alpha} \wp_{\alpha} \left[\ln W_{\alpha} - \ln \frac{W_{\alpha}}{W} \right] \\
&= \sum_{\alpha} \wp_{\alpha} [S_{\alpha} - k_{\mathrm{B}} \ln \wp(\alpha)].
\end{aligned}
$$

(1.19)

This formula holds as well for microstates as for any collective of macrostates. Comparing this to the conventional result, equation (1.1), one must conclude that both cannot be correct.

Clearly equation (1.1) is missing the term $\sum_{\alpha} \wp_{\alpha} S_{\alpha}$, which can be interpreted as the average entropy per macrostate. The term that equation (1.1) does include, $-k_{\mathrm{B}} \sum_{\alpha} \wp_{\alpha} \ln \wp_{\alpha}$, can be interpreted as the uncertainty due to the system having access to multiple macrostates. In any case, one needs both terms to give the total entropy of the system. Further, both terms are necessary for the total entropy of the system to be invariant to the representation: both the functional form of the formula and the numerical result for the total entropy are the same irrespective of the collective of macrostates or microstates that are used.

Figure 1.3 shows the exact entropy per site for the two-dimensional Ising model. This is a classic benchmark in statistical mechanics, with Onsager's solution representing perhaps the first exact solution of a non-trivial, physically realistic model that possesses a phase transition (Baxter 1982). The total entropy, which is the logarithm of the partition function, increases with increasing coupling between the spins. At high coupling this is dominated by the reservoir entropy, which is minus the average energy divided by the temperature. At low coupling it is dominated by the sub-system entropy, which is essentially the weighted number of spin configurations in the most likely energy state. When the spins interact weakly, they are free to randomly and independently orient, which increases their disorder and the entropy of the sub-system.

Figure 1.3. The entropy per site for the Ising model on a square lattice as a function of the coupling parameter (canonical equilibrium system). The curves are the exact result for the total (solid), reservoir (dotted), and sub-system (dashed) entropy. The symbols are the Gibbs–Shannon information entropy using the four-spin probability function (Markov superposition formula (Attard 1999)).
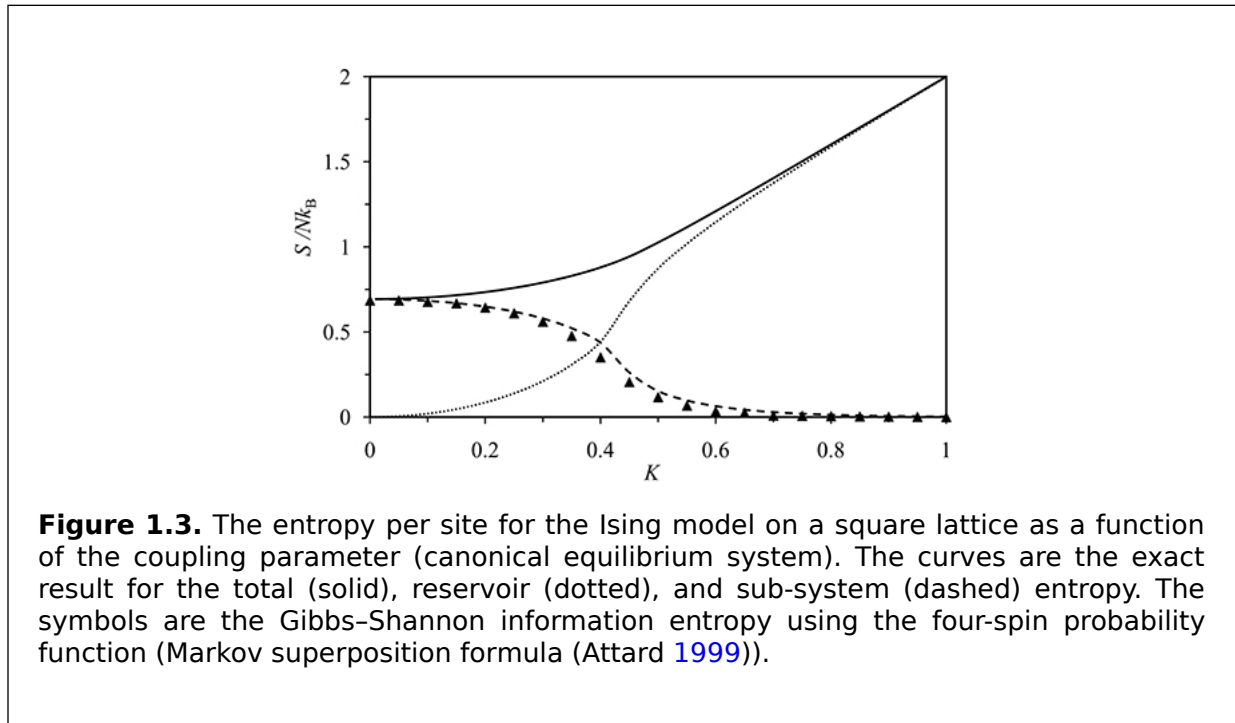
It can be seen in figure 1.3 that the Gibbs–Shannon information entropy, equation (1.1), evaluated in this case using the four-spin probability $\wp^{2,2}(\sigma^4)$, gives only the sub-system part of the total entropy (see section 5.6 for details of the calculation). It neglects the reservoir entropy, which is effectively an internal entropy for each spin microstate. One can see that when the spins are highly correlated, the Gibbs–Shannon entropy is a negligible part of the total entropy.

One of the major uses of equation (1.1) has been in Jaynes' (1957, 2003) maximum entropy theory, in which the probability distribution that maximizes $S[\wp]$ is chosen. A similar use of $S[\wp]$ is made in Shannon's (1948) information theory. But in both cases the procedure is dubious for two reasons. First, the idea is obviously based upon the Second Law of Thermodynamics. We have not yet derived the Second Law of Thermodynamics but when we do we shall show that it holds for specific types of systems and variables. It does not hold for the probability distribution. It does not even make sense to say that the entropy is a maximum with respect to the probability distribution because there is a one-to-one relationship between the entropy of a state S_α and the probability of a state $\wp_\alpha$. The total entropy of the system is fixed by the sum of the entropies of the macrostates and it is pointless to attempt to use the total entropy to find the macrostate probability. The second reason that maximizing equation (1.1) with respect to the probability is pointless is that it

only gives part of the entropy of the system, not the total entropy. Even if the Second Law of Thermodynamics were applicable, then it would refer to the total entropy, not part of the total entropy. What would be the use of maximizing part of the entropy?

In some books one sees it written that equation (1.1) can be shown to be true for equally likely states, as if this justifies the formula more generally. In this case the probability of a microstate is just $\wp_i = 1/n$, where the total number of microstates is $n = \sum_i$. Since the states are equally likely, their weight may be set to unity, $w_i = 1$, and their entropy may be set to zero, $S_i = k_B \ln w_i = 0$. In this case both equations (1.1) and (1.19) give the total entropy correctly as $S = k_B \ln n$. But there is no point in using equation (1.1) for equally likely states because if you know the number of equally likely states then you already know the total entropy. There is not even any point in using equation (1.19) more generally because if you know S_α, which is required by that expression, then you already know the macrostate weight W_α and probability $\wp_\alpha$, the total weight $W = \sum_\alpha W_\alpha$, and the total entropy $S = k_B \ln W$. And finally, the fact that equation (1.1) is true for equally likely states does not make it true in general or useful in particular.

The reader may be interested to pinpoint where Shannon (Shannon and Weaver 1949, appendix 2) and Jaynes (2003, section 11.3) go wrong in their derivation of the entropy as a functional of the probability, equation (1.1). With the microstate probability $\wp_i$ and the macrostate probability $\wp_\alpha$ they say that the entropy of the system can be written as a functional of these with an additional term representing the uncertainty of locating the microstate within each macrostate,

$$S[\wp_i] = S[\wp_\alpha] + \sum_\alpha \wp_\alpha S_\alpha, \quad S_\alpha \equiv S[\wp(i \mid \alpha)].$$

(1.20)

Here the macrostate entropy $S_\alpha \equiv S[\wp(i \mid \alpha)]$ is meant to be the same functional of the probability as $S[\wp_i]$ and $S[\wp_\alpha]$, with the conditional probability of microstates in the macrostate α as its argument.

As I have pointed out in greater detail elsewhere (Attard 2012), this equation has an internal contradiction. The notation in equation (1.20) implies that the microstate functional $S[\wp_i]$ and the macrostate functional $S[\wp_\alpha]$ are the same function of their arguments. But this contradicts the content of equation (1.20), which implies that $S[\wp_i]$ is the total entropy and that $S[\wp_\alpha]$ is only part of the total entropy. The problem is that whilst Shannon and Jaynes accept that macrostates have internal entropy $S[\wp(i \mid \alpha)]$ (and so $S[\wp_\alpha]$ is only part of the total entropy), they assume that microstates do not ($S[\wp_i]$ is the entire total entropy). The last assumption is incorrect for microstates with non-uniform weight. The way to resolve this discrepancy is to rewrite equation (1.20) in terms of the macrostate probability with the total entropy explicit,

$$S_{\text{total}} = S[\wp_\alpha] + \sum_\alpha \wp_\alpha S_\alpha, \quad S_\alpha \equiv S[\wp(i \mid \alpha)].$$

(1.21)

Using the subsequent arguments of Shannon (Shannon and Weaver 1949, appendix 2), and of Jaynes (2003, section 11.3), this can be solved to yield the correct expression for the entropy as a functional of the probability, equation (1.19) (Attard 2012), as will now be shown.

Following Shannon and Jaynes, for uniform microstate, $\wp_i^u = 1/n$, and uniform macrostates, $\wp_\alpha^u = m/n$, suppose that $S[\wp_i^u] = \sigma(n)$, $S[\wp_\alpha^u] = \sigma(n/m)$, $S[\wp^u(i \mid \alpha)] = \sigma(m)$, and that $\sigma(1) = 0$. Then one has

$$S_{\text{total}} = \sigma(n/m) + \sum_{\alpha} \wp_{\alpha}\sigma(m)$$

$$= \sigma(n) + \sum_{i} \wp_i\sigma(1),$$

(1.22)

which has solution $\sigma(n) = k_{\text{B}} \ln n$. For the case of non-uniform macrostates, $\wp_{\alpha} = m_{\alpha}/n$, and uniform microstates $\wp_i^{\text{u}} = 1/n$, then $S[\wp^{\text{u}}(i \mid \alpha)] = \sigma(m_{\alpha})$, and

$$S_{\text{total}} = \sigma(n)$$

$$= S[\wp_{\alpha}] + \sum_{\alpha} \wp_{\alpha}\sigma(m_{\alpha}),$$

(1.23)

or

$$S[\wp_{\alpha}] = k_{\text{B}} \ln n - \sum_{\alpha} \wp_{\alpha}k_{\text{B}} \ln m_{\alpha}$$

$$= -k_{\text{B}} \sum_{\alpha} \wp_{\alpha} \ln \wp_{\alpha}.$$

(1.24)

This shows that the $\wp \ln \wp$ term is just part of the total entropy, and that the full expression is

$$S_{\text{total}} = S[\wp_{\alpha}] + \sum_{\alpha} \wp_{\alpha}S_{\alpha}$$

$$= \sum_{\alpha} \wp_{\alpha}[S_{\alpha} - k_{\text{B}} \ln \wp_{\alpha}],$$

(1.25)

which is of course equation (1.19).

In describing the Gibbs–Shannon expression equation (1.1) as either incorrect, or else, in the uniform probability case, as correct but pointless, one should concede that this formula has long been successfully used in communications and information theory. Perhaps the problem is one of nomenclature. If Shannon had just called *his H*-function the information, as he originally introduced it (Shannon 1948, section 6), rather than calling it the entropy based on its resemblance to a similar function used in statistical mechanics, then it could have been judged solely on its utility for informatics, rather than by whether or not it represented the total entropy of a physical system.

As a concrete example, consider a fair six-sided die. In the first case let the number showing be reported, which means that there are six states each with probability $\wp_i = 1/6$ and internal entropy $S_i = 0$. In this case the Gibbs–Shannon information entropy is the same as the total entropy of the system,

$$S_{\text{GS}} = S_{\text{total}} = k_{\text{B}} \ln 6.$$

(1.26)

In the second case, suppose that the state of the die is reported as 1, 3, 5 or even. The first three states each have probability $\wp_i = 1/6$ and entropy $S_i = 0$, and the final state has probability $\wp_4 = 1/2$ and entropy $S_4 = k_{\text{B}} \ln 3$. The Gibbs–Shannon information entropy is in this case

$$S_{\mathrm{GS}} = k_{\mathrm{B}} \left\{ \frac{3}{6} \ln 6 + \frac{1}{2} \ln 2 \right\} = k_{\mathrm{B}} \ln \sqrt{12} \approx k_{\mathrm{B}} \ln 3.5,$$

which is less than in the first case. The total entropy is

$$S_{\mathrm{total}} = S_{\mathrm{GS}} + k_{\mathrm{B}} \left\{ \frac{3}{6} \ln 1 + \frac{1}{2} \ln 3 \right\} = k_{\mathrm{B}} \ln 6,$$

which is the same as in the first case. Arguably, information has been lost in the second case because the even numbers are not individually identified, and so the Gibbs–Shannon information entropy may be regarded as a realistic quantification of the information in the two systems. Similarly, because the two systems are physically identical, the total entropy must be the same in each, which it is. This example suggests that the Gibbs–Shannon formula should be reserved for informatic applications, and that when applied to physical systems the additional contribution required for the total entropy should be added.

There has been a deal of discussion over the years as to whether or not the Gibbs–Shannon information entropy is in any way related to the thermodynamic or physical entropy. My attitude is that on the one hand in the physical sciences it is incorrect to imply that the Gibbs–Shannon information entropy is 'the' entropy of a physical system. Clearly it is only part of the total entropy, as the above analysis, equation (1.19), and the results in figure 1.3 show. On the other hand it *is* part of the total entropy of a physical system, and it offers a relatively straightforward and quickly converging way to calculate that part (see section 5.6). This part can be difficult to obtain by other methods, whereas the remaining part, the internal entropy of the macrostate, which often turns out to be the reservoir entropy of a physical system, is not difficult to obtain. The latter can be simply added to the Gibbs–Shannon information entropy to get the total entropy. Hence, provided that one is explicit that the Gibbs–Shannon information entropy is only part of the total entropy, it can be quite useful in thermodynamics and statistical mechanics.

1.1.4 Continuum

The above analysis was implicitly for discrete states. In many cases it is worthwhile taking the continuum limit. It is quite straightforward to transform the sums to integrals in the usual fashion. Alternatively, one can derive the results for the continuum directly, as is now done.

Let the state of the system be $\mathbf{x}$, a point in a multidimensional space, and let $\omega(\mathbf{x})$ be the weight density, which is non-negative. (The quantity $\mathbf{x}$ may be regarded as the continuum analogue of microstates, although similar procedures also hold for continuum macrostates.) The total weight of the system is

$$W = \int \mathrm{d}\mathbf{x}\, \omega(\mathbf{x}).$$

The total entropy is just the logarithm of this $S = k_{\mathrm{B}} \ln W$.

The probability density is proportional to the weight density

$$\wp(\mathbf{x}) = \frac{\omega(\mathbf{x})}{W}.$$

This is obviously normalized, $\int d\mathbf{x}\,\wp(\mathbf{x}) = 1$. The probability density has the interpretation that $\wp(\mathbf{x})d\mathbf{x}$ is the probability of the system being within $d\mathbf{x}$ of $\mathbf{x}$.

The average of a function of the state of the system is

$$\langle f \rangle = \int d\mathbf{x}\,\wp(\mathbf{x})f(\mathbf{x}).$$

(1.3
1)

One can introduce an arbitrary volume element $\Delta(\mathbf{x})$, and define the entropy of the state of the system in terms of it as

$$S(\mathbf{x}) = k_B\,\ln\left[\omega(\mathbf{x})\Delta(\mathbf{x})\right].$$

(1.3
2)

With this the probability density is

$$\wp(\mathbf{x}) = \frac{e^{S(\mathbf{x})/k_B}}{\Delta(\mathbf{x})W}.$$

(1.3
3)

The volume element $\Delta(\mathbf{x})$ is solely a matter of convenience that makes the argument of the logarithm dimensionless and gives the probability density the correct dimensions. Obviously the probability density is independent of the choice of $\Delta(\mathbf{x})$, since the one implicit in the state entropy cancels with the one that appears explicitly in the probability density. Since the volume element has no physical consequences, it can be taken to be a constant or it could be ignored altogether.

The total entropy of the system may be written as a functional of the probability density,

$$\begin{aligned}
S &= k_B\,\ln W \\
&= \int d\mathbf{x}\,\wp(\mathbf{x})k_B\,\ln W \\
&= \int d\mathbf{x}\,\wp(\mathbf{x})\left[S(\mathbf{x}) - k_B\,\ln\frac{e^{S(\mathbf{x})/k_B}}{W}\right] \\
&= \int d\mathbf{x}\,\wp(\mathbf{x})[S(\mathbf{x}) - k_B\,\ln\{\wp(\mathbf{x})\Delta(\mathbf{x})\}].
\end{aligned}$$

(1.3
4)

This is the continuum analogue of equation (1.19). Again the arbitrary volume element that appears explicitly here, cancels with that implicit in $S(\mathbf{x})$, and the total entropy is independent of the choice of $\Delta(\mathbf{x})$.

In the case of the continuum, the macrostates are represented as hypersurfaces in the space. These hypersurfaces correspond to constant values of particular observables. In general the observables depend upon the state of the system $\mathbf{x}$.

1.2 A combinatorial example

In this section a simple combinatorial example is given that illustrates the use of entropy. The analysis follows Attard (2002, section 1.2).

1.2.1 Entropy

Suppose that the system consists of a container with N cells and that there are currently $n \leqslant N$ objects in the container, with at most one object per cell. A microstate of the system is a specification of which cells are occupied. There are $W_N(n) = {}^NC_n \equiv N!/(N-n)!n!$ possible microstates. The entropy of the system with this number of cells and objects is therefore

$$S_N(n) = k_\text{B} \ln W_N(n) = k_\text{B} \ln \frac{N!}{(N-n)!n!}.$$

(1.35)

Now add a second container of M cells containing m objects to the system. The number of microstates for this container is $W_M(m) = {}^MC_m$, and the entropy for this second container alone is $S_M(m) = k_\text{B} \ln W_M(m)$. The number of microstates of the total system with the fixed allocation of objects in the two container is

$$W_{N,M}(n,m) = W_N(n)W_M(m).$$

(1.36)

It is the product that appears because for each configuration of one container all configurations of the other container are possible. The entropy of the total system with the fixed allocation is obviously

$$\begin{aligned} S_{N,M}(n,m) &= k_\text{B} \ln W_{N,M}(n,m) \\ &= k_\text{B} \ln W_N(n) + k_\text{B} \ln W_M(m) \\ &= S_N(n) + S_M(m). \end{aligned}$$

(1.37)

This is an explicit example of the linear additive nature of entropy.

1.2.2 Equilibrium allocation

Now introduce a transition rule that allows exchange of objects between the two containers. Choose a cell at random in the first container and a cell at random in the second container and swap their contents. (One can use similar transition rules within each container to change their internal configurations, but this is not necessary for what follows.) Obviously if both cells are occupied or if both cells are empty nothing happens. But if the first cell is occupied and the second cell is empty, then an object is transferred from the first container to the second: $n \Rightarrow n-1$ and $m \Rightarrow m+1$. The total number of objects $n_\text{tot} = n(t) + m(t)$ is conserved by the transfer rule.

Intuitively, one expects that on average objects will move from the container with the greater concentration to that with the lesser. It is concentration of particles rather than number of particles that is relevant because simply doubling the size of the container and doubling the number of particles does not change the probability of a net transfer. Eventually a steady state will be reached where there is no net transfer, and the concentrations are the same in the two containers. (Concentration is an example of an intensive variable, whereas number is an extensive variable. It is quite common in thermodynamics that the optimum state is found by equalizing intensive variables.) The steady state will treat objects and empty cells equally, since the transfer of an object in one direction is equivalent to the transfer of an empty cell in the opposite direction.

In order to makes this intuition quantitative, one needs the expression for the transfer probability. The probability of a cell in the first container being occupied is just n/N, and the probability of it being empty is of course $(N-n)/N$. Analogous expressions hold for the

second container. The probability for an object to be transferred from the first container to the second is therefore

$$\wp(1 \to 2) \equiv \wp(n-1 \mid n) = \frac{n}{N} \frac{M-m}{M}.$$

(1.38)

As the product of two independent normalized probabilities, this is correctly normalized. Obviously the probability of an object going from the second container to the first is

$$\wp(2 \to 1) \equiv \wp(n+1 \mid n) = \frac{N-n}{N} \frac{m}{M}.$$

(1.39)

The probability of no nett exchange is just

$$\wp(n \mid n) = \frac{n}{N} \frac{m}{M} + \frac{N-n}{N} \frac{M-m}{M}.$$

(1.40)

The steady state of no nett transfer is denoted with an over-line, $n = \bar{n}$ and $m = \bar{m}$. At this point the transfer probabilities must be equal, $\wp(\bar{n}-1 \mid \bar{n}) = \wp(\bar{n}+1 \mid \bar{n})$, which is to say

$$\frac{\bar{n}M - \bar{n}\,\bar{m}}{NM} = \frac{\bar{m}N - \bar{n}\,\bar{m}}{NM}, \quad \text{or} \quad \frac{\bar{n}}{N} = \frac{\bar{m}}{M}.$$

(1.41)

As was intuited above, there is no nett transfer on average when the proportion of occupied cells is the same in both containers. Clearly one has the same result for empty cells, $(N - \bar{n})/N = (M - \bar{m})/M$.

1.2.3 Maximum entropy

It is now shown that the entropy of the total system is a maximum in this steady state. Hence it is reasonable to call it the equilibrium state. It will be shown below that the transition rule drives the system to this state, and that it is stable to small fluctuations.

The total entropy for the system constrained to be in the equilibrium state is

$$S_{N,M}(\bar{n}, \bar{m}) = k_{\mathrm{B}} \ln \frac{N!}{(N-\bar{n})!\bar{n}!} + k_{\mathrm{B}} \ln \frac{M!}{(M-\bar{m})!\bar{m}!}.$$

(1.42)

This is the logarithm of the number of distinct arrangements of objects in each container for the equilibrium allocation.

Now consider a state on one or other side of the equilibrium state, $n = \bar{n} + p$ and $m = \bar{m} - p$, and consider a neighboring state, $n\prime = n + 1$, and $m\prime = m - 1$. If $p > 0$ then there are more than the equilibrium number of objects in the first container, $n > \bar{n}$, and if $p < 0$ then there are less than the equilibrium number of objects in the first container, $n < \bar{n}$. The difference between their entropies is

$$S_{N,M}(n\prime,m\prime) - S_{N,M}(n,m) = k_{\mathrm{B}}\,\ln\left\{\frac{(N-n)!n!}{(N-n-1)!(n+1)!}\times\frac{(M-m)!m!}{(M-m+1)!(m-1)!}\right.$$

$$= k_{\mathrm{B}}\,\ln\frac{N-n}{n+1}\,\frac{m}{M-m+1}$$

$$= k_{\mathrm{B}}\,\ln\frac{\alpha-\dfrac{p}{M}}{\alpha+\dfrac{p+1}{N}}\,\frac{1-\alpha-\dfrac{p}{N}}{1-\alpha+\dfrac{1+p}{M}},$$

(1.4 3)

where $\alpha \equiv \bar{n}/N = \bar{m}/M$. If $p > 0$ ($n\prime = n+1$ is further from $\bar{n}$ than n), both fractions are less than unity, the logarithm is negative, and the entropy difference is negative. Hence the entropy decreases moving further away from equilibrium and increases toward equilibrium. If $p < 0$ ($n\prime = n+1$ is closer to $\bar{n}$ than n), both fractions are greater than unity, the logarithm is positive, and the entropy difference is positive. Hence the entropy increases toward equilibrium and decreases moving further away from equilibrium. One concludes that the total entropy is a concave down function that attains its maximum at equilibrium $\bar{n}$.

One has to distinguish between the constrained entropy and the unconstrained entropy. If objects are transferable between the two containers, then the total number of distinct configurations is $W_{N+M}(n+m) = {}^{N+M}C_{n+m}$, since all cells are now accessible to all objects. The total entropy is the usual logarithm of this $S_{N+M}(n+m) = k_{\mathrm{B}}\,\ln W_{N+M}(n+m)$. This is the unconstrained entropy, where n and m are not individually fixed because the containers are effectively combined.

The total number of configurations in this unconstrained case must be greater than any of those of the isolated containers with a fixed allocation of the objects. The latter case is the constrained entropy case. The reason that the unconstrained entropy must be greater than the constrained entropy is that it includes all the configurations of the latter, plus all the configurations with a different allocation of the objects. That is, $W_{N+M}(n+m) \geqslant W_N(n)W_M(m)$ or in terms of the entropy, $S_{N+M}(n+m) \geqslant S_N(n) + S_M(m)$, for any n, m. (This is a strict inequality unless N or M is 0, or unless there are no objects, $n+m = 0$, or no empty cells, $n+m = N+M$.) One concludes that the entropy of the two containers able to exchange objects is greater than the entropy of the two isolated containers with the equilibrium allocation of objects, which is greater than the entropy of any other allocation of objects to the isolated containers,

$$S_{N+M}(n+m) \geqslant S_N(\bar{n}) + S_M(\bar{m}) \geqslant S_N(n) + S_M(m),$$

(1.4 4)

where $\bar{n} + \bar{m} = n+m$, and $\bar{n}/N = \bar{m}/M$. This behavior of the constrained entropy is shown in figure 1.4.
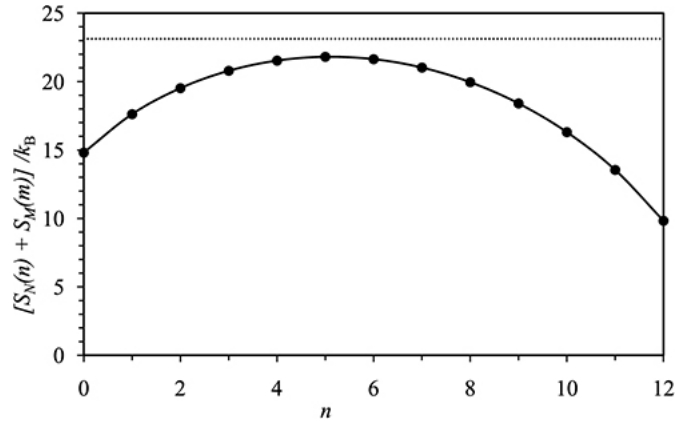
Figure 1.4. The constrained total entropy for two containers, $N = 18$ and $M = 24$, with a total of $n + m = 12$ particles, as a function of n. The maximum $S_N(\bar{n}) + S_M(\bar{m}) = 21.8 k_{\mathrm{B}}$ occurs at $\bar{n} = 5$. The dotted line is the unconstrained total entropy, $S_{N+M}(n + m) = 23.1 k_{\mathrm{B}}$.

Note that it is the constrained entropy that is linear additive.

1.2.4 Stability and the direction of motion

We now prove that the equilibrium state is stable, which is to say that if $n = \bar{n} + p$, $m = \bar{m} - p$, then $\wp(1 \to 2) > \wp(2 \to 1)$ if $p > 0$, and vice versa. For arbitrary p we have

$$
\begin{aligned}
\frac{\wp(1 \to 2)}{\wp(2 \to 1)} &= \frac{n(M - m)}{(N - n)m} \\
&= \frac{(\bar{n} + p)(M - \bar{m} + p)}{(N - \bar{n} - p)(\bar{m} - p)} \\
&= \frac{\left(\alpha + \dfrac{p}{N}\right)}{\left(\alpha - \dfrac{p}{M}\right)} \times \frac{\left(1 - \alpha + \dfrac{p}{m}\right)}{\left(1 - \alpha - \dfrac{p}{N}\right)},
\end{aligned}
$$

(1.45)

where $\alpha \equiv \bar{n}/N = \bar{m}/M$. Since the number of occupied cells cannot be negative, one must have $|\, p\,| \leqslant \bar{n}$, and $|\, p\,| \leqslant \bar{m}$, and since the number of empty cells cannot be negative one must have $|\, p\,| \leqslant N - \bar{n}$, and $|\, p\,| \leqslant M - \bar{m}$. Hence, by inspection, if $p > 0$, then the two fractions are both greater than unity, and if $p < 0$, then the two fractions are both less than unity. Hence one can conclude that

$$
\frac{\wp(1 \to 2)}{\wp(2 \to 1)} = \begin{cases} > 1, & \text{if } n > n, \\ < 1, & \text{if } n < n. \end{cases}
$$

(1.46)

These say that if $n > \bar{n}$, then there is more likely to be a decrease in n than an increase, and vice versa. This shows that the equilibrium state is a stable state, which is to

say that fluctuations in n from $\bar{n}$ are likely to be countermanded by the subsequent transitions.

This proof of the stability of the equilibrium state shows that if the system is displaced from the equilibrium state, either by spontaneous fluctuations, initial preparation, or forces of constraint, then there is a nett driving force toward the equilibrium state. The most likely flux of particles is in the direction of increasing constrained entropy. Although the velocity has not been discussed, on symmetry grounds (vectors are proportional to vectors) one might anticipate that it is proportional to the gradient in entropy.

Finally, there is nothing in this model that precludes a spontaneous fluctuation from the equilibrium state. During such a fluctuation the constrained entropy decreases. This shows that the Second Law of Thermodynamics should not be taken overly literally. For a macroscopic system fluctuations are relatively small and can be neglected. For a macroscopic system prepared in a state away from equilibrium, with overwhelming probability it will evolve in the direction of increasing entropy.

1.2.5 Physical interpretation

The interpretation of these results for this simple combinatorial model gives insight into the behavior of thermodynamic and statistical mechanical systems more generally. One fundamental point is the distinction between microstates and macrostates. In the present example a microstate is a configuration of the objects, which is to say the list of occupied cells. In contrast, a macrostate is the number of occupied cells in a container, irrespective of the specific cells occupied. Obviously each macrostate contains many microstates. A specified macrostate constrains the two containers to hold the given number of objects. If no macrostate is specified, there is no such constraint, objects are free to transfer between the two containers, and the number of occupied cells in each container ranges over all possible values. This means that the number of configurations for the constrained system (i.e. specified macrostate) is less than the number of configurations without constraint (i.e. no specified macrostate). It follows that the entropy of a system constrained to be in a given macrostate (constrained or macrostate entropy) is less than the entropy of an unconstrained system (unconstrained or total entropy).

If the system is initially set up in a macrostate other than the equilibrium one, then there is likely a nett flux of objects toward the equilibrium allocation. This is not driven by any increase in the unconstrained or total entropy because once the total number of objects is fixed and allowed to transfer between the containers the number of possible configurations (microstates) does not change. Instead, the evolution toward equilibrium is a macroscopic flux, with each successive macrostate likely having greater entropy than the preceding one. That is, it is the entropy constrained by the current value of the quantity in flux that increases during the approach to equilibrium. It is the constrained entropy that is a maximum in the equilibrium macrostate.

It ought to be clear that the progress toward equilibrium is a statistical one that holds on average, or in a most likely sense. It is entirely possible for a temporary reversal of the macrostate flux to occur due to the random nature of the microstate transitions. Likewise, once the equilibrium macrostate is attained, there can be temporary fluctuations to nearby macrostates, which by definition have lower entropy.

The above mathematical analysis shows the increase in the constrained entropy during the approach to the equilibrium state. The physical reason for this increase does not lie in the microstates, because the total number of unconstrained configurations is fixed (once the total number of particles and the freedom to exchange is specified) and hence so is the total unconstrained entropy. The microstates do not drive the approach to equilibrium. Transitions between microstates scramble the configurations, but each such microstate transition is as likely as its reverse. There is no flux and no equilibration at the level of microstates.

It is the flux in macrostates that equilibrates the system. The essential point is that the number of microstates differs between macrostates. Because of this the transition between macrostates is asymmetric. If the entropy increases from macrostate 1 to 2 to 3 (i.e. the number of microstates in each macrostate increases in this sequence), then a system in a microstate belonging to macrostate 2 is more likely to make a transition to a microstate in 3 than in 1 simply because there are more target microstates in 3 than in 1.

This point is clear in the present combinatorial example where the transitions are purely stochastic, and the number of possible target microstates for a given microstate is proportional to the size of the target macrostate. An example of this is illustrated in figure 1.5. Although this model is highly idealized, the broader point also holds for the statistical contribution to more physical models with realistic dynamics.

Figure 1.5. All microstate transitions that change the macrostate for a system of two containers, $N = M = 4$ with $n + m = 4$ particles. The initial microstate lies in the macrostate $n = 1$, $m = 3$ (left). There is one transition that decreases n (upper right), and nine transitions that increase n (lower right). The constrained entropy of the initial macrostate is $S_{4,4}(1,3) = k_B \ln 16$, that of the n-decreasing final macrostate is $S_{4,4}(0,4) = 0$, and that of the n-increasing final macrostate is $S_{4,4}(2,2) = k_B \ln 36$, which in fact is the maximum constrained entropy.

The asymmetry in the macrostate transition probability, $\wp(2 \to 3) > \wp(2 \to 1)$ if $S_3 > S_1$, creates the irreversibility in the progress toward equilibrium. Of course the opposite transition may occur at any instant, but microstate transitions that increase macrostate entropy are more probable than the reverse, and in the long run these will occur most frequently.

The equilibrium macrostate is the one with the most microstates. Although fluctuations to nearby macrostates occur, which necessarily decrease the constrained entropy of the system, usually the constrained entropy is such a sharply peaked function of the macrostates that the effects of such fluctuations are relatively negligible.

1.3 Entropy and the Second Law

Let us return to the Second Law as enunciated by Clausius (1850),

> The entropy increases during spontaneous changes in the state of asystem.

Compare this to Boltzmann's (1895, Brush 1964), identification of the physical basis of entropy,

$$\text{The entropy of a state is the logarithm of the number}$$
$$\text{of molecular configurations in the state.} \tag{1.48}$$

The general definition of entropy given earlier in this chapter is slightly broader than Boltzmann's, since it uses weight rather than number, but it reduces to Boltzmann's in the case that the molecular configurations have equal weight, which they do in the phase space of an isolated system, but not more generally.

Boltzmann's insight showed the equivalence of entropy and probability, since for equally likely molecular configurations, the probability of a state is proportional to the number of configurations that it contains. This equivalence was made extant in the general definition of entropy in terms of weight given earlier in the chapter. The relationship of probability to entropy introduces a stochastic element into thermodynamics, with the Second Law now interpreted to mean that

$$\text{Spontaneous transitions are more likely to occur}$$
$$\text{from less probable to more probable states.} \tag{1.49}$$

This does not have quite the zing of Clausius' version, but it does have greater precision. For macroscopic systems the transition probability approaches certainty. This version of the Second Law was implicitly assumed above in the discussion of the macrostate transition probability, and it is intuitively obvious. But whether or not Boltzmann has indeed provided an explanation and a proof of the Second Law of Thermodynamics requires greater scrutiny.

The probabilistic nature of the Second Law of Thermodynamics can be readily seen. An equilibrium system does not change macroscopically in time, but due to molecular motion, fluctuations about the equilibrium state must occur. It will be seen that the relative magnitude of these fluctuation in macroscopic variables scales inversely with the square root of the system size, and so they are generally too small to measure experimentally. Any spontaneous fluctuation away from the equilibrium state must be to a state of lower entropy, which disproves any literal interpretation of Clausius' version of the Second Law. From this one can conclude that the Second Law of Thermodynamics in its literal form should be applied to systems that have been prepared in an initial state far from equilibrium, which is to say a low entropy, spontaneously improbable state. In this case the state of the system will spontaneously evolve toward the equilibrium, maximum entropy, most probable state, and the probability of such evolution approaches certainty for a macroscopic system.

Boltzmann's definition of entropy, equation (1.48), applies to the probability of a state, whereas the Second Law, especially in its more precise form, equation (1.49), applies to the probability of transitions between states. Obviously these are different things. Equilibrium theory depends on Boltzmann's identification of entropy and probability for a state, whereas non-equilibrium theory depends on the nature of transition probabilities. It is slightly perverse that Clausius' Second Law is always regarded as a law of equilibrium thermodynamics, whereas a more pedantic interpretation would be that it applies to not-in-equilibrium systems. In any case in what fashion the transition probabilities are determined by the Second Law is now discussed.

The challenge is to make more precise the intuitively appealing notion that transitions to a more probable state should be more probable then transitions to a less probable state. One would like to deduce something about the transition probability from the state probability and relate this to the spontaneous evolution of entropy and the Second Law.

The unconditional probability of observing the transition from the state α to the state $\alpha\prime$ in the time interval $\tau > 0$ may be denoted $\wp(\alpha\prime, \alpha \mid \tau)$. The probability of the reverse transition is $\wp(\alpha, \alpha\prime \mid \tau)$. Since there is no preferred direction of time in an equilibrium system, one is just as likely to observe the forward transition as the reverse, $\wp(\alpha\prime, \alpha \mid \tau) = \wp(\alpha, \alpha\prime \mid \tau)$. (For simplicity, at this introductory level we assume that the states are even functions of the molecular velocities.) According to Bayes' theorem, which was given above as one of the laws of probability, the unconditional and conditional probabilities are related as $\wp(\alpha\prime, \alpha \mid \tau) = \wp(\alpha\prime \mid \alpha, \tau)\wp(\alpha)$. Here, $\wp(\alpha\prime \mid \alpha, \tau)$ is the probability of the system being in the state $\alpha\prime$ given that it was in the state α at time τ earlier, and $\wp(\alpha)$ is the probability that the system is in the state α. For an equilibrium system the latter does not depend upon time.

For a large time interval, any two states are uncorrelated. In this limit the unconditional probability reduces to the product of the singlet probabilities, $\wp(\alpha\prime, \alpha \mid \tau) \rightarrow \wp(\alpha\prime)\wp(\alpha), \ \tau \rightarrow \infty$. Also, the conditional probability is independent of the initial state, $\wp(\alpha\prime \mid \alpha, \tau) \rightarrow \wp(\alpha\prime), \ \tau \rightarrow \infty$. If one now introduces a third state $\alpha\prime\prime$, then in this limit

$$\frac{\wp(\alpha\prime \mid \alpha, \tau)}{\wp(\alpha\prime\prime \mid \alpha, \tau)} \ \rightarrow \ \frac{\wp(\alpha\prime)}{\wp(\alpha\prime\prime)}, \ \tau \rightarrow \infty.$$

$$(1.50)$$

If the state $\alpha\prime$ is more probable than the state $\alpha\prime\prime$ (Boltzmann would say it has more entropy), then this says that a system in the state α is in the long term more likely to make the transition to $\alpha\prime$ than to the state $\alpha\prime\prime$. In other words transitions to states of higher entropy are more likely to be observed than the reverse. In this sense Boltzmann's physical interpretation of entropy explains Clausius' Second Law of Thermodynamics.

Although illuminating, this argument only applies in the long-time limit, and so it cannot be the whole story. On shorter time intervals correlations between the initial and destination states can be expected. Although not rigorous, one can make a still informative argument for finite time intervals. Let $S(x)$ be the entropy of the state x. We are interested in the microstate transitions, and let $n_{\pm}(x)$ denote the number that increase or decrease the value of x. Now for an equilibrium system, the number of forward transitions must equal the number of reverse transitions, $n_+(x) = n_-(x)$. Let us make the intuitively appealing assumption that the number of such molecular transitions is proportional to the number of molecular configurations, $n_{\pm}(x) \propto \exp S(x)/k_{\mathrm{B}}$. The interpretation of this assumption is that larger states (i.e. those with the greater number of molecular configurations) have a larger number of transitions. Now consider that a state has finite width $\Delta_x > 0$. In this case the excess number of forward transitions is $n_+(x + \Delta_x/2) - n_-(x - \Delta_x/2) \propto \Delta_x \mathrm{d}S(x)/\mathrm{d}x$. One sees that if the states are ordered in terms of increasing entropy, (i.e. $\mathrm{d}S(x)/\mathrm{d}x \geqslant 0$), then there are more forward transitions in the direction of increasing entropy than backward transitions. This is the essence of the Second Law of Thermodynamics.

This argument provides a second link between Boltzmann's interpretation of entropy and Clausius' Second Law of Thermodynamics. The essential points are that the number of transitions between configurations is proportional to the number of configurations themselves. It is also essential that the transitions are steps of finite width, which is to say that this does not apply for infinitesimal transitions. It follows from this second argument that the driving force for the transitions is the gradient in the entropy, as was asserted above in the combinatorial example, section 1.2.4. A more sophisticated version of these arguments is given in section 3.1.5.

1.4 Nature of probability and randomness

Since entropy is in essence the same as probability (because there is a one-to-one relationship between the two), it is worthwhile discussing the conceptual basis of probability in the physical sciences. Such a fundamental understanding of probability determines not just the mathematical formulation of probability theory, but also the quantitative values of the weights that are to be applied to the microstates that occur in statistical mechanics.

Probability is also intimately connected to the notion of randomness. Of course classical statistical mechanics, with which the early parts of this book are concerned, is based on the classical equations of motion (Newton's, or equivalently, Hamilton's). Since these are deterministic, there is an issue as to how randomness arises in the physical sciences.

1.4.1 Probability

There are three philosophical positions that have been adopted for the nature of probability. The first is frequency, which holds that the probability of a state (or an event) is the frequency with which the state occurs in a long sequence of repeated trials. The second is credibility or degree of reasonable belief. This is also known as the subjective interpretation of probability, and it holds that the probability of a state is the strength with which the observer believes that the state will occur. The third is measure (or propensity), which holds that the probability of a state is the weight of the set of conditions that give rise to the state.

Of these three philosophical positions, the present author rejects the first two and embraces the third for the following reasons.

1.4.1.1 Frequency

To say that frequency is probability is to confuse cause and effect. The frequency with which an event or state occurs is a direct consequence of the probability of the event or state. In some cases where it is possible to make a sequence of trials, it is possible to measure probability by the frequency. But in the physical sciences we would say that the probability of the state existed whether or not the trials were carried out, whereas the frequency only exists after the trials are performed. Further, once-only events, such as the winner of a particular horse race, or the weather on a particular day, have a probability associated with them prior to the event, even though it is physically impossible to repeat these in a sequence of trials.

Of course some might say that frequency is just a mental device that allows probability to be explained to the uninitiated, or that allows it to be concretely visualized. But the problem with this defence is that it replaces the real physical origins of probability with an imaginary mental artifice. What artificial properties are created by this picture, and what real properties are overlooked?

This discussion of whether probability causes frequency, or frequency causes probability is not an empty academic exercise carried out solely for the sake of the argument. It does have at least one important consequence in the practice and formulation of statistical mechanics.

The ensemble picture of statistical mechanics is very old, having been used by both Boltzmann and Gibbs. In fact it is so widely used that many workers would be unaware that there is any alternative way of formulating statistical mechanics. The ensemble formulation asserts that the system is but one member of an ensemble of macroscopically identical systems, and that the probability of a particular state of the system is equal to the proportion of systems in the ensemble that are in that state. This ensemble interpretation is applied to classical statistical mechanics, as well as to the density matrix of quantum statistical mechanics. The evolution of the members of the ensemble in time is said to correspond to the evolution of the probability. In so far as the proportion of an ensemble in

a given state is the probability of the state, the ensemble picture of statistical mechanics is obviously just a form of the frequency interpretation of probability.

There are two main objections to the ensemble picture. First, the experimental reality is that measurements are performed on a single system, not an ensemble of systems. (Repeat measurements are not the same as an ensemble: they are not essential, and there are not Avogadros's number of them.) Either the probability theory that underlies statistical mechanics works for a single system, or else the theory is a failure. If probability theory works for a single system, then ensembles are as superfluous as the aether.

Second, the members of the ensemble are implicitly conserved. This implies a conservation law for the probability during its evolution via the molecular equations of motion. Conservation laws are very serious laws of nature, and they should not be postulated without serious reason. That the members of an ensemble are conserved in someone's imagination does not make it true that probability is conserved in a real physical system. In fact it can be shown that for certain non-equilibrium systems weight is not conserved. This is not something that can be deduced from the ensemble picture. However, with hindsight, because ensembles are purely imaginary, any law, real or imaginary, can be added to ensembles to make them behave as they ought to behave. The invocation of arbitrary laws to make ensemble theory fit nature is neither convincing nor is it in accord with the way that science ought to be done.

1.4.1.2 Credibility

The credibility or subjectivist notion of probability holds that it is a measure of the strength of a person's belief that the state will occur. The beliefs must be rational and in accord with the rules of inductive reasoning, which happen to be the same as the rules of probability. However, two people with different information may attribute different probabilities to a particular state. Jaynes (2003) was a particularly strong advocate of the credibility interpretation of probability.

The problem with the subjectivists view is that it divorces probability from the underlying physical causes. For example, in Jaynes' treatment of the canonical probability distribution in statistical mechanics, the temperature is nothing more than a Lagrange multiplier that reflects the observer's ignorance of the energy of the system. This contrasts with the interpretation of most thermodynamicists, who view temperature as a physical attribute independent of what an observer knows or does not know. Similarly, Jaynes views entropy as a measure of the observer's ignorance, whereas I view entropy as the logarithm of the weight of relevant microstates. To most scientists who believe that physical phenomena have an objective reality, the subjectivist view is peculiarly solipsistic.

The credibility interpretation of probability provides the conceptual basis for the maxent theory for obtaining probability distributions, again developed by Jaynes. This has been discussed above in connection with equations (1.1) and (1.19), where it was criticized as being unsound.

1.4.1.3 Measure

In this book an objective interpretation of probability is used, rather similar to the propensity interpretation of Popper (1959). Probability is taken to be a physical property of the system, namely it is proportional to the measure or weight of microstates in the specified macrostate. In the event that the microstates have equal weight, the weight of the macrostate is just the number of microstates in it.

Accepting the physical nature of probability, it remains to give a prescription for determining the microstate weights in a given system. In the simplest case the microstates are discrete and identical apart from their label, as for example, are the six faces of a true die. In such cases the microstates have equal weight, which, without loss of generality, can be assigned unit value so that the weight of a macrostate is simply the number of microstates that it contains.

One should not assume without reason that discrete microstates are equally weighted in a particular case. Similarly, in the case that the microstates form a continuum, the weight density may be uniform or non-uniform, depending on the physical characteristics of the system.

One of the most important cases, and the one upon which is focussed most attention in this book, is when the microstates represent the positions and momenta of all the particles of the system. This is called phase space. In the fundamental case when the system is isolated from the rest of the Universe, the weight density of the total phase space is uniform. This can be established (see chapter 5) by equating phase space averages (i.e. the average of a phase function over phase space using the weight density) to time averages (i.e. the simple average of the values of the phase function at discrete time intervals). This takes it as axiomatic that time is homogeneous and has uniform weight.

Having established the uniformity of classical phase space of the total isolated system, it is straightforward to derive the weight density for that phase space of the particles contained in a sub-system of the total isolated system, which, it turns out, is not uniform. Such sub-systems are the most common application of statistical mechanics.

In the case of quantum statistical mechanics, one can take a similar approach to establish the weight of the discrete quantum states that form the microstates of the system, as is detailed in chapter 7.

1.4.2 Randomness and irreversibility

As mentioned at the beginning of this section, there is a close connection between probability and randomness: if the system were fully determined, one would not need probability and a statistical description. In classical mechanics, the motion of the particles and the consequent evolution of the system is fully deterministic. Hence there is an issue to be addressed regarding how randomness arises from such deterministic equations of motion.

A related problem is that the classical equations of motion are time-reversible: if the velocity of all the particles in the system is reversed, then continuing forward in time the system will exactly retrace its prior history. This is called microscopic reversibility (see sections 5.1.2 and 5.4.2). This means that if the entropy had been increasing, then upon particle velocity reversal it would start decreasing. This reversibility in the equations of motion apparently contradicts the time asymmetry mandated by the Second Law of Thermodynamics, namely that a system evolves only in the direction of increasing entropy.

It turns out that the resolution of these apparent paradoxes revolves around the very formulation of statistical mechanics and thermodynamics. But first, three alternative proposals must be dismissed.

1.4.2.1 Quantum questions?

Some try to simply dismiss these issues as an artefact of classical mechanics that would disappear in a proper quantum treatment. In particular, some argue that quantum mechanics is inherently uncertain and unpredictable, and that this is sufficient to resolve the issue. In fact this does not solve the problem for three reasons. First, in quantum mechanics, the wave function not only evolves in a fully deterministic fashion via Schrödinger's equation, but also the complex conjugate of the wave function evolves in an exactly time-reversed fashion, both of which are analogous to the classical case. Second, most atoms and molecules behave classically due to their size; in practice for most terrestrial condensed matter systems quantum corrections contribute less than 1% for any atom larger than argon. The problems of indeterminacy and irreversibility holds for such systems and the classical equations have to be able to account for them. Third, classical statistical mechanics and thermodynamics exist as disciplines independent of quantum mechanics, as of course is clear from the fact that they were developed more than 50

years before the discovery of quantum theory. Therefore, again one must conclude that quantum mechanics cannot be essential to resolving these two paradoxes.

Of course, since quantum mechanics is the fundamental law of Nature, both classical mechanics and classical statistical mechanics ought to be derived from it. This derivation is carried out in chapter 7, where the physical picture that gives rise to the classical equations turns out to be the same as that discussed here as the origin of randomness and irreversibility.

1.4.2.2 Collective chaos?

A second explanation proposed by some is that of ensembles, which have already been dealt with as an explanation of probability itself, section 1.4.1. Introduced by Boltzmann, an ensemble consists of identical macroscopic copies that differ in their initial microscopic or molecular configuration. The ensemble as a whole evolves according to the deterministic equations of motion of each isolated member. The randomness in this picture arises from the assigned distribution of the initial microstates of the members of the ensemble.

The ensemble picture is today more or less the standard view of probability and randomness in thermodynamics and statistical mechanics, but there are several weaknesses in it, as have already been outlined. First, ensembles are imaginary. They are nothing but a mental image that has no relationship to the reality of a single system being measured or characterized. Second, the ensemble picture does not fix the probability distribution of initial states, but this must be obtained by other physical considerations, which considerations of course could and should be instead applied directly to the real system, not the imaginary ensemble. Third, fixing the number of ensemble members implies a conservation law for probability that is not always true in reality. Just because ensembles are conserved in someone's mental picture doesn't make it true that probability is conserved in the real world. Fourth, ensembles are unnecessary, and that which is unnecessary has no place in science. Instead one should seek the true physical origin of randomness for the single real system being studied, and one should not be bound by some imaginary scheme that obscures understanding and that blocks the mathematical formulation of randomness in statistical mechanics.

1.4.2.3 Initially imprecise?

As a third explanation, some suggest that randomness arises in deterministic equations of motion from imprecisely specified initial conditions. Alternatively, the precise intermolecular interactions may be unknown or too complex to calculate, and these cause indeterminacy in the system's trajectory. Whilst these are undoubtedly practical impediments to the precise calculation and application of the equations of motion, they cannot be the physical principle that underlies randomness. For example, the irreversibility that is embodied in the Second Law of Thermodynamics is an experimental fact that does not require the specification of initial conditions or intermolecular interactions to be measured. This suggests that randomness and irreversibility must be built in to the very equations of motion that determine the evolution of the types of systems to which the Second Law refers.

1.4.2.4 Reservoir randomness

The quantitative approach to randomness and irreversibility that this book takes is based on the so-called reservoir formalism. The idea is sketched in figure 1.6. In this book what is called 'the system' means the total system (what some others might call the Universe), sub-system means that part of the system of direct interest (some others call this the system), and reservoir means the remaining part of the system (what some others call the environment).
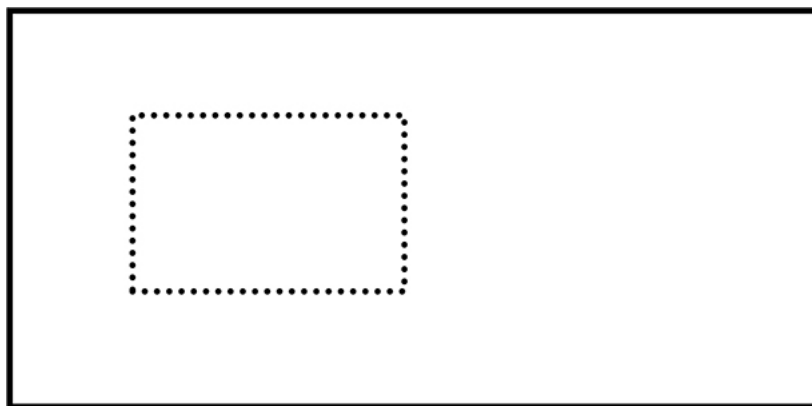
Figure 1.6. A system consisting of a sub-system and a (much larger) reservoir. The total system is isolated; all interactions and exchanges are forbidden across the solid boundary. The type of reservoir is determined by the specific interactions and exchanges that are allowed across the dotted boundary.

It is generally the case that one has a detailed interest in a very small part of the Universe. For example, the shape of a macromolecule may be the focus, in which case the solvent, far-separated solutes, container walls etc are of no interest except in so far as they affect directly the macromolecule. But one cannot ignore these completely, because, for example, the nearby solvent molecules directly affect the conformation of the macromolecule, and the far away solvent molecules act as heat, pressure, and solvent bath. In these circumstances the macromolecule and the nearby solvent molecules form the sub-system, and the far away solvent molecules, container walls, external laboratory etc form the reservoir. The sub-system is followed in molecular detail, and the reservoir is followed at the macroscopic level, which is to say it is characterized solely by intensive thermodynamic parameters such as temperature, pressure, chemical potential, etc.

The reservoir formalism as it gives rise to equilibrium thermodynamics will be set out in chapter 2. Here the relevance of the formalism to randomness and irreversibility will be described. The discussion is in terms of classical mechanics; the application of the formalism to quantum systems is described in chapter 7.

Let $\mathbf{\Gamma}_{\text{tot}}$ denote a point in classical phase space, a microstate, of the total system (i.e. the positions and momenta of all the molecules in the system). Since the latter is isolated it evolves deterministically via Hamilton's equations of motion. The microstate of the total system at time t may be denoted $\mathbf{\Gamma}_{\text{tot}}(t \mid \mathbf{\Gamma}_{\text{tot},0})$, where the initial configuration at $t = 0$ is $\mathbf{\Gamma}_{\text{tot},0}$. (For simplicity, it is assumed that the Hamiltonian of the system is not explicitly time dependent.) This is called the trajectory of the total system. It is both deterministic and reversible.

Deterministic means that the initial point $\mathbf{\Gamma}_{\text{tot},0}$ fully determines all subsequent points. This is to say that if the trajectory were ever to revisit one point, it would also revisit all subsequent points in order, and after the same time intervals. This also means that the trajectory cannot cross itself (i.e. each point on the trajectory is preceded by a unique point, and succeeded by a unique point).

Reversible actually means three symmetries (see section 5.4.2). Time reversibility, which says that if $\mathbf{\Gamma}_2$ transitions from $\mathbf{\Gamma}_1$ after time τ, $\mathbf{\Gamma}_1 \xrightarrow{\tau} \mathbf{\Gamma}_2$, then $\mathbf{\Gamma}_1$ transitions from $\mathbf{\Gamma}_2$ after time $-\tau$, or $\mathbf{\Gamma}_2 \xrightarrow{-\tau} \mathbf{\Gamma}_1$. Conjugate reversibility, which, with a dagger denoting the

state with all velocities reversed, means that the first transition just mentioned, $\mathbf{\Gamma}_1 \xrightarrow{\tau} \mathbf{\Gamma}_2$, implies the velocity reversed transition $\mathbf{\Gamma}_2^\dagger \xrightarrow{\tau} \mathbf{\Gamma}_1^\dagger$. Microscopic reversibility is the combination of these two, namely $\mathbf{\Gamma}_1 \xrightarrow{\tau} \mathbf{\Gamma}_2$, implies $\mathbf{\Gamma}_1^\dagger \xrightarrow{-\tau} \mathbf{\Gamma}_2^\dagger$.

The total phase space comprises that of the sub-system and that of the reservoir, $\mathbf{\Gamma}_{\text{tot}} = \{\mathbf{\Gamma}_s, \mathbf{\Gamma}_r\}$. A microstate of the sub-system is a point in the phase space of the sub-system, $\mathbf{\Gamma}_s$. This gives the position and momenta of all the molecules of the sub-system. The sub-system phase space is a subspace of the total phase space, and a microstate of the sub-system is the projection of a microstate of the total system. Many microstates of the total system, each corresponding to different reservoir microstates, project onto the same microstate of the sub-system. This projection operation has the same effect as what is often called a contracted description.

The projection of the total trajectory onto the sub-system may be written, $\mathbf{\Gamma}_s' = \hat{P}_s \mathbf{\Gamma}(t \mid \mathbf{\Gamma}_{s,0}, \mathbf{\Gamma}_r')$. For a different reservoir but the same sub-system initial point one has a different sub-system trajectory $\mathbf{\Gamma}_s'' = \hat{P}_s \mathbf{\Gamma}(t \mid \mathbf{\Gamma}_{s,0}, \mathbf{\Gamma}_r'')$. Hence the current sub-system configuration alone does not uniquely determine the sub-system evolution; the projection operation makes the sub-system trajectory indeterminant in the sub-system. That is, revisiting the point $\mathbf{\Gamma}_{s,0}$ does not guarantee that subsequent points on the original trajectory will recur exactly, because each $\mathbf{\Gamma}_{s,0}$ could correspond to a different $\mathbf{\Gamma}_r$. Each such configuration of the reservoir molecules has a different influence on the evolution of the sub-system. This means that the projected trajectory can cross itself (figure 1.7) and bifurcate (figure 1.8).

Figure 1.7. A spiral trajectory in three dimensional space (left) becomes a trajectory with a crossing point and loop when projected onto the two-dimensional plane (right).
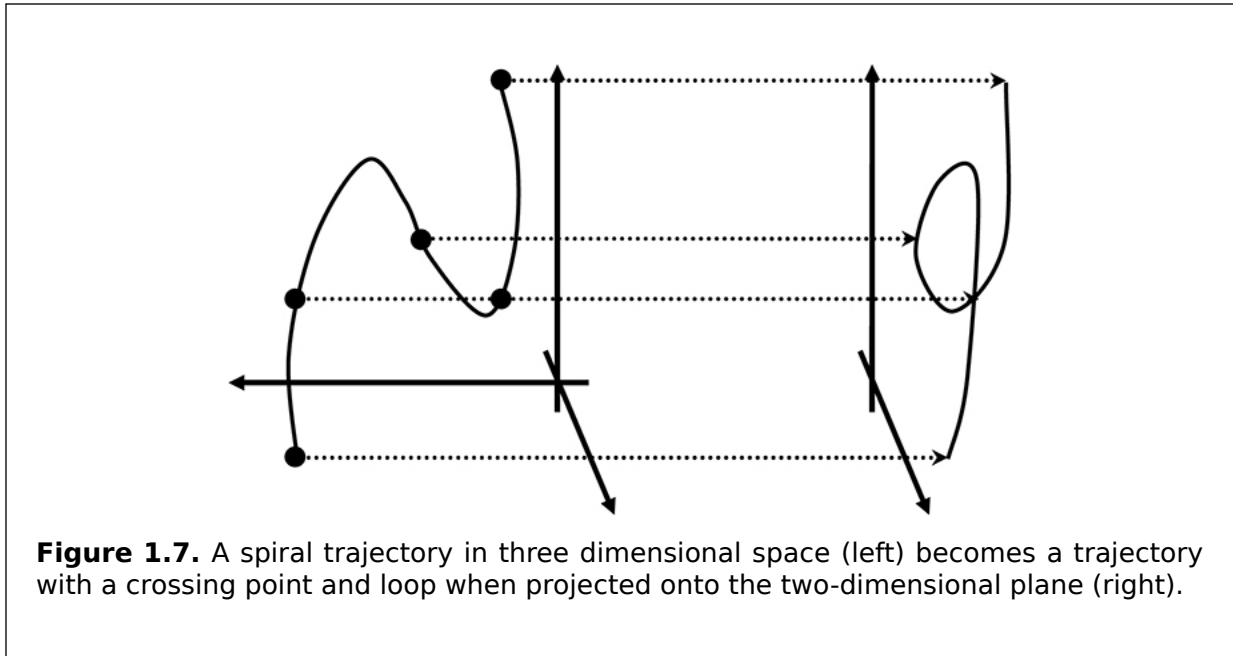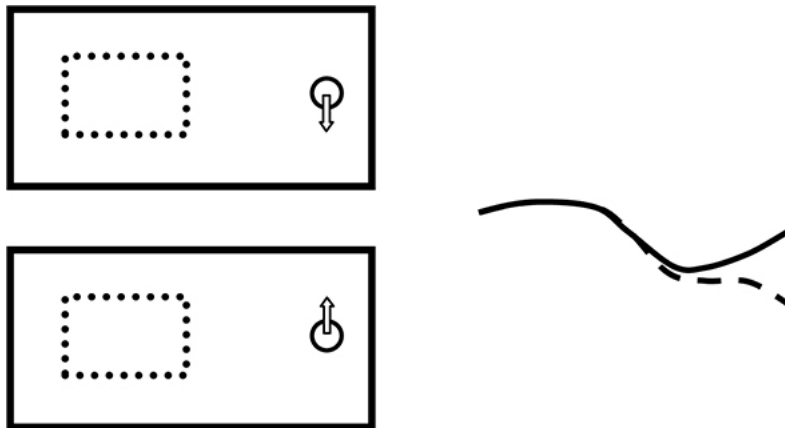
Figure 1.8. Two systems in which the configurations differ only by the velocity of a single reservoir particle far from the sub-system (left). The corresponding sub-system trajectories (right) bifurcate at the moment the perturbing influence of the reservoir particle reaches the sub-system. A similar bifurcation point exists in the past.

This indeterminacy in the sub-system evolution when the reservoir configuration is not treated explicitly is manifest in stochastic terms that are added to the classical picture of Hamilton's deterministic equations of motion, and this gives rise to a probabilistic treatment of the sub-system state. Quantitative treatments of the probability follow from the physical nature of the interaction between the sub-system and the reservoir, as will be given in later chapters (chapters 4 and 6).

This interpretation of randomness in the equations of motion—it arises from the projection of the total system onto the sub-system that is being treated explicitly—also accounts for the second objection that was made to Boltzmann's identification of entropy with the number of molecular configurations, namely that the equations of motion are time-reversible but the Second Law of Thermodynamics gives a preferred direction to time. It turns out that the projection operation contributes both stochastic and dissipative terms to the equations of motion, and that these are time irreversible. It will be shown that the dissipative term is a force that increases the entropy, as legislated by Clausius.

This randomness in the equations of motion leads directly to statistical mechanics and the need for a probabilistic treatment of molecular configurations. There is a direct link between randomness in statistical mechanics and thermodynamics. Different thermodynamic systems are based on the sub-system being able to exchange specified quantities (generally linear additive conserved variables such as energy, volume, number etc) with the reservoir. The mechanism by which this occurs is most usually via intermolecular interactions across the boundary. Therefore, the current configuration of the molecules of the reservoir influences the motion of the molecules of the sub-system near the boundary immediately, and all the molecules of the sub-system ultimately. The physical interpretation of randomness as arising from the projection of external influences leads to the quantitative evaluation of probability.

Summary

- The entropy of a system or state is the logarithm of the weight of the system or state. In the simplest case of equally likely microstates, weight is number. States with high entropy may be thought of as disordered or unpredictable.

- Macrostates are complete and disjoint. Microstates are complete, disjoint, and indivisible.
- The probability of a state is the exponential of its entropy. Probability, like entropy and weight, is a physical, measurable property.
- Conditional transitions to a more probable state are more probable than the reverse. The Second Law of Thermodynamics for conditional transitions is probably true.
- Randomness and irreversibility arise from the projection of the deterministic motion of the total system onto the sub-system of interest.

References

Attard P 1999 *Markov superposition expansion for the entropy and correlation functions in two and three dimensions Statistical Physics on the Eve of the Twenty-First Century* ed Batchelor M T (ed) and Wille L T (ed) (Singapore: World Scientific)

Attard P 2000 The explicit density functional and its connection with entropy maximisation *J. Stat. Phys.* **100** 445

Attard P 2002 *Thermodynamics and Statistical Mechanics: Equilibrium by Entropy Maximisation* (London: Academic)

Attard P 2012 Is the information entropy the same as the statistical mechanical entropy? arXiv:12095500v1 https://arxiv.org/abs/1209.5500

Baxter R J 1982 *Exactly Solved Models in Statistical Mechanics* (London: Academic)

Boltzmann L 1866 Über die mechanische bedeutung des zweiten hauptsatzes der wärmetheorie *Wien. Ber.* **53** 195

Boltzmann L 1895 *Vorlesungen über Gastheorie* **vols 1 and 2** (Leipzig)

Browning E B 1850 *Sonnets from the Portuguese, no. 43*

Brush S G (ed) 1964 *Lectures on Gas Theory* (Berkeley: University of California Press) (Engl. Trans. of Boltzmann 1895)

Clausius R 1850 Über die bewegende kraft der wärme *Ann. Phys.* **79** 368, 500

Clausius R 1851 On the Moving force of heat, and the laws regarding the nature of heat itself which are deducible therefrom *Phil. Mag.* **2** 1, 102 (Eng. trans.)

Gibbs J W 1902 *Elementary Principles in Statistical Mechanics Developed with Special Reference to the Rational Foundation of Thermodynamics* (New Haven, CT: Yale University Press)

Jaynes E T 1957 Information theory and statistical mechanics *Phys. Rev.* **106** 620

Jaynes E T 2003 *Probability Theory: The Logic of Science* ed Bretthorst G L (ed) (Cambridge: Cambridge University Press)

Kline S J 1999 *The Low-Down on Entropy and Interpretive Thermodynamics* (La Cañada, CA: DCW Industries) p 12

Popper K R 1959 The propensity interpretation of probability *J. Philos. Sci.* **10** 25

Shannon C E 1948 A mathematical theory of communication *Bell Syst. Tech. J.* **27** 379–423

Shannon C E and Weaver W 1949 *The Mathematical Theory of Communication* (Urbana, IL: University of Illinois Press)

von Neumann J circa 1939 to Claude E Shannon recounted in Tribus M and McIrvine E C 1971 Energy and information *Sci. Am.* **224** 179

Entropy Beyond the Second Law (Second Edition)

Thermodynamics and statistical mechanics for equilibrium, non-equilibrium, classical, and quantum systems

Phil Attard

Chapter 2

Variational formulation of equilibrium thermodynamics via entropy and free energy

'I propose to call [it] the entropy of the body, from the Greek word 'trope' for 'transformation'. I have intentionally formed the word 'entropy' to be as similar as possible to the word 'energy'; for the two magnitudes to be denoted by these words are so nearly allied in their physical meanings, that a certain similarity in designation appears to be desirable.'

(Clausius 1854)

'Any method involving the notion of entropy, the very existence of which depends on the second law of thermodynamics, will doubtless seem to many far-fetched, and may repel beginners as obscure and difficult of comprehension'

(Gibbs 1873)

'[A reformulation of the postulates] exhibits explicitly the internal consistency of the logical structure...it enables a deeper insight and intuition...and important extensions and generalizations are suggested and made practical'

(Callan 1960)

'One makes a discovery when one sees what everybody else does, but thinks what nobody else had thought before'

(Szent-Györgyi 2011)

The main aim of this chapter is to set out the general thermodynamic procedure for treating equilibrium systems. The procedure for obtaining the thermodynamic potential is given in general, with examples from common equilibrium systems derived. The novel twist on conventional thermodynamics is the relation between total entropy and free energy, from which arises the variational principle based on the Second Law of Thermodynamics, the nature of the constrained quantities that are its subject, and the relation to fluctuation theory.

The so-called reservoir formalism is used. The idea of a reservoir or bath is mainly due to Gibbs (1902), and is standard in thermodynamics (Callan 1960, Sears and Salinger 1986). The equilibrium free energies for various reservoirs derived here are the same as the conventional ones. However, the interpretation of these in terms of total entropy, and the consequent variational principle, are as developed in my previous book (Attard 2002).

As a concrete example to illustrate the general ideas, the canonical equilibrium system is used, which is a sub-system in thermal contact with a heat reservoir. This will bring us to the Helmholtz free energy,

$$F = E - TS.$$

(2.1)

Here E is the energy, T is the temperature and S is the entropy. This expression appears in all books on thermodynamics, but the question is whether or not it is really

understood. In particular, what exactly is 'the' entropy *S* that appears here? An example of what commonly goes wrong is the associated idea that

$$\text{the Helmholtz free energy should be minimized.}$$

(2.2)

By the end of the chapter readers will understand what is wrong with this exhortation, they will appreciate the real physical import of the Helmholtz free energy, and they will be able to write and use the expression more powerfully.

2.1 Entropy of an isolated system

In chapter 1 the First Law of Thermodynamics was introduced as a law for the conservation of energy. Conservation led to counting, counting to linear additivity, and linear additivity to entropy. It was pointed out that although not all linear additive quantities were conserved quantities, linear additivity was a very useful property to have and that it is widely exploited in the physical science.

The fundamental quantities that have proved most useful in thermodynamics are the conserved linear additive quantities energy *E*, number *N*, and volume *V*. Figure 2.1 shows an isolated system that is characterized by the values of these three quantities. The energy is composed of kinetic energy (i.e. due to particle motion) and potential energy (i.e. due to inter-particle interactions, as well as possible interactions with the walls and other fixed external fields). Despite the incessant motion of the particles that make up the system, these three quantities, *E*, *N*, and *V*, are unchanging.

$$NVE$$

Figure 2.1. An isolated system with number of particles *N*, volume *V*, and energy *E*.
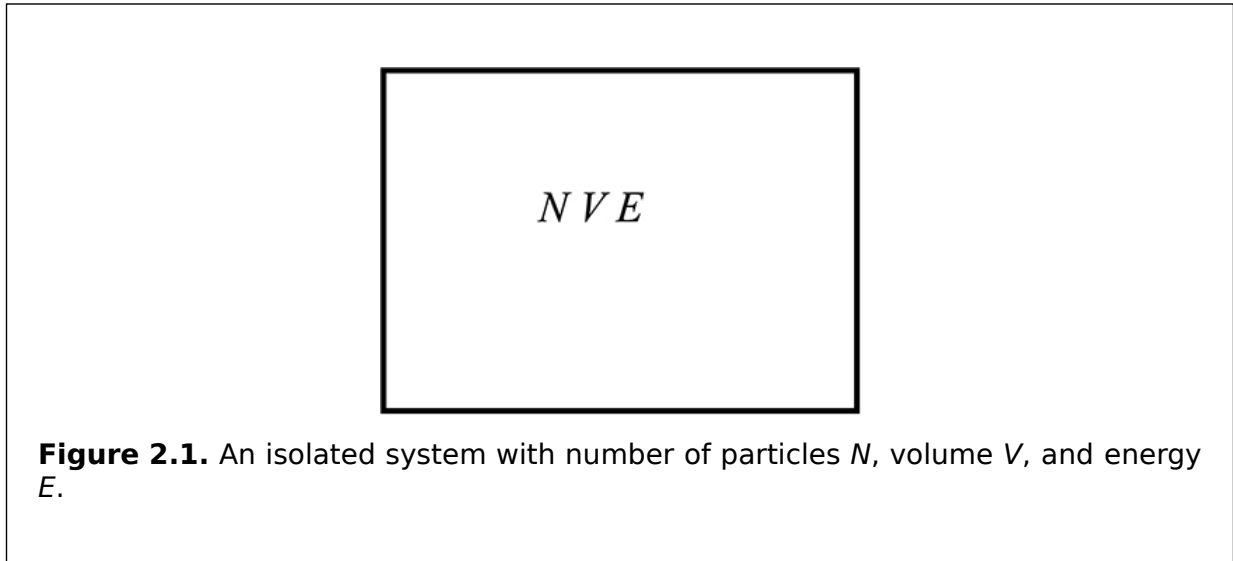
For a multi-component system, one can regard *N* as a vector whose components, N_α say, represent the number of each species.

Other linear additive quantities, such as the linear or angular momentum, can also be constants of the motion if the isolated system is not subject to external forces (and therefore enjoys translational or rotational symmetry). However, they are not widely used in thermodynamics.

Any function of these constants is also a constant of the motion. Hence one can *define* the entropy of the isolated system as dependent on them, $S(E, N, V)$. In the language of chapter 1, the particular state point with values $\{E, N, V\}$ is the

macrostate formed from the intersection of the three macrostates in the three different collectives. The entropy $S(E, N, V)$ is the logarithm of the weight of the molecular microstates in that particular macrostate.

As mentioned, experience has shown the utility of formulating the entropy as dependent solely on extensive variables, and a number of results that flow directly from this will now be established. A further reason will be seen from the discussion of fluctuation theory in section 2.5, where the Gaussian distribution of the fluctuations follows directly from the central limit theorem, since by definition extensive variables are the sum of linear additive variables that themselves are randomly distributed. The same cannot be said of intensive variables.

2.1.1 Extensivity

Figure 2.2 shows two systems, both composed of λ sub-systems. In both cases the total energy is $E_{\text{tot}} = \lambda E_1$, the total number is $N_{\text{tot}} = \lambda N_1$, and the total volume is $V_{\text{tot}} = \lambda V_1$. In the first system, the sub-systems are identical and isolated, which is to say each is in the macrostate $\{E_1, N_1, V_1\}$. Since entropy is linear additive, the entropy of the first system is

$$S(E_{\text{tot}}, N_{\text{tot}}, V_{\text{tot}}; \lambda) = \lambda S(E_1, N_1, V_1).$$

(2.3)

On the left-hand side λ has been included explicitly to emphasize the fact that the total system comprises λ isolated sub-systems. On the right-hand side the isolated sub-system entropy $S(E_1, N_1, V_1)$ includes contributions from the boundaries that contain and isolate it from the rest of the Universe. However, in the thermodynamic limit of large sub-system size, such boundary effects scale with a boundary volume (the surface area times a molecular interaction length), which is negligible compared to the sub-system volume. Therefore, we may call $S(E_1, N_1, V_1)$ 'the' sub-system entropy in the macrostate $\{E_1, N_1, V_1\}$ irrespective of the precise nature of the sub-system boundary.

Figure 2.2. A system with total energy E_{tot}, total number N_{tot}, and total volume V_{tot} composed of $\lambda = 9$ identical isolated sub-systems (left), and composed of $\lambda = 9$ same-sized sub-systems able to exchange energy and number (right).
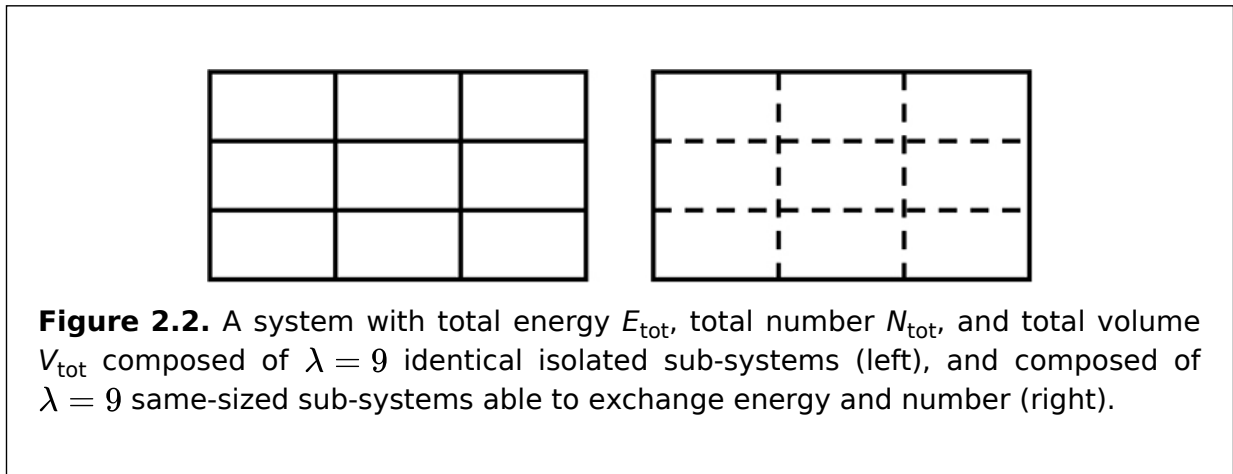
The second system shown in figure 2.2 contains λ equal sized sub-systems that can exchange energy and number with each other. These sub-systems are really imaginary cells that sub-divide the total volume. Because the cells are all the same size, symmetry arguments indicate that most likely they have the same energy

$E_1 = E_{\text{tot}}/\lambda$ and number $N_1 = N_{\text{tot}}/\lambda$. (It will turn out that thermodynamics always deals with the most likely state; in the thermodynamic limit this in fact equals the average state). These most likely values will be signified with an overline in much of the analysis that follows. The entropy of this system is

$$S(E_{\text{tot}}, N_{\text{tot}}, V_{\text{tot}}) \gtrsim \lambda S(E_1, N_1, V_1).$$

(2.4)

On the left-hand side is a single isolated total system. Two assumptions are made to compare this to the right-hand side. First, for a given macrostate, the cells with 'transparent' boundaries that allow interactions and passage across them have the same entropy as isolated cells with 'opaque' boundaries that prohibit interactions and particle transport. This is justified in the thermodynamic limit where boundary conditions make negligible contribution to the total. Second, the left-hand system in figure 2.2 is constrained to be in a specific macrostate, namely the most likely one of equal energy and number in all the sub-systems, whereas the right-hand system can be in this macrostate, and in many other besides. Since the right-hand system is unconstrained in this respect, it must have available more microscopic configurations than the left-hand system and its entropy must be strictly greater than that of the constrained system. Therefore, one must have as a strict inequality

$$S(E_{\text{tot}}, N_{\text{tot}}, V_{\text{tot}}) > S(E_{\text{tot}}, N_{\text{tot}}, V_{\text{tot}}; \lambda) = \lambda S(E_1, N_1, V_1).$$

(2.5)

(For a detailed illustration of this point, see the combinatorial example in section 1.2.3). The difference between the entropies of the two systems is due to the fluctuations in the energy and number of the sub-systems of the right-hand total system (ignoring, as above, other boundary effects).

The entropy due to these fluctuations is negligible compared to the entropy of the most likely macrostate, which, by symmetry, is the state in which the sub-systems all have the same energy and number. The reason that the fluctuation entropy is relatively negligible is because of the large value of the constrained entropy to which it is being added: the most likely macrostate is the state of maximum constrained entropy. This is true by definition, since probability is the exponential of the entropy, and 'most likely' means 'most probable'. More quantitatively, to obtain the unconstrained entropy, one must add to the maximum value of the constrained entropy the entropy due to fluctuations about the maximal macrostate. But, as will be shown below, the fluctuation entropy scales with the square root of the size of the sub-system, whereas the entropy of the sub-system scales with its size, as will next be shown. The former is negligible compared to the latter in the thermodynamic limit of large sub-system size. This is the reason for turning the strict inequality into an approximate inequality. In fact, in the thermodynamic limit one can write

$$S(E_{\text{tot}}, N_{\text{tot}}, V_{\text{tot}}) \approx S(E_{\text{tot}}, N_{\text{tot}}, V_{\text{tot}}; \lambda) = \lambda S(E_1, N_1, V_1).$$

(2.6)

This says that the unconstrained total entropy is approximately equal to the maximum value of the constrained total entropy. This is a general rule that is illustrated by the combinatorial example shown in figure 1.4.

These last three equations try to explain two things. On the one hand the total unconstrained entropy necessarily must be strictly greater than any one value of the constrained total entropy, including the maximal value of the constrained total entropy. On the other hand, in the thermodynamic limit fluctuations about the most

likely state give a negligible contribution, and so the maximal constrained total entropy may be taken to equal the total unconstrained entropy.

2.1.2 Intensive variables

In the final form for the total entropy, λ acts as scale factor. It says that if one scales the extensive (i.e. linear additive) conserved variables all by the same amount, then the total entropy is also scaled by that amount. This means that the entropy is an extensive variable, which is to say that it scales with the size of the system. The extensivity of the entropy may be signified as

$$S(E_1, N_1, V_1) \sim \mathcal{O}(V_1).$$

(2.7)

In view of this it is useful to define the energy and number densities,

$$\varepsilon_1 \equiv \frac{E_1}{V_1}, \quad \text{and} \quad \rho_1 \equiv \frac{N_1}{V_1}.$$

(2.8)

Because of extensivity, the entropy density is a function of these alone,

$$\sigma(\varepsilon_1, \rho_1) \equiv \frac{S(E_1, N_1, V_1)}{V_1}.$$

(2.9)

These densities are intensive variables, which is to say that they are independent of the size of the system, other things being equal (i.e. all extensive independent variables are scaled by the same amount). One reason intensive variables are useful is because they are localized in the sense that they don't depend on the whole system. They are often used as a first approximation for the local properties of an inhomogeneous system. Intensive variables are sometimes called field variables.

The mathematical operation of differentiation is essentially the same as forming a fraction. Hence the derivative of an extensive variable with respect to another extensive variable yields an intensive variable. This is one reason why it is useful to define the isolated system entropy as a function of the extensive variables that form the macrostates, namely the derivatives of the entropy with respect to its independent variables are intensive.

The first such derivative is with respect to energy, which is defined to yield the temperature,

$$T^{-1} \equiv \frac{\partial S(E, N, V)}{\partial E}.$$

(2.10)

The temperature of the isolated system is intensive and dependent on the state of the isolated system, $T(E, N, V)$. At the moment this is just a mathematical definition of the function T, and it remains to show that it has the same physical properties as the temperature of familiar experience.

The derivative with respect to volume yields the pressure,

$$p \equiv T\frac{\partial S(E,N,V)}{\partial V},$$

(2.11)

and that with respect to number yields the chemical potential,

$$\mu \equiv -T\frac{\partial S(E,N,V)}{\partial N}.$$

(2.12)

For a multi-component system, there is a chemical potential conjugate to the number of each species, $\mu_\alpha = -T\partial S(E,\underline{N},V)/\partial N_\alpha$. Again these are intensive variables that are functions of the state of the isolated system, $p(E,N,V)$ and $\mu(E,N,V)$, and these can be shown to have the same properties as the physical quantities of the same name.

With these definitions of the partial derivative of the entropy, the total derivative of the entropy of an isolated system is

$$\mathrm{d}S = \frac{1}{T}\mathrm{d}E + \frac{p}{T}\mathrm{d}V - \frac{\mu}{T}\mathrm{d}N.$$

(2.13)

In the thermodynamic limit, the intensive variables defined by the above entropy derivatives are functions only of the energy and number density. With the entropy density $\sigma(\varepsilon,\rho) \equiv S(E,N,V)/V$, this may be seen explicitly,

$$\frac{1}{T} = \left(\frac{\partial V\sigma(\varepsilon,\rho)}{\partial V\varepsilon}\right)_{N,V} = \left(\frac{\partial \sigma(\varepsilon,\rho)}{\partial \varepsilon}\right)_\rho,$$

(2.14)

$$\frac{-\mu}{T} = \left(\frac{\partial V\sigma(\varepsilon,\rho)}{\partial V\rho}\right)_{E,V} = \left(\frac{\partial \sigma(\varepsilon,\rho)}{\partial \rho}\right)_\varepsilon,$$

(2.15)

and

$$\begin{aligned}
\frac{p}{T} &= \left(\frac{\partial V\sigma(\varepsilon,\rho)}{\partial V}\right)_{E,N} \\
&= \left(\frac{\partial V\sigma(\varepsilon,\rho)}{\partial V}\right)_{\varepsilon,\rho} + \left(\frac{\partial V\sigma(\varepsilon,\rho)}{\partial \varepsilon}\right)_{\rho,V}\left(\frac{\partial \varepsilon}{\partial V}\right)_E + \left(\frac{\partial V\sigma(\varepsilon,\rho)}{\partial \rho}\right)_{\varepsilon,V}\left(\frac{\partial \rho}{\partial V}\right)_N \\
&= \sigma(\varepsilon,\rho) - \frac{\varepsilon}{T} + \frac{\rho\mu}{T}.
\end{aligned}$$

(2.16)

That is, in the thermodynamic limit one has $T(\varepsilon,\rho)$, $p(\varepsilon,\rho)$, and $\mu(\varepsilon,\rho)$.

2.1.3 Concavity of the entropy

For the case of the system consisting of λ identical sub-systems able to exchange energy with each other, the right-hand part of figure 2.2, the equilibrium state is the one where the sub-systems all have the same energy, $E = E_{\text{total}}/\lambda$. (For simplicity,

focus on only energy in the first instance.) This follows from symmetry arguments, or from experience. Any other distribution of energy $\{E_i\}$ must have a lower entropy. Hence

$$\lambda S(E, N, V) \geqslant \sum_{i=1}^{\lambda} S(E_i, N, V)$$

$$= \lambda S(E, N, V) + \frac{\partial S(E, N, V)}{\partial E} \sum_{i=1}^{\lambda} (E_i - E)$$

$$+ \frac{1}{2} \frac{\partial^2 S(E, N, V)}{\partial E^2} \sum_{i=1}^{\lambda} (E_i - E)^2,$$

where a second order Taylor expansion about *E* has been performed. The first term on the right-hand side cancels with the left-hand side. The second term is zero from energy conservation, $\sum_{i=1}^{\lambda} E_i = \lambda E$. These mean that the third term must be negative semi-definite, vanishing if and only if $E_i = E$. Since it contains a sum of non-negative terms, one concludes that the second energy derivative of the entropy must be strictly negative

$$S_{EE} < 0.$$

Hence the entropy of an isolated system is strictly concave with respect to energy.

The argument can be repeated for the other extensive variables, either one at a time or allowing simultaneous exchange. The general conclusion is

$$S_{aa} < 0, \quad \text{and} \quad S_{aa}S_{bb} - S_{ab}^2 > 0,$$

where the subscripts denote the second derivatives with respect to any of the isolated sub-system extensive variables. The second form follows because the eigenvalues of the Jacobean must be negative, and so the determinant of the Jacobean of the extensive variables taken pairwise must be positive. (One can obtain higher order conditions as well.) These are necessary conditions for any thermodynamic state to be stable.

Stable means that the system can exist with uniform densities corresponding to these values of the extensive variables. A system in a uniform state in which this concavity condition is violated can increase its total entropy by sub-dividing into coexisting phases, each with a uniform density that is stable, and with overall densities equal to the original values.

2.2 Heat reservoir and the Helmholtz free energy

We now turn explicitly to the canonical equilibrium system, namely a sub-system able to exchange energy with a heat reservoir, figure 2.3. As mentioned in section 1.4.2, the reservoir formalism is fundamental to how randomness occurs in statistical mechanics, and how statistical mechanics and thermodynamics are structured, and so it is worth defining in detail what is meant by a reservoir.
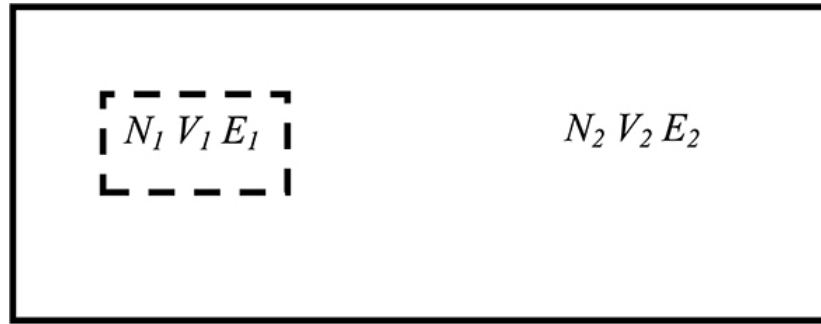
Figure 2.3. A system composed of a sub-system 1 and reservoir 2 able to exchange energy.

A system comprising a sub-system and a reservoir has four characteristics: the system is isolated, the reservoir is infinitely larger than the sub-system, the boundary region between the reservoir and the sub-system is infinitely smaller than the sub-system, and the reservoir and the sub-system can exchange one or more linear additive, conserved variables. (In practice, in order to establish the size of the sub-system there must be at least one linear additive, conserved variable that cannot be exchanged.)

The reason for insisting that the sub-system be large compared to its boundary region (i.e. its surface area times the thickness of direct molecular interactions) is so that boundary effects can be neglected. This condition is also known as the thermodynamic limit. The reason for the reservoir being large compared to the sub-system is so that higher order terms in a Taylor expansion can be neglected, as will be required shortly.

In what follows it will be important to keep track of the dependent and the independent variables. To date energy E has been taken as an independent variable for the isolated system entropy, $S(E, N, V)$. In part of what follows it will be necessary to regard it as dependent on the temperature, at which time it will be denoted $E(N, V, T)$ or by $\bar{E}$. By the partial derivative definition, we have already seen that temperature is a dependent variable, $T(E, N, V)$. As will be discussed shortly, the properties of entropy make the relationship between energy and temperature one-to-one: $T_1 = T(E_1, N, V)$ if and only if $E_1 = E(N, V, T_1)$. In other words, $T_1 = T(E(N, V, T_1), N, V)$.

2.2.1 Constrained total entropy

For the system comprising a sub-system able to exchange energy with a reservoir, figure 2.3, the total energy is fixed, $E_{\text{total}} = E_1 + E_2$. Because the reservoir is infinitely larger than the sub-system, the sub-system energy is much less than the total energy, $E_1 \ll E_{\text{total}}$. This holds in practice for all conceivable configurations of the system. Taking the thermodynamic limit for the sub-system, the boundary conditions become negligible, and the entropy of the total system (in the energy macrostate E_1) is just the sum of the isolated sub-system and isolated reservoir entropies (each in their corresponding macrostate). These facts mean that the entropy of the total system, constrained to be in the macrostate E_1, is

$$
\begin{aligned}
S_{\text{total}}&(E_1 \mid E_{\text{total}}, N_1, V_1, N_2, V_2) \\
&= S_1(E_1, N_1, V_1) + S_2(E_{\text{total}} - E_1, N_2, V_2) \\
&= S_1(E_1, N_1, V_1) + S_2(E_{\text{total}}, N_2, V_2) - E_1 \frac{\partial S_2(E_{\text{total}}, N_2, V_2)}{\partial E_{\text{total}}} + \cdots \\
&= S_1(E_1, N_1, V_1) + \text{const.} - \frac{E_1}{T_2}.
\end{aligned}
$$

The first equality writes the system entropy as the sum of that of the isolated sub-system and isolated reservoir. Here we have allowed for the possibility that the reservoir and sub-system are composed of different materials, and therefore their entropies may be different functions of their arguments, namely S_1 and S_2. This is a detail that is already irrelevant by the final equality.

The second equality performs a Taylor expansion of the reservoir entropy about the total energy. The leading, zeroth order, term, $S_2(E_{\text{total}}, N_2, V_2)$, is a constant completely independent of the sub-system. It therefore has no effect on the state of the sub-system and it can be neglected. It will be recalled from section 1.1.2 that entropy was only defined up to an additive constant.

The first order term is linear in the sub-system energy and the (inverse) reservoir temperature $T_2(E_{\text{total}}, N_2, V_2)$. This term, $-E_1/T_2$, is extensive with the sub-system size, $\mathcal{O}(V_1)$, as is the sub-system entropy itself, $S_1(E_1, N_1, V_1) \sim \mathcal{O}(V_1)$. The second order term in the Taylor expansion, which has been neglected, is

$$
\frac{1}{2} E_1^2 \frac{\partial^2 S_2(E_{\text{total}}, N_2, V_2)}{\partial E_{\text{total}}^2} \sim \mathcal{O}\left(\frac{V_1^2}{V_2}\right).
$$

In the second derivative here, the numerator scales with V_2 and the denominator with V_2^2. Hence the full second order term is $\mathcal{O}(V_1^2/V_2)$, which is a factor of V_1/V_2 smaller than the two terms that are explicitly retained above. (Both $S_1(E_1, N_1, V_1)$ and E_1/T_2 are $\mathcal{O}(V_1)$.) Since the reservoir is infinitely larger than the sub-system, this and higher order terms are completely negligible. This is one of the reasons why it is advantageous to write the entropy as a function of extensive variables only. The net effect of neglecting the higher order terms in the Taylor expansion of the reservoir is that the temperature of the reservoir is fixed and unchanging no matter how much energy is exchanged with the sub-system. In other words, $T_2(E_2, N_2, V_2) = T_2(E_{\text{total}}, N_2, V_2)$, with negligible error $\mathcal{O}(V_1/V_2)$.

As a result of these manipulations, one sees that the heat reservoir only enters the equations via its temperature T_2. Hence one can write the constrained total entropy as

$$
S_{\text{total}}(E_1 \mid N_1, V_1, T_2) = S_1(E_1, N_1, V_1) - \frac{E_1}{T_2}.
$$

Since the temperature is that of the reservoir, and all other properties belong to the sub-system, one can abbreviate this still further and write

$$S_{\text{total}}(E \mid N, V, T) = S_{\text{s}}(E, N, V) - \frac{E}{T}.$$

Of course this concise expression is easy to misinterpret if one is unaware of its derivation. Hopefully the reader will never be misled; T is the temperature of the reservoir and S_{s} is the entropy of the isolated sub-system. The reader will also know that the left-hand side is the entropy of the total system constrained to be in the macrostate that the sub-system has energy E. The reader will also recognize that $-E/T$ is the sub-system-dependent part of the reservoir entropy. Finally, the reader will notice that all four of the arguments of the total entropy are independent.

Just a note on notation. The vertical bar here is the same as that used for the conditional probability. Variables to the left of it are the constrained (or exchangeable, or fluctuating) variables, and variables to the right of it are the conditioning (or fixed) variables. The exponential of the total entropy essentially gives the probability that the sub-system has energy E given the conditions, as will be discussed shortly. This explains the notation that is used here and in similar circumstances throughout.

The energy derivative of the constrained total entropy is

$$\frac{\partial S_{\text{total}}(E \mid N, V, T)}{\partial E} = \frac{\partial S_{\text{s}}(E, N, V)}{\partial E} - \frac{1}{T}.$$

The first term on the right-hand side is $1/T(E, N, V)$, which is the reciprocal of the sub-system temperature, and the second term is the reciprocal the reservoir temperature.

Now one can appeal to the Second Law of Thermodynamics to determine the equilibrium state. The equilibrium state is the most likely value of the sub-system energy. The total entropy is maximized by the energy that makes its derivative vanish. This gives the equilibrium energy $\bar{E}$ as the one satisfying

$$\frac{\partial S_{\text{total}}(E \mid N, V, T)}{\partial E}\Big|_{E=\bar{E}} = 0 \iff T(\bar{E}, V, N) = T.$$

This says that equilibrium (i.e. the macrostate of maximum total entropy) corresponds to temperature equality between the sub-system and the heat reservoir. This is a version of the Zeroth Law of Thermodynamics. This is an implicit equation for the equilibrium energy of the sub-system, $\bar{E} = E(N, V, T)$.

In the case that the sub-system has a lower temperature than the reservoir, $\partial S(E, N, V)/\partial E = 1/T(E, N, V) > 1/T$, then the total constrained entropy increases when energy is transferred from the reservoir to the sub-system, $\partial S_{\text{total}}(E \mid N, V, T)/\partial E > 0$. The opposite occurs in the case that the sub-system is hotter than the reservoir. In common parlance, heat flows from a hot body to a cold body. This is consistent with the fact that an isolated system has entropy that is a concave function of energy, $S_{EE} < 0$, since this means that its temperature must increase with energy, $\partial T(E, N, V)/\partial E = -T^2 S_{EE} > 0$. Hence energy flowing to a cold body increases its temperature.

2.2.2 Constrained Helmholtz free energy

The constrained total entropy is the mathematical representation of the Second Law of Thermodynamics. Constraining the energy *E* fixes the sub-system in a not-in-equilibrium state. When the constraint is relaxed, the system will move toward equilibrium by exchanging energy with the reservoir. These spontaneous changes in the sub-system energy state are in the direction of increasing constrained total entropy.

The equilibrium state is the state of maximum constrained total entropy.

It is important to note the physical basis of the above derivation. The constrained variable in this case is the sub-system energy, and it is with respect to this, and only this, that the constrained total entropy is maximized in the equilibrium state.

The present variational principle maximizes the constrained total entropy. Alternatively, one can minimize a thermodynamic potential. The general definition of this is simply the negative of the (reservoir) temperature times the constrained total entropy. The negative sign converts the maximum into a minimum. Multiplying the entropy by the temperature gives the units of energy.

In the present case the thermodynamic potential is the constrained Helmholtz free energy,

$$F(E \mid N, V, T) \equiv -TS_{\text{total}}(E \mid N, V, T)$$
$$= E - TS_{\text{s}}(E, N, V).$$

(2.26)

By design, this is a minimum in the equilibrium state, and the energy flow is down the constrained free energy gradient.

The same rule applies in the general case: the constrained thermodynamic potential is the negative of the temperature times the constrained total entropy, $F(X \mid Y, x, T) = -TS_{\text{total}}(X \mid Y, x, T)$, where the *X* are the exchangeable variables, the *x* are the conjugate reservoir field variables, the *Y* are the fixed sub-system variables, and *T* is the temperature of the reservoir. Concrete examples will be given in section 2.4.

Because the relationship between the thermodynamic potential and the total entropy is trivial, the properties of the former are essentially the same as the latter. It is a mathematical and physical fact that the thermodynamic potential is redundant. The present author's main argument against it is that it is unnecessary, and it obscures the role of entropy and its physical origin. Such rational arguments, however, carry little weight against convention and historical tradition. It is more likely that energy will flow from a cold body to a hot body than it is that the free energy will be replaced by the total entropy in textbooks and the literature.

The *equilibrium thermodynamic potential* is defined as the minimum value of the constrained thermodynamic potential. For the present heat reservoir this is the Helmholtz free energy, and the minimum obviously occurs at $\bar{E} = E(N, V, T)$,

$$\bar{F}(N, V, T) \equiv F(\bar{E} \mid N, V, T) = \bar{E} - TS_{\text{s}}(\bar{E}, N, V).$$

(2.27)

The Helmholtz free energy is a function of just three independent variables, namely *N*, *V*, and *T*, since $\bar{E} = E(N, V, T)$ is a dependent variable. The entropy that appears on the right of the definition of the Helmholtz free energy is that of the isolated sub-system with the equilibrium energy $\bar{E}$. The overline on the Helmholtz free energy indicates that it is an equilibrium property.

Note that as a short-hand notation, sometimes we write $\bar{S}_\mathrm{s}(N, V, T)$ instead of $S_\mathrm{s}(\bar{E}(N, V, T), N, V)$. This is not ideal notation because entropy is fundamentally a function of extensive variables. Also, it risks confusion with the unconstrained total entropy $S_\mathrm{total}(N, V, T)$, which is essentially the logarithm of the partition function. This possible problem is exacerbated when the subscript s for sub-system is dropped. In any case it should always be kept in mind that the entropy that appears explicitly in the formula for the Helmholtz free energy is the entropy of the isolated sub-system in the equilibrium energy macrostate.

The equilibrium thermodynamic potential, $F_\mathrm{TD}(N, V, T)$, contains less information than the constrained thermodynamic potential, $F(E \mid N, V, T)$, since it is a function of three independent variables instead of four. The constrained thermodynamic potential describes the approach to energy equilibrium and energy fluctuations about the equilibrium state, whereas the equilibrium thermodynamic potential only describes the equilibrium state. Because the constrained thermodynamic potential describes the excursions of the system to not-in-equilibrium states (see section 2.5), it may also be called the *fluctuation potential*.

The three types of total entropy and the respective free energies are summarized in table 2.1. The partition function is the total weight (see section 2.5.3),

$$Z(N, V, T) = \sum_E \mathrm{e}^{S_\mathrm{total}(E \mid N, V, T)/k_\mathrm{B}}.$$

(2.28)

By definition, the logarithm of the total weight is the total unconstrained entropy $S_\mathrm{total}(N, V, T) = k_\mathrm{B} \ln Z(N, V, T)$. Since the free energy in general is minus the temperature times the total entropy, the statistical mechanical free energy follows as $F_\mathrm{SM}(N, V, T) = -T S_\mathrm{total}(N, V, T)$. The thermodynamic free energy is the minimum value of the constrained free energy, and in the present canonical equilibrium case it is called the Helmholtz free energy, $F_\mathrm{TD}(N, V, T) = -T S_\mathrm{total}(\bar{E} \mid N, V, T) = \bar{E} - T S_\mathrm{s}(\bar{E}, N, V)$. One has the inequalities

$$S_\mathrm{total}(N, V, T) > S_\mathrm{total}(\bar{E} \mid N, V, T) \geqslant S_\mathrm{total}(E \mid N, V, T).$$

(2.29)

In the thermodynamic limit, fluctuations are relatively negligible, and the inequality can be replaced by an equality and $S_\mathrm{total}(N, V, T) \approx S_\mathrm{total}(\bar{E} \mid N, V, T)$, or $F_\mathrm{SM}(N, V, T) \approx F_\mathrm{TD}(N, V, T)$.

Table 2.1. Canonical equilibrium total entropies and free energies.

	Entropy	Free energy	Formula
Constrained	$S_\mathrm{total}(E \mid N, V, T)$	$F(E \mid N, V, T)$	$= E - T S_\mathrm{s}(E, N, V)$
Maximum constrained	$S_\mathrm{total}(\bar{E} \mid N, V, T)$	$F_\mathrm{TD}(N, V, T)$	$= \bar{E} - T S_\mathrm{s}(\bar{E}, N, V)$
Unconstrained	$S_\mathrm{total}(N, V, T)$	$F_\mathrm{SM}(N, V, T)$	$= -k_\mathrm{B} T \ln Z(N, V, T)$

One can compare the conventional expression for the Helmholtz free energy given at the beginning of this chapter, equation (2.1) with the equilibrium equation equation (2.27) or the more general not-in-equilibrium form, equation (2.26). In the equilibrium comparison, which is the most direct, one sees that the conventional expression does not avert to the dependence of the energy on the number, volume and temperature, $\bar{E} = E(N, V, T)$. In this sense equation (2.1) may be criticized as being imprecise. One also sees that it does not avert to the fact that the entropy that appears explicitly, $S(\bar{E}, N, V)$, is the entropy for the isolated sub-system in the energy macrostate $\bar{E}$. In this sense it is misleading to call this 'the' entropy. In the same vein it would be better to explicitly recognize that the Helmholtz free energy is just minus the temperature times the maximum value of the constrained total entropy. Finally, comparing the conventional expression for the Helmholtz free energy, equation (2.1), to the more general constrained thermodynamic potential equation (2.26), one sees that the conventional expression for the Helmholtz free energy neither recognizes nor exhibits its variational nature with respect to energy.

One sometimes sees it asserted that the Helmholtz free energy is a minimum, as in the exhortation that is equation (2.2), which implies that it gives a variational principle. There are three things wrong with this assertion. First, the Helmholtz free energy is $\bar{F}(N, V, T)$, and this does not obey any variational principle. Second, the constrained thermodynamic potential $F(E \mid N, V, T)$ does obey a variational principle with respect to the sub-system energy, and it is essential that the variational parameter be stated and understood explicitly. This is not mere nit picking. It is not uncommon in the literature to see the Helmholtz free energy formulated for a particular problem and then for it to be minimized with respect to some parameter of interest. Unless that parameter is the energy, such a procedure is a mathematical and physical absurdity. Third, the variational principle for the constrained thermodynamic potential derives directly from the Second Law of Thermodynamics via its relationship to the total entropy. This essential physical basis and its further consequences are missed entirely by the simplistic incantation that the Helmholtz free energy should be minimized.

2.2.3 Derivatives of the Helmholtz free energy

A concrete example of the power of the constrained thermodynamic potential over the usual free energy expressions is in the ease with which it can be differentiated. Because it provides a variational principle for equilibrium, (i.e. it is a minimum with respect to variations in sub-system energy), the energy can effectively be regarded as a fixed independent variable rather than a dependent variable during differentiation.

For example, the temperature derivative of the Helmholtz free energy is

$$
\begin{aligned}
\frac{\partial \bar{F}(N, V, T)}{\partial T} &= \left(\frac{\partial F(\bar{E}(N, V, T) \mid N, V, T)}{\partial T} \right)_{N, V} \\
&= \frac{\partial F(E \mid N, V, T)}{\partial T} \Big|_{E=\overline{E}} + \frac{\partial F(E \mid N, V, T)}{\partial E} \Big|_{E=\overline{E}} \frac{\partial \overline{E}(N, V, T)}{\partial T} \\
&= \frac{\partial F(E \mid N, V, T)}{\partial T} \Big|_{E=\overline{E}} \\
&= - S_{\mathrm{s}}(\bar{E}, N, V).
\end{aligned}
$$

The result is that the temperature derivative of the Helmholtz free energy is the negative of the sub-system entropy in the equilibrium state. (2.30)

The third equality follows from the variational principle for the constrained thermodynamic potential, $\partial F(E \mid N, V, T)/\partial E|_{E=\bar{E}} = 0$. Hence even though the equilibrium sub-system energy is a dependent variable, $\partial \bar{E}(N, V, T)/\partial T \neq 0$, this dependence can be ignored because it is multiplied by a derivative that vanishes at equilibrium. In other words, because the constrained thermodynamic potential is optimised at equilibrium, differentiating the Helmholtz free energy is the same as differentiating the constrained thermodynamic potential *holding $\bar{E}$ fixed*. This is a general feature of constrained thermodynamic potentials that can be exploited in all similar derivatives.

Now let us differentiate the Helmholtz free energy with respect to volume and number. One can use the corresponding partial derivatives of the isolated system entropy, equations (2.11) and (2.12), since one can again hold $\bar{E}$ fixed. The volume derivative gives the pressure,

$$
\begin{aligned}
\left(\frac{\partial \bar{F}(N, V, T)}{\partial V} \right)_{T,N} &= \left(\frac{\partial F(\bar{E} \mid N, V, T)}{\partial V} \right)_{\bar{E},T,N} \\
&= -T \left(\frac{\partial S_{\mathrm{s}}(\bar{E}, N, V)}{\partial V} \right)_{\bar{E},N} \\
&= -\bar{p},
\end{aligned}
$$

(2.31)

and the number derivative gives the chemical potential,

$$
\begin{aligned}
\left(\frac{\partial \bar{F}(N, V, T)}{\partial N} \right)_{T,V} &= \left(\frac{\partial F(\bar{E} \mid N, V, T)}{\partial N} \right)_{\bar{E},T,V} \\
&= -T \left(\frac{\partial S_{\mathrm{s}}(\bar{E}, N, V)}{\partial N} \right)_{\bar{E},V} \\
&= \bar{\mu}.
\end{aligned}
$$

(2.32)

It will shortly be shown that the equilibrium state is unique. Hence the pressure that appears here $\bar{p}(N, V, T)$ is equal to that of the isolated sub-system in the equilibrium energy state, $p(\bar{E}(N, V, T), N, V)$. This is necessary for thermodynamics to be internally consistent. One must have the volume derivative of the Helmholtz free energy for a sub-system in contact with a heat reservoir giving the same pressure as an isolated system with the equilibrium energy. Similar comments apply to the chemical potential, $\bar{\mu}(N, V, T) = \mu(\bar{E}(N, V, T), N, V)$ and to the sub-system entropy $\bar{S}_{\mathrm{s}}(N, V, T) = S_{\mathrm{s}}(\bar{E}(N, V, T), N, V)$. The overline here denotes the fact that these are properties of the sub-system in thermal equilibrium.

These results for the partial derivatives give the total differential of the Helmholtz free energy,

$$\mathrm{d}\bar{F}(N,V,T) = -\bar{S}_{\mathrm{s}}(N,V,T)\,\mathrm{d}T - \bar{p}(N,V,T)\,\mathrm{d}V + \bar{\mu}(N,V,T)\,\mathrm{d}N.$$

(2.33)

Finally, there is one more example that exploits the variational nature of the constrained thermodynamic potential for a sub-system in contact with a heat reservoir. Dividing both sides of the Helmholtz free energy by temperature, $\bar{F}/T = \bar{E}/T - \bar{S}_{\mathrm{s}}(\bar{E},N,V)$, and differentiating with respect to T^{-1} one obtains

$$\left(\frac{\partial(\bar{F}(N,V,T)/T)}{\partial(1/T)}\right)_{N,V} = \bar{E}.$$

(2.34)

Again this follows because $\bar{E}$ and hence $S_{\mathrm{s}}(\bar{E}(N,V,T),N,V)$ can be held constant.

2.2.4 Concavity, uniqueness, and thermal stability

2.2.4.1 Concavity of the entropy

It is now shown that the entropy of the isolated system is a concave function of its arguments. The derivation is based on the extensivity of the entropy and the fact that the constrained total entropy is a maximum at equilibrium.

In figure 2.4 are shown two systems each comprising two sub-systems that are identical except for energy. In the left-hand system the sub-systems are isolated from each other and have the same volume *V*, number *N*, but different energies E_1 and E_2. The entropies of the sub-systems are $S_1 = S(E_1, N, V)$ and $S_2 = S(E_2, N, V)$. From the linear additive nature of the entropy, the total entropy with the insulating partition in place is $S_{\mathrm{total}} = S_1 + S_2$.
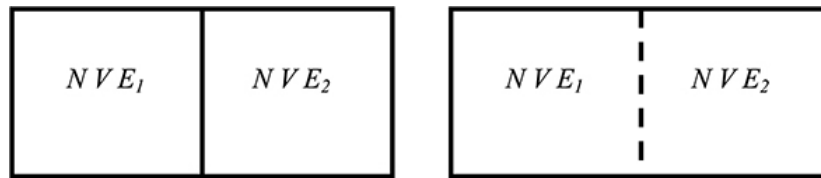
Figure 2.4. Two isolated systems, each composed of two sub-systems identical except for energy. In the system on the left, the sub-systems are separated by an insulating wall that prevents energy exchange. In the system on the right, the sub-systems are separated by a conducting wall that allows energy exchange.

If the insulating partition is replaced by a conducting partition, as in the right-hand diagram, energy irreversibly flows from the sub-system with more energy to that with less, as is obvious on symmetry grounds. The total entropy at the end of this process must be greater than at the beginning,

$$S_{\mathrm{total}}(E_1 + E_2, 2N, 2V) \geqslant S(E_1, N, V) + S(E_2, N, V).$$

This is of course demanded by the Second Law of Thermodynamics: energy flows in the direction of entropy increase. At the microscopic level, the number of microstates must be greater when the constraint is relaxed. This is a strict inequality unless initially $E_1 = E_2$.

The equilibrium state is the one in which each sub-system has the same energy, namely $(E_1 + E_2)/2$, and this is the state of maximum entropy. Once equilibrated, there is no further nett energy flow, and the removal and insertion of the partition becomes a reversible process. In this case the entropy does not change and one must have

$$S_{\text{total}}(E_1 + E_2, 2N, 2V) = 2S\left(\frac{E_1 + E_2}{2}, N, V\right).$$

(2.36)

This expression assumes that fluctuations about the equilibrium state make relatively negligible contribution to the total entropy on the left-hand side. This result is the $\lambda = 2$ case that was derived above in the discussion of extensivity; see section 2.1.1 and figure 2.2.

These two equations combined yield

$$S\left(\frac{E_1 + E_2}{2}, N, V\right) > \frac{1}{2}\left[S(E_1, N, V) + S(E_2, N, V,)\right].$$

(2.37)

excluding the case $E_1 = E_2$. This shows that entropy is a concave function of energy, since any chord to the entropy curve lies below the curve. A Taylor expansion as $E_1 \to E_2$ shows that this is equivalent to

$$\left(\frac{\partial^2 S(E, N, V)}{\partial E^2}\right)_{N,V} < 0.$$

(2.38)

Again, a negative second derivative defines a concave function. Analogous steps can be carried out individually for volume and for particle number. One can conclude in general that the entropy is a concave function of its arguments.

It is not much harder to derive the result for simultaneous variations. For example, considering two isolated systems that are identical apart from their energies and volumes, and allowing exchange of these, one can readily show that

$$S_{\text{total}}\left(\frac{E_1 + E_2}{2}, N, \frac{V_1 + V_2}{2}\right) > \frac{1}{2}\left[S(E_1, N, V_1) + S(E_2, N, V_2)\right].$$

(2.39)

Performing a Taylor expansion of the right-hand side about the equilibrium energy and volume, the zeroth and linear terms cancel, which leaves

$$S_{EE}\,\mathrm{d}E^2 + 2S_{EV}\,\mathrm{d}E\,\mathrm{d}V + S_{VV}\,\mathrm{d}V^2 < 0,$$

(2.40)

where the subscripts denote partial derivatives, and $\mathrm{d}E$ and $\mathrm{d}V$ are arbitrary. Again a negative Jacobian defines a concave function of several variables. This result is consistent with equation (2.19). We shall use this result below.

2.2.4.2 Uniqueness of the thermodynamic state

In thermodynamics it is sometimes useful to swap dependent and independent variables. For example, to date we have described an isolated system by its entropy, which is a function of the extensive variables energy, number and volume, $S(E, N, V)$. For mathematical reasons it is sometimes convenient to instead take energy as the dependent variable, $E(S, N, V)$.

In general one can interchange variables, such as

$$x_1 = x(y_1) \Leftrightarrow y_1 = y(x_1),$$

(2.41)

if the relationship between the conjugate variables is one-to-one. This is the same as saying that the function is invertible,

$$x = f(y) \Leftrightarrow y = f^{-1}(x).$$

(2.42)

The condition for a continuous function to be invertible is that its derivative must be finite and nonzero on the interior of its domain of definition. Since entropy is a strictly concave function of its arguments, it is indeed an invertible.

First the energy–entropy swap. (We shall assume that entropy is a continuous function of energy.) As we have seen, entropy is strictly concave, $S\prime\prime < 0$. Therefore, it cannot have more than one turning point, which we can assume occurs at E_1, with $S\prime(E_1) = 0$. This turning point, if it exists, must be a maximum. Obviously entropy is monotonic increasing, $S\prime(E) > 0$, for $E < E_1$, which means that $T(E) > 0$. It is monotonic decreasing, $S\prime(E) < 0$, for $E > E_1$, which is to say that $T(E) < 0$. Since systems with negative temperature are unstable, we shall only consider systems with positive temperatures here and throughout, $E < E_1$. Denote the lowest energy or ground state by E_0. Because the entropy is strictly concave, its gradient may possibly be infinite here, but nowhere else. An infinite gradient, $S\prime(E_0) = \infty$, corresponds to $T = 0$, which is absolute zero. These considerations show that on the domain (E_0, E_1) there is a one-to-one relationship between the entropy and the energy. Hence the entropy function is invertible and one can write $S(E, N, V)$ or $E(S, N, V)$ as convenient.

Now the energy–temperature swap. On the physical domain (E_0, E_1), the concavity of the entropy, $S_{EE} < 0$, means that $\partial(1/T) / \partial E < 0$, or $\partial T / \partial E > 0$. Since temperature is a function of energy, $T(E)$, its energy derivative cannot be infinite (otherwise there would be more than one temperature for a given energy). Hence the relationship between energy and temperature must be one-to-one, which is to say that it is invertible, $T_1 = T(E_1, N, V) \Leftrightarrow E_1 = E(T_1, N, V)$.

Entirely analogous arguments can be made for the other variables, namely pressure and volume can be swapped, as can number and chemical potential.

2.2.4.3 Thermal stability

For the present case of a sub-system in thermal contact with a heat reservoir, the constrained thermodynamic potential, $F(E \mid N, V, T) = E - TS(E, N, V)$, is a convex function of energy (because $S(E, N, V)$ is a concave function of energy, and the term E is linear). Similarly, it is also a convex function of volume and number, so that we can write generically $F\prime\prime = -TS\prime\prime > 0$. As discussed at length above, the

Second Law of Thermodynamics confers upon the constrained thermodynamic potential a variational principle that means that it is minimized by the equilibrium energy. Its minimum value is equal to the Helmholtz free energy.

As mentioned above, the Helmholtz free energy $\bar{F}(N,V,T) = F(\bar{E} \mid N,V,T)$ does not obey a variational principle. It is not a minimum with respect to number, volume, or temperature.

The thermal stability of matter may be deduced from the convexity of the constrained thermodynamic potential. Suppose that in a small region of the system there is a local increase in energy, with volume and particle number remaining constant. This increases the constrained thermodynamic potential for that region, considering it as a quasi-isolated sub-system. Equivalently, it decreases the local constrained total entropy. This is unfavorable, which is to say improbable, and the change is most likely to be counteracted by energy flowing back out of the region. This restores it to equilibrium. Conversely a local fluctuation to a lower energy is most likely followed by a countervailing back flow of energy.

It is the convexity of the constrained thermodynamic potential that makes matter thermally stable by damping such local fluctuations in the energy. Thermal stability does not arise from the Helmholtz free energy, but rather from the variational nature of the underlying constrained thermodynamic potential.

In the next section analogous optimization principles will be derived for volume and particle number reservoirs. It will be similarly shown that matter is mechanically stable to volume and particle number fluctuations.

2.3 Various reservoirs

The canonical equilibrium system that was just treated in detail may be called a constant temperature system, because the reservoir sets the temperature of the sub-system with which it can exchange energy. All that was required to obtain the constrained total entropy (equivalently, the constrained thermodynamic potential) for that system was to select the relevant linear additive, conserved variable, in this case energy, write the total entropy as the sum of that of the sub-system and that of the reservoir, and do a first order Taylor expansion of the latter, discarding the constant term.

In this section this procedure is repeated for several common reservoirs.

2.3.1 Constant pressure

An isobaric system is one in which the reservoir sets the pressure of the sub-system. The arrangement of the sub-system and reservoir remains the same as in figure 2.3, except that now the boundary is permeable to both energy and volume. This means that not only is the boundary made of conducting material, but also that it is flexible or moveable so that the volume of the sub-system can change at the expense of that of the reservoir. Thus the total energy $E_{\text{total}} = E_1 + E_2$ and total volume $V_{\text{total}} = V_1 + V_2$ are fixed, but not their partitioning between the two systems.

As the second system is an energy and volume reservoir (i.e. it is very much larger than the sub-system), the Taylor expansion of its entropy about the total energy and volume may be terminated at the linear term. The zeroth order term, $S_2(E_{\text{total}}, N_2, V_{\text{total}})$, is a constant independent of the sub-system that may be discarded. Hence the sub-system-dependent part of the total entropy $S_{\text{total}} = S_1 + S_2$ is (dropping the subscripts)

$$S_{\text{total}}(E, V \mid N, p, T) = S_{\text{s}}(E, N, V) - \frac{E}{T} - \frac{pV}{T}.$$

This uses the volume derivative of the entropy, equation (2.11), for the pressure p of the reservoir, and the energy derivative, equation (2.10), for the temperature T of the reservoir. The reservoir only enters through its temperature and pressure. The constrained total entropy is extensive.

This is the constrained total entropy for the sub-system with arbitrary (i.e. not in equilibrium) values of energy or volume. By the Second Law of Thermodynamics, a spontaneous change in the volume must increase the total constrained entropy, or

$$\begin{aligned} \Delta S_{\text{total}}(E, V \mid N, p, T) &= \frac{\partial S_{\text{total}}(E, V \mid N, p, T)}{\partial V} \Delta V \\ &= \left[\frac{p(E, N, V)}{T(E, N, V)} - \frac{p}{T} \right] \Delta V \\ &\geqslant 0. \end{aligned}$$

Suppose that the system is in thermal equilibrium, $T(E, N, V) = T$, but not in mechanical equilibrium, $p(E, N, V) \neq p$. In this case this equation says that when the internal sub-system pressure is greater than the external reservoir pressure, $p(E, N, V) > p$, the sub-system spontaneously grows, $\Delta V > 0$, and vice versa. This is of course what one would expect on physical grounds and it justifies identifying the mathematical object p with the physical object 'pressure'.

As usual, the equilibrium state is the one that maximizes the constrained total entropy, and is denoted $\bar{E}(N, p, T)$ and $\bar{V}(N, p, T)$. This variational procedure can be flipped into a minimization problem by defining the constrained thermodynamic potential as the negative of the temperature times the constrained total entropy,

$$\begin{aligned} G(E, V \mid N, p, T) &\equiv -T S_{\text{total}}(E, V \mid N, p, T) \\ &= E - T S_{\text{s}}(E, N, V) + pV \\ &= F(E \mid N, V, T) + pV. \end{aligned}$$

By design this constrained thermodynamic potential is a convex function of the sub-system volume and energy. It is appropriate for a sub-system with moveable conducting walls in contact with a temperature and pressure reservoir. In the event that thermal equilibrium is established, $E = \bar{E}(N, V, T)$, one can define $\tilde{G}(V \mid N, p, T) = \bar{F}(N, V, T) + pV$ as the constrained thermodynamic potential that determines volume fluctuations and equilibration.

Minimizing $G(E, V \mid N, p, T)$ with respect to energy yields $T(E, V, N) = T$, which is an implicit equation for the equilibrium sub-system energy for a given (constrained) volume, $\bar{E}(N, V, T)$. The minimization with respect to volume yields $p(E, V, N)/T(E, V, N) = p/T$, which is an implicit equation for the sub-system volume for a given sub-system energy and reservoir temperature and pressure, $\bar{V}(E \mid N, p, T)$. The simultaneous solution of these gives the equilibrium quantities, $\bar{E}(N, p, T)$ and $\bar{V}(N, p, T)$. The equilibrium sub-system volume also follows directly

from the minimization of the thermally equilibrated constrained thermodynamic potential, $\partial \tilde{G}(V \mid N, p, T)/\partial V = 0$, or

$$\frac{\partial \bar{F}(N, V, T)}{\partial V}\Big|_{V=\bar{V}} = -p.$$

(2.46)

The minimum value of the constrained thermodynamic potential defines Gibbs free energy,

$$\bar{G}(N, p, T) \equiv G(\bar{V}, \bar{E} \mid N, T, p)$$
$$= \tilde{G}(\bar{V} \mid N, T, p)$$
$$= F(N, \bar{V}, T) + p\bar{V}.$$

(2.47)

Since the Gibbs free energy must be extensive, and since it is a function of only one extensive variable, N, then it must be proportional to it, $\bar{G} \propto N$. The proportionality constant will be derived shortly.

Again one can exploit the power of the constrained thermodynamic potential by invoking its variational nature when taking derivatives. That is, differentiating the Gibbs free energy is the same as differentiating the constrained thermodynamic potential while holding $V = \bar{V}$ and $E = \bar{E}$ fixed. Accordingly, the pressure derivative gives the equilibrium sub-system volume,

$$\left(\frac{\partial \bar{G}(N, p, T)}{\partial p}\right)_{N,T} = \bar{V},$$

(2.48)

the number derivative gives the equilibrium sub-system chemical potential,

$$\left(\frac{\partial \bar{G}(N, p, T)}{\partial N}\right)_{p,T} = \left(\frac{\partial \bar{F}(N, \bar{V}, T)}{\partial N}\right)_{\bar{V},T} = \bar{\mu},$$

(2.49)

and the temperature derivative gives the negative of the equilibrium sub-system entropy,

$$\left(\frac{\partial \bar{G}(N, p, T)}{\partial T}\right)_{N,p} = \left(\frac{\partial \bar{F}(N, \bar{V}, T)}{\partial T}\right)_{N,V} = -\bar{S}_{\text{s}}.$$

(2.50)

Hence the total differential of the Gibbs free energy is

$$d\bar{G} = \bar{V}dp + \bar{\mu}dN - \bar{S}_{\text{s}}dT.$$

(2.51)

The second of the above derivatives combined with the above extensivity argument shows that

$$\bar{G}(N, p, T) = \bar{\mu}(p, T)N.$$

(2.52)

Finally, dividing both sides of the definition by *T*, one also has that

$$\left(\frac{\partial(\bar{G}/T)}{\partial(1/T)} \right)_{N,p} = \bar{E} + p\bar{V}.$$

(2.53)

The quantity on the right-hand side is called the enthalpy, and it will recur below.

2.3.2 Constant chemical potential

Now consider a reservoir that can exchange particles and energy with the sub-system of interest. This set-up is called an open system or, better, an open sub-system, or a grand canonical system. It can also be called a constant chemical potential system, as this and the temperature are fixed by the reservoir.

As usual, the total entropy is the sum of that of the sub-system and reservoir, each considered as isolated and in the designated macrostate, $S_{\text{total}} = S_1 + S_2$. The total energy $E_{\text{total}} = E_1 + E_2$ and the total particle number $N = N_1 + N_2$ are fixed, as well as the individual volumes, V_1 and V_2. Again the relative size of the reservoir allows all terms in the Taylor expansion of the reservoir entropy to be neglected except for the linear one. (The zeroth term can again be dropped because it is a constant that is independent of the sub-system.) Hence dropping the subscripts, the constrained total entropy is

$$S_{\text{total}}(N, E \mid \mu, V, T) = S_{\text{s}}(E, V, N) - \frac{E}{T} + \frac{\mu}{T}N.$$

(2.54)

This gives the total entropy for a system when the sub-system of fixed volume *V* has energy *E* and particle number *N*, which are not necessarily the equilibrium values, while in thermal and diffusive contact with a reservoir of temperature *T* and chemical potential μ.

The constrained total entropy has derivatives

$$\frac{\partial S_{\text{total}}(N, E \mid \mu, V, T)}{\partial N} = \frac{-\mu(E, N, V)}{T(E, N, V)} + \frac{\mu}{T},$$

(2.55)

and

$$\frac{\partial S_{\text{total}}(N, E \mid \mu, V, T)}{\partial E} = \frac{1}{T(E, N, V)} - \frac{1}{T}.$$

(2.56)

The first term on the right-hand side in each case is the derivative of the isolated sub-system entropy, and they give the chemical potential and the temperature of the sub-system in the specified macrostate. The maximum of the constrained total entropy of course corresponds to the vanishing of its derivatives. The implicit equations for the equilibrium energy $\bar{E}(\mu, V, T)$ and the equilibrium particle number $\bar{N}(\mu, V, T)$ that

this gives are $T(\bar{E}, \bar{N}, V) = T$ and $\mu(\bar{E}, \bar{N}, V) = \mu$. Obviously, equilibrium corresponds to temperature and chemical potential equality between the sub-system and the reservoir.

Spontaneous changes in the sub-system particle number ΔN must lead to an increase in the total entropy

$$
\begin{aligned}
\Delta S_{\text{total}}(N, E \mid \mu, V, T) &= \frac{\partial S_{\text{total}}(N, E \mid \mu, V, T)}{\partial N} \Delta N \\
&= \left[\frac{-\mu(E, N, V)}{T(E, N, V)} + \frac{\mu}{T} \right] \Delta N \\
&\geqslant 0.
\end{aligned}
$$

(2.57)

Assuming thermal equilibration, $T(\bar{E}, N, V) = T$, when the chemical potential of the sub-system is greater than that of the reservoir, $\mu(\bar{E}(N, V, T), N, V) > \mu$, this says that $\Delta N < 0$, which is to say particles spontaneously flow from the sub-system to the reservoir.

In words, this says that particles move down the chemical potential gradient. This is analogous to what happens in a thermal system: energy moves down a temperature gradient. It is the opposite of what happens in an isobaric system, where volume moves up a pressure gradient. The similarities and differences in the three cases can be accounted for by whether or not the definition of the field variable in terms of an entropy derivative involves a minus sign, or whether or not the definition refers to the reciprocal of the field variable.

The corresponding constrained thermodynamic potential is the negative of the total entropy times the temperature,

$$
\begin{aligned}
\Omega(E, N \mid \mu, V, T) &\equiv -T S_{\text{total}}(E, N \mid \mu, V, T) \\
&= E - T S_{\text{s}}(E, N, V) - \mu N \\
&= F(E \mid N, V, T) - \mu N.
\end{aligned}
$$

(2.58)

This is a convex function of N and of E that is minimized by their equilibrium values. Thus this is the minimization variational principle for a sub-system able to exchange energy and particles with a reservoir of chemical potential μ and temperature T.

If thermal equilibration is faster than diffusive equilibration, then one can define $\tilde{\Omega}(N \mid \mu, V, T) \equiv \Omega(\bar{E}(N, V, T), N \mid \mu, V, T) = \bar{F}(N, V, T) - \mu N$. This involves the Helmholtz free energy, and it provides a variational principle for a sub-system at the same temperature as the reservoir, but not in diffusive equilibrium with it.

The concavity of the isolated sub-system entropy means that the constrained thermodynamic potential is a convex function of particle number and energy, which can be figuratively denoted $\Omega'' = -T S_{\text{s}}'' > 0$. The minimum value of the constrained thermodynamic potential of this open sub-system is the equilibrium free energy, which is this case is called the grand potential. It is

$$\bar{\Omega}(\mu, V, T) \equiv \Omega(\bar{E}, \bar{N} \mid \mu, V, T)$$
$$= \bar{F}(\bar{N}, V, T) - \mu \bar{N}.$$

The properties of the variational procedure allow the otherwise dependent variables $\bar{E}$ and $\bar{N}$ to be held fixed during differentiation of the grand potential. Hence one has

$$\left(\frac{\partial \bar{\Omega}(\mu, V, T)}{\partial \mu} \right)_{V,T} = -\bar{N},$$

$$\left(\frac{\partial \bar{\Omega}(\mu, V, T)}{\partial V} \right)_{\mu,T} = \left(\frac{\partial \bar{F}(\bar{N}, V, T)}{\partial V} \right)_{\bar{N},T} = -\bar{p},$$

and

$$\left(\frac{\partial \bar{\Omega}(\mu, V, T)}{\partial T} \right)_{\mu,V} = \left(\frac{\partial \bar{F}(\bar{N}, V, T)}{\partial T} \right)_{\bar{N},V} = -\bar{S}_{\mathrm{s}}.$$

This last quantity is the entropy of the isolated sub-system with the equilibrium energy and particle number, $\bar{S}_{\mathrm{s}}(\mu, V, T) \equiv S_{\mathrm{s}}(\bar{E}(\mu, V, T), \bar{N}(\mu, V, T), V)$. It follows that the total differential of the grand potential is

$$\mathrm{d}\bar{\Omega} = -\bar{N}\,\mathrm{d}\mu - \bar{p}\,\mathrm{d}V - \bar{S}_{\mathrm{s}}\,\mathrm{d}T.$$

One also has

$$\left(\frac{\partial (\bar{\Omega}/T)}{\partial (1/T)} \right)_{\mu,V} = \bar{E} - \mu \bar{N}.$$

Finally, because the grand potential must be extensive, and because volume is the only one of its arguments that is extensive, the grand potential must be proportional to the volume. From the above volume derivative one concludes that

$$\bar{\Omega}(\mu, V, T) = -\bar{p}V.$$

2.3.2.1 Multi-component system

It is not unusual to have several particle species present. Let N_α be the number of particles of type α in the sub-system. The chemical potential for this species is defined as

$$\mu_\alpha \equiv -T\left(\frac{\partial S_{\rm s}}{\partial N_\alpha}\right)_{E,V,N_{\gamma\neq\alpha}}.$$

(2.66)

The formulae given above for a single species are changed to include the sum over species. For example, for m different types of particles, the total differential of the isolated system entropy becomes

$$\mathrm{d}S_{\rm s}(E,\underline{N},V) = \frac{1}{T}\,\mathrm{d}E + \frac{p}{T}\,\mathrm{d}V - \frac{1}{T}\sum_{\alpha=1}^{m}\mu_\alpha\,\mathrm{d}N_\alpha.$$

(2.67)

It is often convenient to use vector notation, and to replace the sum by a scalar product. For example, the constrained thermodynamic potential may be written as

$$\Omega(E,\underline{N}\mid\underline{\mu},V,T) = E - TS_{\rm s}(E,\underline{N},V) - \underline{\mu}\cdot\underline{N}.$$

(2.68)

The diffusive equilibrium condition is obviously chemical potential equality between the sub-system and the reservoir for each species, $\bar{\mu}_\alpha \equiv \mu_\alpha(\bar{E},\overline{\underline{N}},V) = \mu_\alpha$.

2.3.3 Constant enthalpy

We now consider an isolated sub-system with a moveable wall or piston, figure 2.5. The number of particles in the sub-system N_1 is fixed. Volume can be exchanged with the surrounding reservoir, $\mathrm{d}V_1 = -\mathrm{d}V_2$. Even though the walls are insulated (adiabatic, adiathermal, no heat flow), energy can still exchange between the sub-system and the reservoir via so-called *pV*-work. The total energy is fixed, $E_{\rm total} = E_1 + E_2$, which means that $\mathrm{d}E_1 = -\mathrm{d}E_2$.

Figure 2.5. An isolated sub-system 1 with moveable piston.
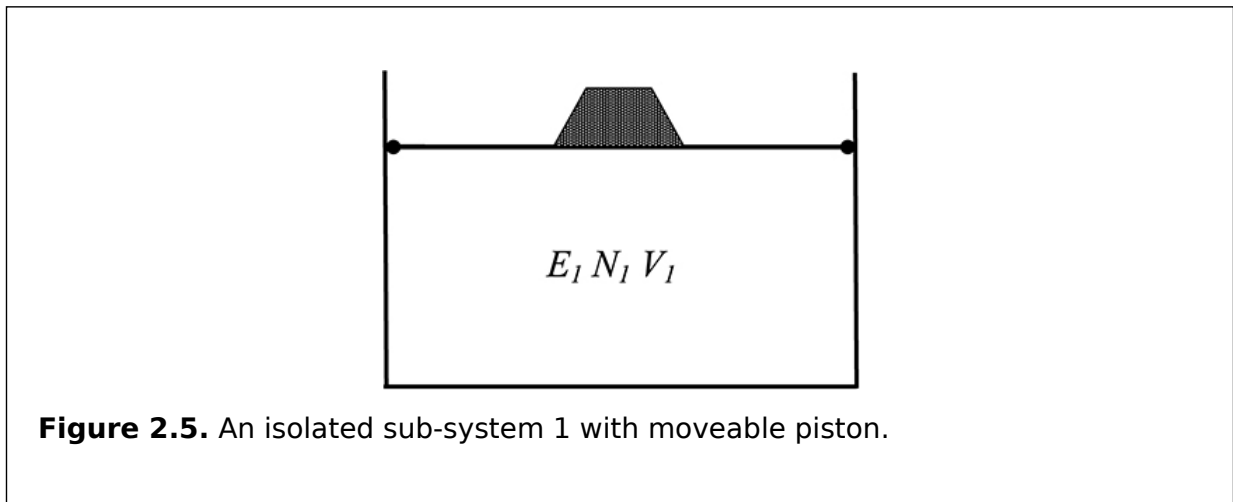
The energy of the sub-system changes as its volume changes, $E_1(V_1)$. As the wall or piston moves, the change in energy of the reservoir is linearly proportional to the change in sub-system volume, $\mathrm{d}E_2 = p_2\mathrm{d}V_1$. (This can be seen in mechanical terms: if the piston has area A_1 and mass M, and if the acceleration due to gravity is g, then the

constant external pressure is $p_2 = Mg/A_1$, and the change in energy is proportional to the change in height, $dE_2 = Mg dh_1$.) In any event, the external pressure p_2 is taken to be constant. With this it may be seen that changes in the volume and energy of the sub-system at constant particle number are related by

$$\frac{dE_1}{dV_1} = -p_2, \quad \text{or} \quad E_1 + p_2 V_1 = \text{const.}$$

(2.69)

The quantity that is constant during volume changes without heat flow is like the enthalpy,

$$H = E_1 + p_2 V_1.$$

(2.70)

(Usually the difference between the sub-system pressure and the reservoir pressure is ignored.) The enthalpy previously appeared as essentially the inverse temperature derivative of the Gibbs free energy, equation (2.53). The fact that this is constant in the present system of no heat flow is consistent with a form of the First Law of Thermodynamics, namely pV-work at constant pressure changes the heat of the system as $dQ = dE + p\,dV = d(E + pV)$, which is just the change in enthalpy. In the present case this is zero because the system has been defined to have no heat flow.

In a reversible change, the internal pressure balances the external pressure, $p_1(E_1, N_1, V_1) = p_2$. In this case the enthalpy of the sub-system is constant, $dH_1 = dE_1 + p_1 dV_1 = 0$. More generally, $p_1 \neq p_2$, and it is the enthalpy-like quantity $E_1 + p_2 V_1$ that is constant, where the constant externally applied pressure is used, not the internal pressure of the sub-system.

The total entropy is that of the sub-system alone, since the entropy of the reservoir is constant

$$dS_2(E_2, N_2, V_2) = \frac{1}{T_2} dE_2 + \frac{p_2}{T_2} dV_2$$

$$= \frac{1}{T_2}[p_2 dV_1 - p_2 dV_1]$$

$$= 0.$$

(2.71)

One could have guessed that this would be the case because the external pressure could have been made mechanical in origin. Hence the total entropy is $S_{\text{total}}(E, V \mid N, p) = S_1(E, V, N)$, with $E + pV = \text{const.}$, where the subscripts have been dropped. Hence the change in total entropy due to a change in sub-system volume is

$$\Delta S_{\text{total}} = \frac{\partial S_1}{\partial E} \frac{dE}{dV} \Delta V + \frac{\partial S_1}{\partial V} \Delta V$$

$$= \frac{p(E, N, V) - p}{T(E, N, V)} \Delta V.$$

(2.72)

For the total entropy to increase, the volume of the sub-system has to increase when its pressure is greater than the external pressure, and the volume has to decrease when the internal pressure is less than the external pressure. This was also found to be the case for the isobaric system, which allowed heat flow, equation (2.44). Equilibrium, of course, corresponds to pressure equality.

Since the enthalpy $H = E + pV$ is constant, it may be taken as one of the independent variables for the insulated, variable volume system. The constrained thermodynamic potential may be defined as

$$\Phi(V \mid H, N, p; T) = -TS_\mathrm{s}(E, N, V), \quad E \equiv H - pV.$$

(2.73)

The temperature here is arbitrary, as it is introduced solely to give the potential the dimensions of energy. It enters as a trivial multiplicative factor whose value has no physical consequences. By construction Φ is a convex function of volume minimized at the equilibrium volume $\bar{V}(H, N, p)$. The volume derivative is

$$\frac{1}{T}\frac{\partial \Phi(V \mid H, N, p; T)}{\partial V} = -\frac{\partial S_\mathrm{s}(E, N, V)}{\partial E}\frac{\partial (H - pV)}{\partial V} - \frac{\partial S_\mathrm{s}(E, N, V)}{\partial V}$$

$$= \frac{p - p(E, N, V)}{T(E, N, V)}, \quad E \equiv H - pV.$$

(2.74)

Again, this vanishes at equilibrium, when the internal pressure of the sub-system equals the external pressure of the reservoir, $p(\bar{E}, \bar{V}, N) = p$, where $\bar{E} = H - p\bar{V}$.

As usual, the thermodynamic potential of this system is the minimum value of the constrained thermodynamic potential, $\bar{\Phi}(H, N, p; T) = \Phi(\bar{V} \mid H, N, p; T)$. And again as usual, the variational nature of Φ, means that $\bar{\Phi}$ can be differentiated at fixed $\bar{V}$. Hence one has

$$\frac{1}{T}\left(\frac{\partial \bar{\Phi}}{\partial N}\right)_{H,p,T} = \frac{1}{T}\left(\frac{\partial \Phi}{\partial N}\right)_{H,p,T,\bar{V}} = \frac{\bar{\mu}}{\bar{T}},$$

(2.75)

$$\frac{1}{T}\left(\frac{\partial \bar{\Phi}}{\partial H}\right)_{N,p,T} = \frac{1}{T}\left(\frac{\partial \Phi}{\partial H}\right)_{N,p,T,\bar{V}} = \frac{-1}{\bar{T}},$$

(2.76)

and

$$\frac{1}{T}\left(\frac{\partial \bar{\Phi}}{\partial p}\right)_{H,N,T} = \frac{1}{T}\left(\frac{\partial \Phi}{\partial p}\right)_{H,N,T,\bar{V}} = \frac{\bar{V}}{\bar{T}}.$$

(2.77)

Together with the trivial result $\partial \bar{\Phi}/\partial T = \bar{\Phi}/T$, these give the total differential,

$$\mathrm{d}\bar{\Phi} = -\frac{T}{\bar{T}}\,\mathrm{d}H + \frac{\bar{V}T}{\bar{T}}\,\mathrm{d}p + \frac{\bar{\mu}T}{\bar{T}}\,\mathrm{d}N + \frac{\bar{\Phi}}{T}\,\mathrm{d}T.$$

Finally, it is worth mentioning that the change in internal energy was taken above to be equal and opposite to that of the reservoir, $dE_1 = -dE_2 = -p_2 dV_1$. However, since no heat flows, the change in energy of the sub-system is just the work done by it on the piston, $dE_1 = -p_1 dV_1$. There is an obvious contradiction between these two in the case that $p_1 \neq p_2$.

The problem can be resolved by noting that the piston acquires kinetic energy as it accelerates due to the initial force imbalance. This has been neglected in the above analysis. Also neglected has been any heat created by friction or by viscous dissipation. If the viscous dissipation of the reservoir is comparable to that of the sub-system, one cannot say how the kinetic energy of the piston is divided between the two when it dies out. Because of the adiathermal walls, the arbitrary internal heating does not equilibrate between them. Although equilibrium still corresponds to pressure equality, one does not know the final energy, temperature, or volume of the sub-system. This indeterminacy of linked isenthalpic systems has been noted by Callan (1960, appendix C).

2.3.4 Constant entropy

So far physical reservoirs have been considered. These allow the exchange of energy, volume, and particles, which are the independent variables of the sub-system entropy. We now wish to consider a case in which the sub-system entropy itself is an independently specified variable. This is done more to illustrate the mathematical procedures than for any application to a real physical system.

Suppose that the sub-system can exchange energy and volume with a reservoir while keeping the sub-system entropy constant. In this case we can regard the volume as independent and write the energy as $E(S, V, N)$. As usual, the constrained total entropy is

$$S_{\text{total}}(V \mid S, p, N; T) = S - \frac{E(S, V, N)}{T} - \frac{pV}{T},$$

where T and p are the temperature and pressure of the reservoir, respectively. Only the quantities that depend on the sub-system are kept here. Again as usual, the constrained thermodynamic potential is

$$H^*(V \mid S, p, N; T) = -TS_{\text{total}}(V \mid S, p, N; T)$$
$$= E(S, V, N) + pV - TS.$$

One can discard the final constant term since we are interested only in variations of the potential with the constrained volume at constant sub-system entropy,

$$H(V \mid S, p, N) = E(S, V, N) + pV.$$

This is just the enthalpy, and it describes volume fluctuations of an isentropic sub-system.

Now for an isolated system at constant entropy and number,

$$0 = dS(E, N, V) = \frac{1}{T} dE + \frac{p}{T} dV.$$

Hence $\partial E(S,V,N)/\partial V = -p(S,V,N)$. It follows that the volume derivative of the constrained thermodynamic potential is

$$\left(\frac{\partial H}{\partial V}\right)_{S,p,N} = -p(S,V,N) + p.$$

(2.8 3)

As expected, the equilibrium volume, $\bar{V}(S,p,N)$, equalizes the sub-system and reservoir pressures. Being derived from the total entropy, the constrained thermodynamic potential is a convex function of volume. Hence the extremum is a minimum,

$$\left(\frac{\partial^2 H}{\partial V^2}\right)_{S,p,N} = \left(\frac{\partial^2 E}{\partial V^2}\right)_{S,N} = -\left(\frac{\partial p(S,V,N)}{\partial V}\right)_{S,N} > 0.$$

(2.8 4)

The equilibrium enthalpy is the minimum value of the constrained thermodynamic potential,

$$\bar{H}(S,p,N) = E(S,\bar{V},N) + p\bar{V}.$$

(2.8 5)

This is the equilibrium thermodynamic potential for this isentropic system. Holding $\bar{V}$ fixed as per the variational principle, one has

$$\left(\frac{\partial \bar{H}(S,p,N)}{\partial p}\right)_{N,S} = \bar{V},$$

(2.8 6)

$$\left(\frac{\partial \bar{H}(S,p,N)}{\partial N}\right)_{p,S} = \left(\frac{\partial E(S,\bar{V},N)}{\partial N}\right)_{\bar{V},S} = \bar{\mu},$$

(2.8 7)

and

$$\left(\frac{\partial \bar{H}(S,p,N)}{\partial S}\right)_{N,p} = \left(\frac{\partial E(S,\bar{V},N)}{\partial S}\right)_{N,\bar{V}} = \bar{T}.$$

(2.8 8)

These may be summarised by the total differential,

$$d\bar{H} = \bar{V}dp + \bar{\mu}dN + \bar{T}dS.$$

(2.8 9)

In the case of the full thermodynamic potential, $\bar{H}^*(S,p,N;T)$, the results remain the same, except that the final one is replaced by

$$\left(\frac{\partial \bar{H}^{*}(S, p, N; T)}{\partial S}\right)_{N, p; T} = \bar{T} - T.$$

(2.90)

Note that there is no requirement for the temperature of the sub-system to equal that of the reservoir. In addition one has the entropy itself

$$\left(\frac{\partial \bar{H}^{*}(S, p, N; T)}{\partial T}\right)_{S, p, N} = -S,$$

(2.91)

and the equilibrium enthalpy

$$\left(\frac{\partial (\bar{H}^{*}(S, p, N; T)/T)}{\partial T^{-1}}\right)_{S, p, N} = -(\bar{E} + p\bar{V}).$$

(2.92)

The results for the various reservoirs are summarized in table 2.2.

Table 2.2. Summary of the common thermodynamic systems (Attard 2002).

System	Constrained potential parameters	Equilibrium potential differential
Isolated	—-	Entropy, S
	N, V, E	$T dS = dE + p\, dV - \mu\, dN$
Isothermal	$F = E - TS_{\mathrm{s}}(E, N, V)$	Helmholtz, $\bar{F}$
	$E \mid N, V, T$	$d\bar{F} = -\bar{S}_{\mathrm{s}}\, dT - \bar{p}\, dV + \bar{\mu}\, dN$
Open	$\Omega = E - TS_{\mathrm{s}}(E, N, V) - \mu N$	Grand, $\bar{\Omega}$
	$N, E \mid \mu, V, T$	$d\bar{\Omega} = -\bar{S}_{\mathrm{s}}\, dT - \bar{p}\, dV - \bar{N}\, d\mu$
Isobaric	$G = E - TS_{\mathrm{s}}(E, N, V) + pV$	Gibbs, $\bar{G}$
	$E, V \mid N, p, T$	$d\bar{G} = -\bar{S}_{\mathrm{s}}\, dT + \bar{V}\, dp + \bar{\mu}\, dN$

System	Constrained potential parameters	Equilibrium potential differential
Isenthalpic	$\Phi = -TS_s(H - pV, N, V)$	$\bar{\Phi}$
	$V \mid N, H, p; T$	$\bar{T}\,\mathrm{d}\bar{\Phi}/T = -\mathrm{d}H + \bar{V}\,\mathrm{d}p + \bar{\mu}\,\mathrm{d}N$
		$+\bar{\Phi}\,\bar{T}\,\mathrm{d}T/T^2$
Isentropic	$H = E(S, V, N) + pV$	Enthalpy, $\bar{H}$
	$V \mid N, S, p$	$\mathrm{d}\bar{H} = \bar{V}\,\mathrm{d}p + \bar{\mu}\,\mathrm{d}N + \bar{T}\,\mathrm{d}S$

2.4 Thermodynamic second derivatives

The second derivatives of the equilibrium thermodynamic potentials provide measurable physical parameters such as the heat capacity, compressibility, thermal expansivity etc. These also determine the probability of the statistical fluctuations about equilibrium, as will be shown in section 2.5.

The thermodynamic systems analyzed above each have a characteristic set of dependent and independent variables. However, because there is a one-to-one mapping between the equilibrium values of the dependent variables and the independent variables, which is to say that the thermodynamic state is unique, one is free to choose any set of three variables as the independent ones. To simplify the notation in this section, equilibrium quantities will not be over-lined, and the dependent and independent variables should be gleaned from the context.

2.4.1 Concavity of the thermodynamic potentials

The thermodynamic reservoir formalism is based on the exchange between the sub-system and the reservoir of one or more conserved, linear additive quantities. The constrained thermodynamic potential is convex with respect to these constrained extensive variables, as follows from the fact that it is the negative of the constrained total entropy. The equilibrium thermodynamic potential, which is the minimum value of the constrained thermodynamic potential, effectively replaces as independent variables these exchangeable extensive variables of the sub-system with the intensive variables of the reservoir that are their thermodynamic conjugates. The equilibrium free energy or thermodynamic potential is concave with respect to these conjugate intensive variables, as is now shown.

Let X be the conserved extensive variable and let the conjugate intensive (field) variable be

$$x = -\left(\frac{\partial E}{\partial X}\right)_S,$$

(2.93)

or

$$x = T\frac{\partial S}{\partial X}.$$

Typically X is E, N, or V, in which case x is 1, $-\mu$, or p, respectively. The variable X could also be some additional extensive, exchangeable system parameter. The part of the total entropy that depends on the sub-system for a reservoir at x and T, with X being exchanged, is

$$S_{\text{total}}(X \mid x, T) = S_{\text{s}}(X) - \frac{xX}{T}.$$

Now as has already been shown, the total entropy is concave with respect to the extensive variable,

$$\left(\frac{\partial^2 S_{\text{total}}(X \mid x, T)}{\partial X^2} \right)_x < 0.$$

The equilibrium value of X is $\bar{X}(x, T)$, and this gives the maximum total entropy, $\bar{S}_{\text{total}}(x, T) = S_{\text{total}}(\bar{X}(x, T) \mid x, T)$. This equilibrium value of the total entropy is a convex function of the reservoir field x/T, as is now proven.

One has

$$\frac{\mathrm{d}^2 S_{\text{total}}(\bar{X} \mid x, T)}{\mathrm{d}(x/T)^2} = \frac{\mathrm{d}}{\mathrm{d}(x/T)} \left[\frac{\partial S_{\text{s}}(X)}{\partial X} \frac{\mathrm{d}\bar{X}}{\mathrm{d}(x/T)} - \frac{x}{T} \frac{\mathrm{d}\bar{X}}{\mathrm{d}(x/T)} - \bar{X} \right]$$

$$= \frac{-\mathrm{d}\bar{X}(x, T)}{\mathrm{d}(x/T)},$$

The first two terms in the brackets cancel because at equilibrium, the sub-system intensive variables equal those of the reservoir, $x\left(\bar{X}(x, T) \right) = x$ and $T\left(\bar{X}(x, T) \right) = T$. For any value of the constrained exchangeable variable, say $X\prime$, not equal to the equilibrium value, $X\prime \neq \bar{X}$, then

$$S_{\text{total}}(X\prime \mid x, T) < S_{\text{total}}(\bar{X} \mid x, T).$$

This follows because equilibrium is defined as the maximum of the total entropy. Writing out the two sides of this inequality explicitly and re-arranging gives

$$\frac{x\bar{X}}{T} - \frac{xX\prime}{T} < S_{\text{s}}(\bar{X}) - S_{\text{s}}(X\prime).$$

The second order expansion of the sub-system entropy about $X\prime$ is

$$S_{\mathrm{s}}(\bar{X}) = S_{\mathrm{s}}(X\prime) + (\bar{X} - X\prime)S_{\mathrm{s},X}(X\prime) + \frac{(\bar{X} - X\prime)^2}{2}S_{\mathrm{s},XX}(X\prime)$$

$$= S_{\mathrm{s}}(X\prime) + (\bar{X} - X\prime)\frac{\bar{x}\prime}{\bar{T}\prime} + \frac{(\bar{X} - X\prime)^2}{2}S_{\mathrm{s},XX}(X\prime).$$

(2.100)

Inserting this into the inequality yields

$$\left[\frac{x}{T} - \frac{\bar{x}\prime}{\bar{T}\prime}\right](\bar{X} - X\prime) < \frac{(\bar{X} - X\prime)^2}{2}S_{\mathrm{s},XX}(X\prime).$$

(2.101)

Since the isolated system entropy is a concave function of its arguments, the right-hand side is negative. Hence divide both sides by $(\bar{X} - X\prime)^2$ and take the limit $X\prime \to \bar{X}$ to obtain

$$\frac{\mathrm{d}(\bar{x}/\bar{T})}{\mathrm{d}X} < 0.$$

(2.102)

This is equivalently but more conveniently written

$$\frac{\mathrm{d}\bar{X}}{\mathrm{d}(x/T)} < 0.$$

(2.103)

This shows that the equilibrium total entropy (i.e. its maximum value) is a convex function of x/T,

$$\frac{\mathrm{d}^2 S_{\mathrm{total}}(\bar{X}(x) \mid x)}{\mathrm{d}(x/T)^2} > 0.$$

(2.104)

Since the equilibrium thermodynamic potential is essentially the negative of the equilibrium total entropy $\bar{F}(x) \equiv F(\bar{X}(x) \mid x) = -T S_{\mathrm{total}}(\bar{X}(x) \mid x)$, the convexity of the latter translates into concavity of the former,

$$\frac{\mathrm{d}^2(\bar{F}(x)/T)}{\mathrm{d}(x/T)^2} < 0.$$

(2.105)

Finally, one can readily identify X and x in a given case by noting that

$$\frac{\mathrm{d}(\bar{F}(x)/T)}{\mathrm{d}(x/T)} = \bar{X}(x).$$

(2.106)

2.4.2 Heat capacity

The heat capacity tells how much energy is required to change the temperature of a sub-system. It is actually defined as the reciprocal of this, namely the rate of change of heat with temperature.

Heat flow can be thought of as the rate of microscopic or statistical energy change. Work is the rate of macroscopic energy change, the most common example being pressure–volume or *pV*-work. Neither heat not work are state functions (one cannot say that in a given thermodynamic state the system has a certain amount of heat, or a certain amount of work), but changes in the sum of the two together give the change in the energy, which is a state function.

From the First Law of Thermodynamics, the change in energy of the system is the change in heat of the system plus the work done on the system, $dE = dQ + dW$. Hence the change of energy differs if the change is at constant volume ($dW = 0$) or at constant pressure ($dW = -p\, dV$).

The heat capacity at constant volume is

$$C_V = \left(\frac{\partial Q}{\partial T} \right)_{V,N} = \left(\frac{\partial \bar{E}}{\partial T} \right)_{V,N},$$

where this is the equilibrium energy of a sub-system in contact with a heat reservoir of temperature T, $\bar{E}(N, V, T)$. Since the equilibrium energy is the derivative of the Helmholtz free energy, $\bar{F}(N, V, T)$, equation (2.34), this is the second derivative

$$C_V = \frac{-1}{T^2} \left(\frac{\partial^2 (\bar{F}/T)}{\partial (1/T)^2} \right)_{V,N}.$$

The heat capacity is evidently an extensive variable.

In the preceding section, it was established that in general the equilibrium thermodynamic potentials were concave with respect to the conjugate intensive variables of the reservoir. The present case corresponds to $x \equiv 1$ (and $X \equiv E$), and so one concludes that the heat capacity at constant volume is positive,

$$C_V > 0.$$

The heat capacity at constant pressure is

$$C_p = \left(\frac{\partial Q}{\partial T}\right)_{p,N}$$

$$= \left(\frac{\partial \bar{E}}{\partial T}\right)_{p,N} + p\left(\frac{\partial \bar{V}}{\partial T}\right)_{p,N}$$

$$= \frac{-1}{T^2}\left(\frac{\partial^2(\bar{G}/T)}{\partial(1/T)^2}\right)_{p,N},$$

where the enthalpy $E + pV$ has been written as the derivative of the Gibbs free energy, equation (2.53).

Since $\partial\bar{G}(N,p,T)/\partial T = -\bar{S}$, the heat capacity at constant pressure can also be written as the temperature derivative of the entropy

$$C_p = \frac{-1}{T^2}\frac{\partial^2(\bar{G}(N,p,T)/T)}{\partial(1/T)^2}$$

$$= -T^{-2}(-T^2)\frac{\partial}{\partial T}\left[(-T^2)\frac{\partial(\bar{G}(N,p,T)/T)}{\partial T}\right]$$

$$= \frac{\partial}{\partial T}\left[\bar{G}(N,p,T) - T\frac{\partial\bar{G}(N,p,T)}{\partial T}\right]$$

$$= \frac{\partial\bar{G}(N,p,T)}{\partial T} - \frac{\partial\bar{G}(N,p,T)}{\partial T} - T\frac{\partial^2\bar{G}(N,p,T)}{\partial T^2}$$

$$= T\frac{\partial\bar{S}(N,p,T)}{\partial T}.$$

Again invoking the concavity established in the preceding section, one identifies $x \equiv 1$ and $X = H$, and one concludes that the heat capacity at constant pressure is positive,

$$C_p > 0.$$

Further analysis shows that $C_p > C_V$. This signifies that additional heat is required to raise the temperature of a sub-system at constant pressure because extra work is required to expand the volume against the external pressure.

2.4.3 Compressibility

The compressibility is the rate of change of volume with pressure. In this case one has a choice of making the change at constant temperature or at constant entropy. The isothermal compressibility is defined to be

$$\chi_T = \frac{-1}{\bar{V}} \left(\frac{\partial \bar{V}}{\partial p} \right)_{T,N}.$$

The compressibility is evidently an intensive variable. The volume here is $\bar{V}(N, p, T)$, which is the pressure derivative of the Gibbs free energy, so that this may be rewritten

$$\chi_T = \frac{-1}{\bar{V}} \left(\frac{\partial^2 \bar{G}}{\partial p^2} \right)_{T,N} = \frac{-1}{\bar{V}T} \left(\frac{\partial^2 (\bar{G}/T)}{\partial (p/T)^2} \right)_{T,N}.$$

Invoking the concavity result and identifying $x \equiv p/T$ and $X \equiv V$, it follows that the isothermal compressibility is positive, $\chi_T > 0$.

For the case of constant entropy, the adiabatic compressibility is

$$\chi_S = \frac{-1}{\bar{V}} \left(\frac{\partial \bar{V}}{\partial p} \right)_{S,N}.$$

Since the equilibrium volume is the pressure derivative of the enthalpy at constant entropy, equation (2.86), this can be written as the second derivative of the enthalpy,

$$\chi_S = \frac{-1}{\bar{V}} \left(\frac{\partial^2 \bar{H}}{\partial p^2} \right)_{S,N}.$$

Again this is positive.

2.4.4 Thermal expansivity

The thermal expansivity is defined as

$$\alpha = \frac{-1}{\bar{\rho}} \frac{\partial \bar{\rho}(p, T)}{\partial T}.$$

This is usually positive, liquid water below 4 °C being a well-known exception. This can also be written as the cross second derivative of the Gibbs free energy,

$$\alpha = \frac{1}{\bar{V}} \left(\frac{\partial \bar{V}}{\partial T} \right)_{N,p} = \frac{1}{\bar{V}} \left(\frac{\partial^2 \bar{G}}{\partial T \partial p} \right)_N,$$

which explains why it does not have to have a definite sign

2.4.5 Maxwell relations

The four examples given above were 'pure' second derivatives of equilibrium thermodynamic potentials. One can also form the mixed second derivatives.

Interchanging the order of differentiation gives rise to the Maxwell relations.

For example, the two ways of taking the temperature and volume derivatives of the Helmholtz free energy can be equated to each other,

$$\frac{\partial^2 \bar{F}(N,V,T)}{\partial V \partial T} = \frac{\partial^2 \bar{F}(N,V,T)}{\partial T \partial V}.$$

(2.119)

Performing the inner derivatives, equations (2.30) and (2.31) yield

$$\left(\frac{\partial \bar{S}_{\mathrm{s}}(N,V,T)}{\partial V} \right)_{T,N} = \left(\frac{\partial \bar{p}(N,V,T)}{\partial T} \right)_{V,N}.$$

(2.120)

Another example, which will be of use later, what is essentially the energy per particle in a canonical equilibrium system is

$$\begin{aligned}
\frac{\partial \bar{E}(N,V,T)}{\partial N} &= \frac{\partial^2 (\bar{F}(N,V,T)/T)}{\partial N \partial (1/T)} \\
&= \frac{\partial (\bar{\mu}/T)}{\partial (1/T)} \\
&= \bar{\mu} - T \frac{\partial \bar{\mu}}{\partial T}.
\end{aligned}$$

(2.121)

The second equality follows by changing the order of differentiation in the cross second derivative of the Helmholtz free energy.

There are many such Maxwell relations. One is best advised to derive each as it is required rather than attempting to remember them all.

2.5 Probability and fluctuation theory

In this section the probability of fluctuations in the state of a sub-system able to exchange with a reservoir is analyzed in a generic fashion.

We shall denote by Y the set of extensive sub-system variables that are fixed, and by X the extensive sub-system variables that can be exchanged with the reservoir. The conjugate intensive variables are defined by the derivatives of the isolated sub-system entropy,

$$x = T \frac{\partial S(X,Y)}{\partial X} \quad \text{and} \quad y = T \frac{\partial S(X,Y)}{\partial Y}.$$

(2.122)

The most usual pairs of conjugate variables are $\{x, X\} = \{1, E\}$, $\{p, V\}$, and $\{-\mu_\alpha, N_\alpha\}$. The X and the x are vectors in the case that more than one quantity is exchangeable with the reservoir.

2.5.1 Exchange with a reservoir

Using a subscript s for the sub-system and r for the reservoir, as usual the total entropy is

$$S_{\text{total}}(X_{\text{s}} \mid Y_{\text{s}}, Y_{\text{r}}, X_{\text{total}}) = S_{\text{s}}(X_{\text{s}}, Y_{\text{s}}) + S_{\text{r}}(X_{\text{total}} - X_{\text{s}}, Y_{\text{r}})$$
$$= S_{\text{s}}(X_{\text{s}}, Y_{\text{s}}) - \frac{x_{\text{r}}}{T_{\text{r}}} X_{\text{s}}.$$

(2.123)

Only that part of the total entropy that depends on the sub-system, and terms extensive with the sub-system, have been kept here. Dropping the subscripts since the reservoir only enters through its intensive variables this is

$$S_{\text{total}}(X \mid Y, x, T) = S_{\text{s}}(X, Y) - \frac{x}{T} X.$$

(2.124)

The temperature that appears here is that of the reservoir. In the event of multiple exchangeable parameters, $xX \Rightarrow \mathbf{x} \cdot \mathbf{X}$.

The probability that the sub-system is in the macrostate X (i.e. has an amount X of exchangeable material) is simply proportional to the exponential of the total entropy,

$$\wp(X \mid Y, x, T) = \frac{1}{Z(Y, x, T)} e^{S_{\text{total}}(X\mid Y,x,T)/k_{\text{B}}}$$
$$= \frac{1}{Z(Y, x, T)} e^{S_{\text{s}}(X,Y)/k_{\text{B}}} e^{-\beta x X},$$

(2.125)

where $\beta \equiv 1/k_{\text{B}}T$ is often called the inverse temperature. The normalizing factor $Z(Y, x, T)$ is called the partition function.

As we have seen, the total entropy is a maximum when the conjugate sub-system field variable equals that of the reservoir,

$$\frac{\partial S_{\text{total}}(X \mid Y, x, T)}{\partial X} \Big|_{X=\overline{X}} = 0 \Leftrightarrow x_{\text{s}}(\overline{X}_{\text{s}}, Y_{\text{s}}) = x_{\text{r}}.$$

(2.126)

Obviously this corresponds to the maximum of the probability distribution, and $\overline{X}$ is the most likely macrostate of the system, or, equivalently, the most likely value of the exchangeable variable of the sub-system.

2.5.1.1 Concavity of the entropy

It was argued in section 2.1.3 that the entropy of an isolated sub-system must be a concave function of its extensive arguments in order for the thermodynamic state to be stable. This condition leads to

$$S_{ii} < 0, \quad \text{and} \quad S_{ii}S_{jj} - S_{ij}^2 > 0,$$

(2.127)

where the subscripts denote the second derivatives with respect to any of the isolated sub-system X_i. (The second condition comes from the fact that the eigenvalues of the Jacobean matrix must all be negative, and, for two exchangeable parameters, the product of the two eigenvalues is the determinant.)

In the present case of exchange with a reservoir, for a stable equilibrium state, the total entropy must be a concave function of its constrained arguments. But this is guaranteed by the concavity of the isolated system entropy, since the part of the total entropy connected to the reservoir is a linear function of the exchangeable variables, and concavity reflects the second derivative.

The connection between the two cases is not so surprising since a macroscopic isolated system forms a reservoir for any part of itself.

2.5.1.2 Gaussian probability

The concavity of the total entropy means that the probability distribution for the constrained variable that the sub-system can exchange with a reservoir has a well-defined peak at $\bar{X}$. One can make a second order expansion of the exponent about this peak to obtain a Gaussian distribution,

$$\wp(X \mid Y, x, T) \approx \frac{1}{Z_{\mathrm{G}}(Y, x, T)} e^{S_{\mathrm{s}}\prime\prime(X - \bar{X})^2 / 2 k_{\mathrm{B}}}.$$

(2.128)

where $S_{\mathrm{s}}\prime\prime \equiv S_{\mathrm{s}}\prime\prime(\bar{X}(Y, x, T), Y)$ is the Jacobian matrix of second derivatives of the isolated sub-system entropy evaluated in the equilibrium state.

Note that the contributions from higher order derivatives are negligible, as can be seen from extensivity arguments. (The nth entropy derivative $S_{\mathrm{s}}^{(n)}$ scales with V^{1-n}, whereas $(X - \bar{X})^n$ scales with $V^{n/2}$, so that their product decreases in magnitude with increasing volume for $n > 2$.) Hence this Gaussian form for the probability distribution is exact in the thermodynamic limit, $V \to \infty$.

This result is essentially the central limit theorem, which in essence says that the probability distribution of variables that are the sum of random variables is Gaussian. The utility of Gaussian distributions is another strong argument for formulating the entropy as dependent solely on extensive variables.

From the easily proved properties of the Gaussian distribution, the average value is equal to the most likely value,

$$\langle X \rangle_{Y, x, T} = \bar{X}(Y, x, T).$$

(2.129)

This result is essential to the consistency of thermodynamics and statistical mechanics. Thermodynamics always refers to the most likely state or value. Statistical mechanics mainly deals with average values. By this result the two are the same.

2.5.2 Constrained and equilibrium thermodynamic potential

As we have seen, in order to make more direct contact with conventional thermodynamics, it is useful to introduce the constrained thermodynamic potential as the negative of the temperature times the constrained total entropy,

$$\begin{aligned} F(X \mid Y, x; T) &= -T S_{\mathrm{total}}(X \mid Y, x, T) \\ &= xX - T S_{\mathrm{s}}(X, Y). \end{aligned}$$

(2.130)

(The reservoir temperature T that appears here affects the potential only trivially in the event that energy is not an exchangeable parameter.) This may also be called

the fluctuation potential, for reasons that will become clearer shortly. It follows that the probability distribution is essentially the exponential of this constrained thermodynamic potential

$$\wp(X \mid Y, x, T) = \frac{1}{Z(Y, x, T)} e^{-\beta F(X \mid Y, x, T)}.$$

(2.131)

By design, the constrained thermodynamic potential is a convex function of the exchangeable parameters that reaches its minimum at the equilibrium value, $X = \bar{X}(Y, x, T)$. This minimum value is called the equilibrium thermodynamic potential,

$$\bar{F}(Y, x, T) = F(\bar{X} \mid Y, x, T) = x\bar{X} - TS_{\mathrm{s}}(\bar{X}, Y).$$

(2.132)

The utility of the constrained thermodynamic potential is that it provides a variational principle for the exchanged variable. Amongst other things, this makes differentiation of the equilibrium thermodynamic potential particularly simple because one may effectively hold the exchanged variable constant during differentiation. For example

$$
\begin{aligned}
\frac{\partial \bar{F}(Y, x, T)}{\partial x} &= \frac{\partial F(\bar{X} \mid Y, x, T)}{\partial x} + \frac{\partial F(X \mid Y, x, T)}{\partial X}\Big|_{X=\overline{X}} \frac{\partial \overline{X}(Y, x, T)}{\partial x} \\
&= \frac{\partial F(\bar{X} \mid Y, x, T)}{\partial x} \\
&= \bar{X}.
\end{aligned}
$$

(2.133)

For the same reason differentiation with respect to one of the non-exchangeable sub-system variables yields

$$
\begin{aligned}
\frac{\partial \bar{F}(Y, x, T)}{\partial Y} &= \frac{\partial F(\bar{X} \mid Y, x, T)}{\partial Y} + \frac{\partial F(X \mid Y, x, T)}{\partial X}\Big|_{X=\overline{X}} \frac{\partial \bar{X}(Y, x, T)}{\partial Y} \\
&= \frac{\partial F(\bar{X} \mid Y, x, T)}{\partial Y} \\
&= -T\frac{\partial S(\bar{X}, Y)}{\partial Y} \\
&= -T\frac{y}{T_{\mathrm{s}}},
\end{aligned}
$$

(2.134)

where the conjugate intensive variable of the sub-system and also the sub-system temperature appear. Dividing by temperature and differentiating by inverse temperature yields

$$\frac{\partial(\bar{F}(Y,x,T)/T)}{\partial T^{-1}}$$

$$= \frac{\partial(F(\bar{X}\mid Y,x,T)/T)}{\partial T^{-1}} + \frac{\partial(F(X\mid Y,x,T)/T)}{\partial X}\Big|_{X=\bar{X}}\frac{\partial\bar{X}(Y,x,T)}{\partial T^{-1}}$$

$$= \frac{\partial(F(\bar{X}\mid Y,x,T)/T)}{\partial T^{-1}}$$

$$= x\bar{X}.$$

In the event of multiple exchangeable variables, the right-hand side is $\mathbf{x}\cdot\bar{\mathbf{X}}$.

The equilibrium thermodynamic potential is a convex function of the non-exchangeable extensive variables, (because its second derivative is the negative of the second derivative of the isolated sub-system entropy). It is a concave function of the reservoir field variables conjugate to the exchangeable variables. Hence

$$\frac{\partial^2\bar{F}}{\partial x_i^2} < 0, \quad \text{and} \quad \frac{\partial^2\bar{F}}{\partial x_i^2}\frac{\partial^2\bar{F}}{\partial x_j^2} - \left(\frac{\partial^2\bar{F}}{\partial x_i\partial x_j}\right)^2 > 0.$$

The curvature of the equilibrium thermodynamic potential determines the sign of physical quantities such as the heat capacity or the compressibility.

2.5.3 Partition function

The partition function that normalizes the probability distribution is

$$Z(Y,x,T) = \sum_X e^{S_{\text{total}}(X\mid Y,x,T)/k_B}$$

$$= \sum_X e^{S_s(X,Y)/k_B}e^{-\beta xX}$$

$$= \int dX\, e^{S_s(X,Y)/k_B}e^{-\beta xX}.$$

In the integral form one often introduces an appropriate constant factor that makes the partition function dimensionless. This is unimportant as it has no non-trivial consequences. Since the partition function is the weighted sum over all macrostates, its logarithm is the unconstrained total entropy,

$$S_{\text{total}}(Y,x,T) \equiv k_B\, \ln Z(Y,x,T).$$

This explains why any constant factor multiplying the partition function has no physical consequences: entropy is only defined up to an additive constant.

2.5.3.1 Averages
The partition function acts as a generating function for the system. The average value of various properties of the sub-system may be obtained as derivatives of the

logarithm of the partition function. The latter is the unconstrained total entropy of the system.

The average of an exchangeable variable of the sub-system is

$$
\begin{aligned}
\langle X \rangle_{Y,x,T} &= \int \mathrm{d}X \, \wp(X \mid Y, x, T) \, X \\
&= \frac{1}{Z(Y,x,T)} \int \mathrm{d}X \, \mathrm{e}^{S_\mathrm{s}(X,Y)/k_\mathrm{B}} \mathrm{e}^{-\beta x X} \, X \\
&= \frac{1}{Z(Y,x,T)} \int \mathrm{d}X \, \mathrm{e}^{S_\mathrm{s}(X,Y)/k_\mathrm{B}} \frac{-\partial \mathrm{e}^{-\beta x X}}{\partial(\beta x)} \\
&= \frac{-\partial \, \ln \, Z(Y,x,T)}{\partial(\beta x)}.
\end{aligned}
$$

(2.139)

Similarly, the sub-system field parameters conjugate to the non-exchangeable variables, $y = T_\mathrm{s} \partial S(X,Y)/\partial Y$, have average value

$$
\begin{aligned}
\langle \beta_\mathrm{s} y \rangle_{Y,x,T} &= \int \mathrm{d}X \, \wp(X \mid Y, x, T) \, \beta_\mathrm{s} y \\
&= \frac{1}{Z(Y,x,T)} \int \mathrm{d}X \, \mathrm{e}^{S_\mathrm{s}(X,Y)/k_\mathrm{B}} \mathrm{e}^{-\beta x X} \frac{\partial S_\mathrm{s}(X,Y)}{k_\mathrm{B} \partial Y} \\
&= \frac{1}{Z(Y,x,T)} \int \mathrm{d}X \, \mathrm{e}^{-\beta x X} \frac{\partial \mathrm{e}^{S_\mathrm{s}(X,Y)/k_\mathrm{B}}}{\partial Y} \\
&= \frac{\partial \, \ln \, Z(Y,x,T)}{\partial Y}.
\end{aligned}
$$

(2.140)

The average of the sub-system field parameters conjugate to the exchangeable variables, is

$$
\begin{aligned}
\langle \beta_\mathrm{s} x_\mathrm{s} \rangle_{Y,x,T} &= \int \mathrm{d}X \, \wp(X \mid Y, x, T) \, \beta_\mathrm{s} x_\mathrm{s} \\
&= \frac{1}{Z(Y,x,T)} \int \mathrm{d}X \, \mathrm{e}^{S_\mathrm{s}(X,Y)/k_\mathrm{B}} \mathrm{e}^{-\beta x X} \frac{\partial S_\mathrm{s}(X,Y)}{k_\mathrm{B} \partial X} \\
&= \frac{1}{Z(Y,x,T)} \int \mathrm{d}X \, \mathrm{e}^{-\beta x X} \frac{\partial \mathrm{e}^{S_\mathrm{s}(X,Y)/k_\mathrm{B}}}{\partial X} \\
&= \frac{-1}{Z(Y,x,T)} \int \mathrm{d}X \, \mathrm{e}^{-\beta x X} \frac{\partial \mathrm{e}^{-\beta x X}}{\partial X} \\
&= \beta x.
\end{aligned}
$$

(2.141)

An integration by parts has been performed here, with the integrated part vanishing because extreme (i.e. at the integral limits) values of the exchangeable parameter have vanishingly small probability. This says that the average of the sub-system intensive variable conjugate to the exchangeable parameter is equal to that of the reservoir. This is what one would expect from the above finding that average values equal most likely values.

2.5.3.2 Fluctuations

As we have just seen, the derivative of the unconstrained total entropy, which is the logarithm of the partition function, give the average of various properties of the sub-system. Likewise, the second derivative gives the fluctuations in those quantities. One has

$$
\begin{aligned}
\frac{\partial^2 \ln Z(Y,x,T)}{\partial(\beta x)^2} &= \frac{-\partial\langle X\rangle_{Y,x,T}}{\partial(\beta x)} \\
&= \frac{-\partial}{\partial(\beta x)}\left\{ \frac{1}{Z(Y,x,T)} \int dX\, e^{S_s(X,Y)/k_B} e^{-\beta x X}\, X \right\} \\
&= \langle X^2\rangle_{Y,x,T} - \langle X\rangle^2_{Y,x,T} \\
&= \left\langle (\Delta X)^2 \right\rangle_{Y,x,T}.
\end{aligned}
$$

(2.142)

Here the fluctuation or departure from the norm is $\Delta X \equiv X - \langle X\rangle_{Y,x,T}$. The square root of the average of the square of the fluctuation is a measure of the extent to which the system can be expected to depart at any instant from its average value.

The average of the square of the departure is evidently positive. Hence the unconstrained total entropy is a convex function of the reservoir conjugate field variables. The fluctuation in X measures the width of the probability distribution $\wp(X\,|\,Y,x,T)$: a small fluctuation means a sharply peaked distribution. The numerator on the left-hand side is extensive (the total entropy scales with the size of the sub-system), and the denominator is intensive. Hence the average of the square of the fluctuation is extensive. The expected relative error in a measurement of X is

$$
\frac{\sqrt{\left\langle (\Delta X)^2 \right\rangle_{Y,x,T}}}{\langle X\rangle_{Y,x,T}} \sim \mathcal{O}(V^{-1/2}),
$$

(2.143)

since the denominator is extensive. This vanishes in the thermodynamic limit, $V \to \infty$. Hence the probability distribution for the exchangeable parameter becomes infinitely sharply peaked in the thermodynamic limit. This means that the average value is the same as the most likely value. It also means that the relative statistical error in the measured value of the exchangeable quantity is negligible for a macroscopic system.

Similarly, for two exchangeable quantities, the cross second derivative is

$$\frac{\partial^2 \ln Z(Y, x, T)}{\partial(\beta x_i)\partial(\beta x_j)} = \langle X_i X_j \rangle_{Y,x,T} - \langle X_i \rangle_{Y,x,T} \langle X_j \rangle_{Y,x,T}$$

$$= \langle \Delta X_i \, \Delta X_j \rangle_{Y,x,T}.$$

(2.1 44)

This is called the cross-correlation, and it can be positive or negative. If it vanishes then the two exchangeable quantities are said to be uncorrelated with each other.

It is possible to combine the cross- and self-correlation in such a way that a positive quantity is obtained,

$$\left(\frac{\partial}{\langle X_i \rangle_{Y,x,T}\partial(\beta x_i)} - \frac{\partial}{\langle X_j \rangle_{Y,x,T}\partial(\beta x_j)} \right)^2 \ln Z(Y, x, T)$$

$$= \left\langle \left(\frac{X_i}{\langle X_i \rangle_{Y,x,T}} - \frac{X_j}{\langle X_j \rangle_{Y,x,T}} \right)^2 \right\rangle_{Y,x,T}.$$

(2.1 45)

The averages on the left-hand side are to be regarded here as constants that are not differentiated.

Differentiation with respect to the inverse temperature yields

$$\frac{\partial^2 \ln Z(Y, x, T)}{\partial \beta^2} = \frac{-\partial \langle xX \rangle_{Y,x,T}}{\partial \beta}$$

$$= \left\langle (\Delta(xX))^2 \right\rangle_{Y,x,T}.$$

(2.1 46)

This gives the fluctuation of the exchangeable work term; for multiple quantities, $xX \Rightarrow \mathbf{x} \cdot \mathbf{X}$.

One can similarly analyze the non-exchangeable extensive variables of the sub-system. This yields

$$\frac{\partial^2 \ln Z(Y, x, T)}{\partial Y^2} = \frac{\partial \langle \beta_s y \rangle_{Y,x,T}}{\partial Y}$$

$$= \left\langle (\beta_s y)^2 \right\rangle_{Y,x,T} - \langle \beta_s y \rangle_{Y,x,T}^2 + \left\langle \frac{\partial(\beta_s y)}{\partial Y} \right\rangle_{Y,x,T}$$

$$= \left\langle (\Delta(\beta_s y))^2 \right\rangle_{Y,x,T} + \left\langle \frac{\partial^2 S_s(X, Y)/k_B}{\partial Y^2} \right\rangle_{Y,x,T}.$$

(2.1 47)

The unconstrained total entropy is extensive, and so the left-hand side is $\mathcal{O}(V^{-1})$. The final term on the right-hand side is similarly $\mathcal{O}(V^{-1})$. Hence the average of the square of the fluctuation in the sub-system intensive variables (the ones not conjugate to the exchangeable variables) must scale as $\mathcal{O}(V^{-1})$, which in the thermodynamic limit is negligible compared to $\langle \beta_s y \rangle_{Y,x,T} \sim \mathcal{O}(V^0)$.

The first term on the right-hand side is positive, whereas the second term is negative due to the concavity of the entropy of the isolated sub-system. Hence the sign of the left-hand side is indeterminate. However, the result may be rewritten as

$$\frac{\partial^2 \ln Z(Y,x,T)}{\partial Y^2} - \left\langle \frac{\partial^2 S_s(X,Y)/k_B}{\partial Y^2} \right\rangle_{Y,x,T} = \left\langle (\Delta(\beta_s y))^2 \right\rangle_{Y,x,T},$$

(2.148)

which says that the left-hand side is positive.

2.5.3.3 Partition function and thermodynamic potential

The logarithm of the partition function is essentially the unconstrained total entropy,

$$S_{\text{total}}(Y,x,T) = k_B \ln Z(Y,x,T).$$

(2.149)

This plays an important role in the formalism as its derivatives generate average values and their fluctuations.

In conventional approaches instead of this it is the equilibrium thermodynamic potential that is equated to the logarithm of the partition function. There is no inconsistency here because for macroscopic systems (i.e. in the thermodynamic limit) the two are equal because fluctuations are relatively negligible. This can be seen explicitly,

$$-k_B T \ln Z(Y,x,T)$$
$$= -k_B T \ln \int dX \, e^{S_s(X,Y)/k_B} e^{-\beta x X}$$
$$\approx -k_B T \ln \left[e^{S_s(\bar{X},Y)/k_B} e^{-\beta x \bar{X}} \int dX \, e^{S_s''(\bar{X},Y)(X-\bar{X})^2/2k_B} \right]$$
$$= \beta x \bar{X} - T S_s(\bar{X},Y) - \frac{k_B T}{2} \ln \left[\text{Det}\left\{ -S_s''(\bar{X},Y)/2\pi k_B \right\} \right]$$
$$= \bar{F}(Y,x,T) - \frac{k_B T}{2} \ln \left[\text{Det}\left\{ -S_s''(\bar{X},Y)/2\pi k_B \right\} \right].$$

(2.150)

For multiple exchangeable parameters, $x\bar{X} \Rightarrow \underline{x} \cdot \underline{\bar{X}}$, and $S_s''(\bar{X},Y)$ is the Jacobean matrix of second derivatives. In this derivation the exponent has been expanded to second order and the Gaussian integral evaluated. This gives the equilibrium thermodynamic potential or free energy plus a logarithmic correction. The latter arises from the contribution of the fluctuations to the unconstrained total entropy, and for a macroscopic system it can be neglected. (This is in essence the same result as was given for the combinatorial example in section 1.2.3). One concludes that to leading order the conventional expression is correct: the equilibrium thermodynamic potential is essentially the logarithm of the partition function,

$$\bar{F}(Y,x,T) = -k_B T \ln Z(Y,x,T).$$

(2.151)

This result means that the logarithmic derivative of the partition function equals essentially the derivative of the equilibrium thermodynamic potential. Hence average values, which are given by the former, equal most likely values, which are given by the latter. Similarly for the second derivatives, so that the sign of the average square fluctuation determines the sign corresponding to equilibrium derivatives.

Summary

- The constrained total entropy is the sum of the sub-system entropy and the reservoir entropy, each isolated and in the specified macrostate. The reservoir entropy upon Taylor expansion is the negative of the sub-system constrained exchangeable parameters times the conjugate reservoir field variables. The constrained thermodynamic potential is the negative of the reservoir temperature times the constrained total entropy.
- The constrained total entropy is maximized by the equilibrium or most likely value of the constrained parameter. The corresponding minimum value of the constrained thermodynamic potential is the equilibrium thermodynamic potential, which is commonly called a free energy.
- The concavity of the isolated system entropy leads to the concavity of the constrained total entropy, the stability of matter, the uniqueness of the thermodynamic state, and the equivalence of the equilibrium state obtained with different reservoirs. Exchangeable variables flow up the entropy gradient, which makes the conjugate field variable more uniform spatially. In common parlance, energy flows from a hot body to a cold body, thereby decreasing the temperature disparity.
- The constrained total entropy, constrained thermodynamic potential, and equilibrium thermodynamic potential are all extensive.
- The equilibrium thermodynamic potential is a concave function of its field variables and a convex function of its extensive variables. These determine the sign of its second derivatives, which give common thermodynamic parameters such as the heat capacity, compressibility, etc.
- The values of the sub-system variables exchangeable with a reservoir fluctuate about their equilibrium values. The logarithm of the partition function gives the total unconstrained entropy, which is approximately equal to the maximum value of the constrained total entropy, which is essentially the equilibrium thermodynamic potential.
- Fluctuations of extensive variables are Gaussian distributed with a variance determined by the second derivative of the equilibrium thermodynamic potential (equivalently, the isolated system entropy). Fluctuations are relatively negligible for a macroscopic system, to which extent the Second Law of Thermodynamics may be regarded as a deterministic rather than a probabilistic law.

References

Attard P 2002 *Thermodynamics and Statistical Mechanics: Equilibrium by Entropy Maximisation* (London: Academic)

Callan H B 1960 *Thermodynamics* (New York: Wiley)

Clausius R 1854 Über eine veränderte form des zweiten hauptsatzes der mechanischen wärmetheorien *Ann. Phys. Chem.* **93** 481

Gibbs J W 1873 Graphical methods in the thermodynamics of fluids *Trans. Conn. Acad. Arts Sci.* **2** 309

Gibbs J W 1902 *Elementary Principles in Statistical Mechanics Developed with Special Reference to the Rational Foundation of Thermodynamics* (New Haven, CT: Yale University Press)

Sears F W and Salinger G L 1986 *Thermodynamics, Kinetic Theory, and Statistical Thermodynamics* 3rd edn (Reading, MA: Addison-Wesley)

Szent-Györgyi A 2011 *Drive and Curiosity* ed Hargittai I (ed) (New York: Prometheous) p 22

Entropy Beyond the Second Law (Second Edition)

Thermodynamics and statistical mechanics for equilibrium, non-equilibrium, classical, and quantum systems

Phil Attard

Chapter 3

The entropy for transitions and non-equilibrium systems

'the decay of a system from a given non-equilibrium state produced by a spontaneous fluctuation obeys, *on the average*, the (empirical) law for the decay from the same state back to equilibrium, when it has been produced by a constraint which is then suddenly removed'

(Onsager and Machlup 1953)

'Entropy' is a word that seems to attract the crackpots of the pseudo-scientific societies ... the second law is highly attractive to those who are rather more philosophic and hand-waving than is acceptable in the normal circles of the hard-sciences'

(Paltridge in Kleidon and Lorenz 2005)

'The mistakes of the common herd are usually in the same direction. Like sheep, they all follow a single leader'

(Fisher 1906)

The main aim of this chapter is to formulate non-equilibrium thermodynamics as a variational principle of the appropriate entropy. The particular theory has been developed by me (Attard 2005a), and the presentation follows previous work (Attard 2012). The approach is illustrated by treating in detail the canonical non-equilibrium system, namely steady heat flow.

The most wide-spread competing approach to non-equilibrium thermodynamics is a variational principle based on the rate of entropy production,

$$\delta \dot{S}\big|_{\bar{J}} = 0.$$

(3.1)

This asserts that the optimum non-equilibrium state (symbolized by the flux $\bar{J}$) is the one that extremizes the rate of entropy production, $\dot{S}$, which is also called the dissipation. There are also variants of this based on extremizing different functions of the dissipation. This is claimed to be the variational principle upon which the field of non-equilibrium thermodynamics must be based. It is meant to be the generalization of the Second Law of Thermodynamics to non-equilibrium systems. This formula has quite a long history (see below), and it has been advocated and used by at least three Nobel Laureates (Rayleigh, Onsager, Prigogine), and other authorities besides.

3.1 Entropy for non-equilibrium systems

In this section the way to formulate an entropy for non-equilibriums systems is given in the generality of sets and states. But first is briefly discussed the need to develop a principle beyond the Second Law as a basis for non-equilibrium thermodynamics, which leads to a historical review of the above canard.

3.1.1 The Second Law is not enough

Chapter 2 analyzed the variational principle for equilibrium thermodynamics, namely the Second Law of Thermodynamics that entropy is maximized. An equilibrium system does not change macroscopically with time, although of course microscopic fluctuations occur due to the incessant molecular motion.

A non-equilibrium system is one that changes macroscopically with time. Such changes can be due to an applied time-dependent mechanical field, or to an applied thermodynamic gradient. An example of the latter is steady heat flow, where the energy of the hot reservoir decreases and that of the cold reservoir increases over time.

In fact equilibrium and non-equilibrium systems are more closely connected than is implied by the above distinction. As we shall see, the fluctuations in time of an equilibrium system are intimately related to the flux or motion in a non-equilibrium system.

The reader may be puzzled by the opening paragraph of this subsection in which the Second Law of Thermodynamics was referred to as the basis of equilibrium thermodynamics. Surely the fact that it refers to spontaneous changes in a system is a signifier of a non-equilibrium system?

It is certainly true that the Second Law refers to macroscopic changes, and that these are the stuff of non-equilibrium thermodynamics. It is also undeniably true that the Second Law provides the variational principle that underlies the quantitative treatment of equilibrium systems, as evidenced by the results in chapter 2. The point is, however, that whilst the Second Law sets the direction of change for a system that is not in equilibrium, it does not quantitatively determine the rate of change. In short, the Second Law gives the future direction of time's arrow but not its speed. Any thermodynamic theory for non-equilibrium systems must involve time in a quantitative not just qualitative sense.

3.1.2 Principle of extreme dissipation

'these [reciprocal] relations can be summarized in a variation-principle...an extension of Lord Rayleigh's *principle of the least dissipation of energy*'...the rate of increase of the entropy plays the rôle of a potential' (Onsager 1931)

This brings us to the history and motivation for the principle of extreme dissipation, equation (3.1). The absence of quantitative time in the Second Law of Thermodynamics has led many scientists to the same general conclusion: the Second Law must be modified for non-equilibrium systems.

In order to insert time quantitatively, almost the only approach that has been tried is to replace the entropy that appears there by its time rate of change. The latter is also known as the rate of entropy production, or the rate of dissipation, or just the dissipation.

In all cases the argument has been made by analogy with the Second Law rather than from first principles, and so workers pursuing such an approach have had to guess at the variational principle. As a result two contradictory versions have been promulgated. A substantial fraction of the advocates of this approach assert that the general principle that underlies non-equilibrium thermodynamics is that the rate of entropy production is a *minimum*. Or by analogy to the Second Law,

$$\text{The rate of entropy production decreases during}$$
$$\text{spontaneous changes of the system.}$$

(3.2)

Onsager, in his Noble-prize winning paper on the reciprocal relations, calls this 'The Principle of Least Dissipation'. He characterizes the dissipation by generalizing the energy dissipation functions given earlier by Rayleigh (Strutt 1871), and by Kelvin (Onsager 1931). Rayleigh (Strutt 1913) attributes it to Helmholtz (1971) and to Korteweg (1883). This principle of minimal entropy production has also been asserted by Prigogine (1967), de Groot and Mazur (1984), Biot (1955, 1975), Paltridge (1979), and many others besides.

Contradicting this, more than half of the advocates of extremizing the rate of entropy production assert the exact opposite. They say that the rate of entropy production is a *maximum*, which, by analogy to the Second Law, can be formulated as

$$\text{The rate of entropy production increases during}$$
$$\text{spontaneous changes of the system.}$$

(3.3)

Examples of scientists arguing for this include Prigogine (1967), Paltridge (1979), Swenson and Turvey (1991), Schneider and Kay (1994), and Dewar (2003). For other examples and reviews see Kleidon and Lorenz (2005) and Martyushev and Seleznev (2006). Some scientists, including the Noble laureate Prigogine, have advocated both principles.

Of some influence has been the variational functional given by Onsager and Machlup (1953), based on different quadratic forms for the entropy production, and whose extremum is designed to give the optimum steady state of a non-equilibrium system. They chose the extremum to be a minimum. The variational functionals given by Hashitsume (1952, 1956), Gyarmati (1968, 1970), and Bochkov and Kuzovlev (1980), are in essence identical to that given by Onsager and Machlup (1953). Some of the proponents of this functional say the extremum should be a maximum, and others say is should be a minimum.

The reader will doubtless be troubled by the fact that two directly opposed criteria have been proposed and argued for with equal fervor. Surely there should be a unique general principle that determines the optimum non-equilibrium state, and this principle should be self-evident, incontrovertible, and backed by evidence. It is perhaps surprising that those who argue for a principle based on the rate of entropy production cannot agree amongst themselves whether it should be a maximum or a minimum. Despite the evident contradiction and the dearth of convincing evidence for either, there has been little criticism in the literature of the principle of extreme dissipation. Doubtless most workers have chosen discretion over valor, preferring to follow each other rather than to question authority. The few honorable exceptions that I am aware of are Gage *et al* (1966), Keizer and Fox, (1974), Lavenda (1985), Hunt *et al* (1987), Ross and Vlad (2005), and of course myself (Attard 2006).

I won't give here a detailed critique of the principle of extreme dissipation (for such, see Attard (2006)) since it is preferable to devote this chapter to the correct non-equilibrium variational principle rather than to be distracted by a fundamentally flawed approach. Briefly however, it can be noted that the rate of change of an extensive variable $\dot{X}$ is a flux, and multiplying by the conjugate field variable x/T, a thermodynamic force, gives the rate of entropy production,

$$\dot{S} = \frac{x}{T}\dot{X}.$$

(3.4)

There are many details that are skipped here, but the essence of the argument can be readily appreciated. Since this is a function bilinear in the flux and the force, extremizing it with respect to either can only ever yield $\pm\infty$. Onsager and Machlup (1953) get around this problem by adding quadratic terms in the force and in the flux, which also give the entropy dissipation in the optimum state, and these serve to produce an extremum at finite values (for detailed analysis, see Attard (2006)). There are of course an infinite number of variational functionals that one can construct to give the known optimum steady state. What is required is the unique variational formula that comes from the physical principles that generalize the Second Law of Thermodynamics. This matter is further discussed following equation (3.95).

3.1.3 Entropy of joint macrostates

In section 1.1.2, entropy and probability were defined in terms of the weight of macrostates and microstates, which is basically set theory. It will be recalled that microstates are disjoint, indivisible, and form a complete set, and that macrostates are disjoint and also form a complete set. With w_i the weight of microstate i, the weight of the macrostate α is

$$W_\alpha = \sum_i w_i \delta(A_i - A_\alpha) = \sum_{i \in \alpha} w_i,$$

(3.5)

and the total weight of the system is

$$W = \sum_i w_i = \sum_\alpha W_\alpha.$$

(3.6)

The entropy is essentially the logarithm of the weight, so that the microstate, macrostate, and total entropy is respectively

$$S_i = k_B \ln w_i, \quad S_\alpha = k_B \ln W_\alpha, \quad \text{and} \quad S = k_B \ln W.$$

(3.7)

Since macrostates from different collectives, labeled by say α and β, are not disjoint, their joint weight can be calculated as the weight of their intersection,

$$W(\alpha\beta) = \sum_i w_i \delta(A_i - A_\alpha)\delta(B_i - B_\beta) = \sum_{i \in A_\alpha \cap B_\beta} w_i.$$

(3.8)

This is sketched in figure 1.1. This is the unconditional weight. If α and β belong to the same collective, this is $W(\alpha\beta) = W_\alpha \delta(\alpha - \beta)$.

Because the collectives are complete, summing over one of the macrostates must reduce the joint weight to the weight of the remaining macrostate,

$$\sum_\beta W(\alpha\beta) = W_\alpha.$$

(3.9)

This is a type of conservation law for weight. The idea can again be seen in figure 1.1, where it is evident that the α macrostate is covered by all the intersecting β macrostates.

It follows that summing over the joint weight must yield the total weight

$$\sum_{\alpha,\beta} W(\alpha\beta) = W.$$

(3.10)

The entropy of the joint macrostate is of course the logarithm of its weight, and the joint probability is

$$\wp(\alpha\beta) = \frac{W(\alpha\beta)}{W} = \frac{e^{S(\alpha\beta)/k_B}}{W}.$$

(3.11)

This is the unconditional probability, which is to say the probability of the system being in the state α and in the state β. The conditional probability $\wp(\alpha \mid \beta)$, which is to say the

probability that the system is in the state α given that it is in the state β, was discussed in section 1.1.2. It is

$$\wp(\alpha \mid \beta) = \frac{\wp(\alpha\beta)}{\wp(\beta)} = \frac{W(\alpha\beta)}{W(\beta)} = \mathrm{e}^{[S(\alpha\beta)-S(\beta)]/k_\mathrm{B}}.$$

(3.1 2)

3.1.4 Second entropy for transitions

Consider the transition between two microstates in the time interval τ, $i \xrightarrow{\tau} j$. This is the key to the analysis. Just as the microstates i themselves provided the elements for the analysis of equilibrium systems, it is these transitions between microstates that are the elements of non-equilibrium theory.

As the microstates have weight, so also do the transitions between them, and the manipulation of these transition weights is based on set theory rather similar to the analysis of the microstate weights above and in section 1.1.2. The state with the system in the microstate i at time t_1 and j at time t_2 may be described as a transition microstate. Attached to it is the weight, $w(j, t_2; i, t_1) = w(j, i \mid t_{21})$, where $t_{21} \equiv t_2 - t_1$. That this depends only on the time interval is true for an equilibrium system, which is homogeneous in time, and which is the sole concern here.

This is the unconditional weight, and one can just as well reverse the order of the arguments and times,

$$w(j, t_2; i, t_1) = w(i, t_1; j, t_2) \Leftrightarrow w(j, i \mid t_{21}) = w(i, j \mid t_{12}).$$

(3.1 3)

These all give the weight for the system to be in the microstate i at time t_1 and j at time t_2. This is the unconditional weight, and the unconditional transition probability is proportional to it.

In common parlance, if $t_2 > t_1$, this would normally be called the transition from i to j. However it is important in the present theory to also consider transitions backward in time. Because of the symmetry of the unconditional weight, forward and backward unconditional transitions have equal weight. However, we know from the Second Law of Thermodynamics that there must be an asymmetry or irreversibility in time. For this reason we shall try to avoid using the word transition in association with the unconditional weight, and instead restrict it to conditional weights.

Conditional weights are defined in the usual way, via

$$w(j, i \mid \tau) = w(j \mid i, \tau) \, w(i).$$

(3.1 4)

That the microstate weight $w(i)$ is independent of time is true for the present equilibrium system. The conditional weight $w(j \mid i, \tau)$ is to be read as the weight associated with the fact the system is in the microstate j given that it was in the microstate i a time τ earlier. Here i is the initial microstate and j is the destination microstate. This is the weight for the unconditional transition i to j in time τ, which is denoted $i \xrightarrow{\tau} j$. This may be a forward or a backward transition, according to the sign of τ.

The conditional weight is in general asymmetric in time. This follows since $w(j, i \mid \tau) = w(i, j \mid -\tau)$, which may be rearranged as

$$\frac{w(j \mid i, \tau)}{w(i \mid j, -\tau)} = \frac{w(j)}{w(i)}.$$

This says that the weights of the forward and backward conditional transitions have the same ratio as the weights of the destination microstates.

In general microstate transitions are stochastic, and their probability is reflected in their weight. Of course saying that the transition is stochastic does not rule out the possibility that it has a deterministic component, or even that the random component may be a small perturbation on the deterministic component. In the event that the transition is fully deterministic, then the weight becomes a δ-function. In general the stochastic contribution to the microstate transitions of a sub-system comes from the interactions with the reservoirs, which are treated in projection. It may also come from internal motion of the microstates that is not treated explicitly. (Although the microstates are defined to be indivisible, this is a statement of how the system is to be treated mathematically rather than a statement of the physical reality.)

Continuing in the same vein as the equilibrium formulation, one defines the macrostate transition, $\alpha \xrightarrow{\tau} \beta$ (cf figure 3.1). These can belong to the same or to different collectives. The unconditional transition weight for this is

$$W(\beta, \alpha \mid \tau) = \sum_{i,j} w(j, i \mid \tau)\delta(A_i - A_\alpha)\delta(B_j - B_\beta)$$

$$= \sum_{i \in \alpha} \sum_{j \in \beta} w(j, i \mid \tau).$$

It ought to be clear that even if the microstate transitions were fully deterministic, the macrostate transitions must contain a stochastic element because specifying the initial macrostate does not fix the initial microstate that is required to determine the subsequent transition.

Figure 3.1. A transition in time τ from the α macrostate to the β macrostate of a collective.
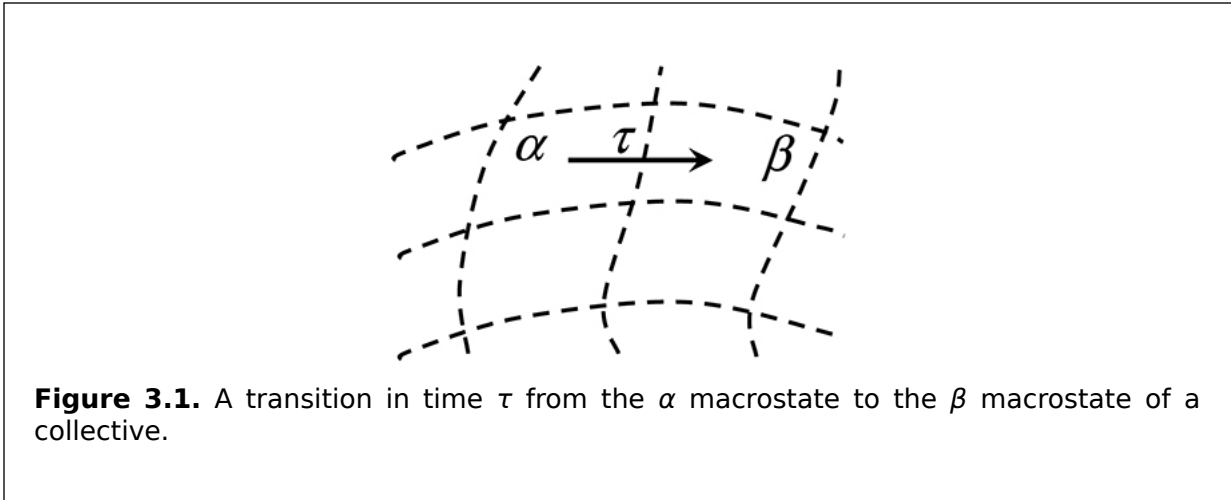
In equation (3.9), the weight of joint (i.e. simultaneous) macrostates was reduced to the weight of one of them by summing over the other, $\sum_\beta W(\alpha\beta) = W_\alpha$. The present transition weights can be taken to obey a similar reduction condition. We take it as a theorem (shortly proven by a time average) that for an equilibrium system the unconditional microstate weight is conserved in a microstate transition,

$$\sum_j w(j, i \mid \tau) = w(i) \quad \text{and} \quad \sum_i w(j, i \mid \tau) = w(j).$$

The first says that all the weight in microstate i is distributed without loss or gain amongst the possible target microstates j. (It is possible to scale the right-hand side by a positive constant, but this is only material in some non-equilibrium cases.) The second says that all the weight in the target microstates come from the possible initial microstates. The second is the same as the first with τ changed to $-\tau$. For the conditional microstate weight the conservation rule is equivalent to

$$\sum_j w(j \mid i, \tau) = 1.$$

This is plausible because the system must be in one and only one microstate at a time.

The proof of the conservation of weight during a transition follows from the axiom that time is uniform, at least in an equilibrium system. By this is meant that a time average over nodes of equal weight gives the same result as a statistical average using the microstate weights. Since the system must be in one and only one microstate at a time, one can define its trajectory in time through the microstates as $i(t)$. Let a label the nodes, $t_a = a\tau$, and let the ath microstate on the trajectory be $i_a \equiv i(a\tau)$. The microstate weight is proportional to the number of times the trajectory visits the microstate,

$$w(i) \propto n(i) = \sum_a \delta(i_a - i).$$

The unconditional transition weight is proportional to the number of times the transition occurs on the trajectory,

$$w(j, i \mid \tau) \propto n(j, i \mid \tau) = \sum_a \delta(i_a - i)\, \delta(i_{a+1} - j).$$

Summing over the destination microstates gives

$$\begin{aligned}
\sum_j n(j, i \mid \tau) &= \sum_j \sum_a \delta(i_a - i)\, \delta(i_{a+1} - j) \\
&= \sum_a \delta(i_a - i) \\
&= n(i).
\end{aligned}$$

The penultimate equality follows because the set of microstates is complete and disjoint (and hence non-degenerate). Hence the weight is conserved during a transition. Obviously one gets the second result in equation (3.17) by summing over i instead of j.

For certain time-varying non-equilibrium systems it is necessary to include a time-dependent scale factor in this conservation law (see chapter 6).

Each microstate i has a conjugate microstate $i^\dagger$. The physical interpretation of this is that in the conjugate state all the velocities of the particles in the system have their velocities reversed. (And also any velocity-derived external fields like magnetic fields or Coriolis forces, if present.) For an equilibrium system, the weight of a microstate and its conjugate are equal,

$$w(i^\dagger) = w(i).$$

Since reversing all the velocities is the same as reversing the direction of time, this says that time has no preferred direction in an equilibrium system. Note that in the sum over microstates, $\sum_i$, (for example, to obtain the total weight) both j and $j^\dagger$ each appear once.

Because reversing the velocities is the same as reversing time, one must have for the unconditional transition weight

$$w(j, i \mid \tau) = w(j^\dagger, i^\dagger \mid -\tau).$$

(3.23)

This might be called microscopic equilibrium reversibility. Combining these two with equation (3.15) yields

$$\frac{w(j \mid i, \tau)}{w(j^\dagger \mid i^\dagger, -\tau)} = \frac{w(i^\dagger)}{w(i)} = 1, \quad \text{or} \quad w(j \mid i, \tau) = w(j^\dagger \mid i^\dagger, -\tau).$$

(3.24)

This means that the forward transition has the same conditional weight as the backward transition between the conjugate microstates, This is called microscopic reversibility.

Microscopic reversibility may be confirmed explicitly for an isolated system evolving adiabatically with Hamilton's equations of motion. It holds also for stochastic dissipative transitions for an equilibrium system (see section 5.4.2).

Each macrostate α has a conjugate macrostate $\alpha^\dagger$. Consider the physical observable A, with A_α its value in the macrostate α. If A is an even function of the velocities (for example, energy, number etc), then $A_\alpha = A_{\alpha^\dagger}$. Since degenerate macrostates are precluded in the formalism (there is a one-to-one relationship between the macrostate labels and the values of the physical observable), this means that $\alpha = \alpha^\dagger$, which is to say that in this even parity case the macrostates are self-conjugate. If the physical observable is an odd function of the velocities, for example, the total momentum of the system, then $A_\alpha = -A_{\alpha^\dagger}$, and $\alpha \neq \alpha^\dagger$. If an observable is not of even parity (i.e. it is of odd or mixed parity), then both α and $\alpha^\dagger$ occur in the sum over macrostates. In view of this, whether the parity is even, odd, or mixed, the reverse transition between the conjugate macrostates has weight

$$\begin{aligned}
W(\alpha^\dagger, \beta^\dagger \mid \tau) &= \sum_{i^\dagger \in \alpha^\dagger} \sum_{j^\dagger \in \beta^\dagger} w(i^\dagger, j^\dagger \mid \tau) \\
&= \sum_{i \in \alpha} \sum_{j \in \beta} w(i^\dagger(i), j^\dagger(j) \mid \tau) \\
&= \sum_{i \in \alpha} \sum_{j \in \beta} w(i^\dagger, j^\dagger \mid \tau) \\
&= \sum_{i \in \alpha} \sum_{j \in \beta} w(i, j \mid -\tau) \\
&= W(\alpha, \beta \mid -\tau).
\end{aligned}$$

(3.25)

The second equality follows because for every $i^\dagger \in \alpha^\dagger$ there is an $i \in \alpha$ and for every $j^\dagger \in \beta^\dagger$ there is a $j \in \beta$. The third equality follows because $i^\dagger(i) = i^\dagger$ and $j^\dagger(j) = j^\dagger$. The fourth equality results from microscopic reversibility. The first and last equalities are just the definition of the macrostate transition weight. The derivation shows that macroscopic equilibrium reversibility, $W(\alpha^\dagger, \beta^\dagger \mid \tau) = W(\alpha, \beta \mid -\tau)$, is a direct consequence of microscopic equilibrium reversibility $w(i^\dagger, j^\dagger \mid \tau) = w(i, j \mid -\tau)$. As in that case it says that

the forward and reverse transitions between macrostates have the same conditional weight,

$$W(\alpha^\dagger \mid \beta^\dagger, \tau) = W(\alpha \mid \beta, -\tau).$$

(3.26)

If the macrostates have even parity, $\alpha^\dagger = \alpha$ and $\beta^\dagger = \beta$ then this becomes $W(\alpha, \beta \mid \tau) = W(\alpha, \beta \mid -\tau)$. This is a stronger result but it will seldom be required for what follows.

Macroscopic reversibility appears to contradict the Second Law of Thermodynamics. Macroscopic reversibility says that there is no preference between forward and reverse transitions, whereas the Second Law says that there is a preferred direction for time. This apparent contradiction is addressed below.

The conservation law for the microstate transition weight, equation (3.17) also follows for the macrostate transition weights,

$$\sum_\alpha W(\alpha, \beta \mid \tau) = W(\beta), \quad \text{and} \quad \sum_\beta W(\alpha, \beta \mid \tau) = W(\alpha).$$

(3.27)

From these conservations laws it follows that the total unconditional transition weight is just the total weight itself,

$$\sum_{i,j} w(j, i \mid \tau) = \sum_i w(i) = W.$$

(3.28)

This is independent of the time interval τ. Evidently the same result holds for the total macrostate transition weight,

$$\sum_{\alpha, \beta} W(\beta, \alpha \mid \tau) = \sum_\beta W(\beta) = W.$$

(3.29)

As in the formulation of entropy for the equilibrium case, one can define the second entropy as the logarithm of the transition weight,

$$S^{(2)}(\beta, \alpha \mid \tau) \equiv k_{\mathrm{B}} \ln W(\beta, \alpha \mid \tau).$$

(3.30)

There is an analogous expression for the microstate transitions. This could also be called the transition entropy, or the two-time entropy.

Obviously the unconditional transition probability is essentially the exponential of the second entropy,

$$\wp(\beta, \alpha \mid \tau) = \frac{W(\beta, \alpha \mid \tau)}{W} = \frac{1}{W} e^{S^{(2)}(\beta, \alpha \mid \tau)/k_{\mathrm{B}}}.$$

(3.31)

The conditional transition probability is the probability that the system will be in the macrostate β at a time τ, given that it is currently in the macrostate α. It is given by

$$\wp(\beta \mid \alpha, \tau) \equiv \frac{\wp(\beta, \alpha \mid \tau)}{\wp(\alpha)} = \frac{W(\beta, \alpha \mid \tau)}{W(\alpha)}.$$

(3.32)

If $\tau > 0$ this a forward transition, and if $\tau < 0$ this a backward transition.

The macroscopic reversibility condition on the transition weight, equation (3.25), yields a similar result for the unconditional transition probability,

$$\wp(\alpha^\dagger, \beta^\dagger \mid \tau) = \wp(\alpha, \beta \mid -\tau) = \wp(\beta, \alpha, \mid \tau).$$

(3.33)

The interpretation of this is that the forward transition between two macrostates is as likely to occur as the reverse transition between their conjugate states. Applying this to the conditional transition probability yields

$$\wp(\alpha^\dagger \mid \beta^\dagger, \tau) = \frac{\wp(\alpha^\dagger, \beta^\dagger \mid \tau)}{\wp(\beta^\dagger)} = \frac{\wp(\beta, \alpha \mid \tau)}{\wp(\beta)} = \wp(\beta \mid \alpha, \tau) \frac{\wp(\alpha)}{\wp(\beta)}.$$

(3.34)

This may be rearranged as

$$\frac{\wp(\alpha^\dagger \mid \beta^\dagger, \tau)}{\wp(\beta \mid \alpha, \tau)} = \frac{\wp(\alpha)}{\wp(\beta)} = e^{[S(\alpha) - S(\beta)]/k_{\mathrm{B}}}.$$

(3.35)

Recall that from microscopic reversibility, $\wp(\alpha^\dagger \mid \beta^\dagger, \tau) = \wp(\alpha \mid \beta, -\tau)$. One can conclude that the ratio of forward and reverse conditional transition probabilities equals the exponential of the entropy difference of the two macrostates.

The conditional transition probability in terms of the second entropy is

$$\wp(\beta \mid \alpha, \tau) = e^{\left[S^{(2)}(\beta, \alpha \mid \tau) - S(\alpha)\right]/k_{\mathrm{B}}}.$$

(3.36)

One could define the exponent to be the conditional second entropy, but this does not appear to be particularly useful.

A quantity that will prove of some interest is the most likely conditional transition state. In particular, given the current macrostate β, the most likely macrostate a time τ distant may be denoted $\bar{\alpha} \equiv \bar{\alpha}(\tau \mid \beta)$. This is the macrostate that maximizes the second entropy,

$$\frac{\partial S^{(2)}(\alpha, \beta \mid \tau)}{\partial \alpha}\Big|_{\alpha = \bar{\alpha}} = 0.$$

(3.37)

Maximizing the second entropy $S^{(2)}(\alpha, \beta \mid \tau)$ is the same as maximizing the transition weight $W(\alpha, \beta \mid \tau)$. This definition of $\bar{\alpha}$, together with the conservation condition on the transition weight, equation (3.27), can be used to 'reduce' the second entropy to the first or ordinary entropy. One has

$$\begin{aligned}
S(\beta) &= k_{\mathrm{B}} \ln W(\beta) \\
&= k_{\mathrm{B}} \ln \sum_\alpha W(\alpha, \beta \mid \tau) \\
&\approx k_{\mathrm{B}} \ln W(\bar{\alpha}(\tau \mid \beta), \beta \mid \tau) \\
&= S^{(2)}(\bar{\alpha}(\tau \mid \beta), \beta \mid \tau).
\end{aligned}$$

observe a fluctuation *from* equilibrium as to observe the reverse fluctuation *to* equilibrium, in an unconditional sense. In other words, an entropy increasing transition $\beta \xrightarrow{\tau} \alpha$, is just as unconditionally likely an entropy decreasing transition, $\alpha \xrightarrow{\tau} \beta$.

The assertion in the Second Law of Thermodynamics that spontaneous transitions increase the entropy refers to the conditional future transition probability from an unlikely (not-in-equilibrium) state. The Second Law does not apply to the unconditional transition probability. In general the Second Law is probabilistic, not deterministic, and so spontaneous transitions from the equilibrium macrostate to states of lower entropy in the conditional sense may well be unlikely but in general they are not impossible.

In macroscopic systems, fluctuations in the transitions are relatively negligible, and the transitions become effectively deterministic. Hence if the system is set up in a not-in-equilibrium macrostate, it will spontaneously evolve toward macrostates of higher entropy. If the system is set up in the equilibrium or most likely macrostate, it will stay there. In this macroscopic case the Second Law holds for future conditional transitions with probability approaching unity.

3.1.6 Two points on the arrow of time

'If as we follow the arrow we find more and more of the random element in the state of the world, then the arrow is pointing towards the future; if the random element decreases the arrow points towards the past.' (Eddington 1928)

There is an interesting twist on the arrow of time argument popularized by Eddington. For a system currently observed in a not-in-equilibrium macrostate, the usual formulation asks 'where will the system go to?'. But what if the obverse question were asked, namely 'from whence came the system?'. Since an answer is impossible if the previous states were set by some artificial unspecified process, one must assume that the current state arose by a spontaneous fluctuation of the system itself. In this case the answer is 'most likely the system came from a state of higher entropy'.

This answer is entirely consistent with the quantitative analysis above, where it can be noted that nothing was said about the sign of τ.

One must conclude that the usual formulation of the Second Law of Thermodynamics,

$$\text{All systems evolve to states with greater entropy,} \tag{3.41a}$$

is valid, and equally valid is its obverse,

$$\text{All systems have evolved from states with greater entropy.} \tag{3.41b}$$

Alternatively, the formulation

$$\text{The entropy increases during spontaneous transitions from the current state,} \tag{3.41c}$$

is as correct as

$$\text{The entropy decreases during spontaneous transitions to the current state.} \tag{3.41d}$$

These statements hold in a probabilistic sense (i.e. they are most likely true), they are conditional on the current state alone, and they hold for a system undergoing spontaneous fluctuations (i.e. without outside interference or specific preparation). From these alternative formulations it is clear that the notion that the arrow of time points in the direction of increasing entropy is a selective interpretation of the Second Law that entirely ignores history.

The fact of the matter is that the arrow of time is symmetric with regard to the change in entropy. If one defines a fixed arrow of time as pointing from what we call the past to the future, then entropy most likely increases in a forward transition to the future (i.e. in the direction of the arrow of time), and it most likely increases in a backward transition to the past (i.e. against the arrow of time). The present results contradict the quote from Eddington given at the beginning of this subsection. In view of this symmetry, one must conclude that there is no such thing as the arrow of time as defined by Eddington (1928).

3.2 Pure parity fluctuation analysis

3.2.1 Gaussian distribution

As was shown in section 2.5.3, for an equilibrium system the probability distribution for an extensive exchangeable parameter becomes infinitely sharply peaked in the thermodynamic limit. This means it can be accurately modeled as a Gaussian distribution, with the variance determined by the second derivative of the isolated system entropy.

In fact the prevalence of Gaussian distributions in physical systems is more general than this. Gaussian distributions are stable because any variable that is the sum of variables that are Gaussian distributed is itself Gaussian distributed. And if the variables are arbitrarily distributed, then by the central limit theorem their sum tends toward a Gaussian distribution as their number increases. Since the physical observables that define macrostates have multiple stochastic influences, it is no surprise that they have a Gaussian distribution.

Let X be a vector of extensive variables. The present analysis is simplified by taking all the components to be of the same parity, $X^{\dagger} = X$ or else $X^{\dagger} = -X$. That is, upon reversal of the velocities of all particles in the Universe, the values of the observables comprising X are either all unchanged or else all negated. The more general mixed parity case will be analyzed in section 3.3 below.

We are primarily interested in non-conserved, extensive variables, which is to say variables whose value can change with time in an isolated system. The spatial distribution of energy or number are examples. Furthermore the variable is defined such that its most likely or average value in the isolated system vanishes,

$$\bar{X} = 0.$$

(3.4 2)

For example, the difference in energy between the two halves of an isolated system most likely vanishes, because a uniform energy distribution is the one that maximizes the entropy. Obviously one can always subtract off any nonzero most likely value and interpret X as the departure from the most likely value. A nonzero value of X is called a fluctuation. Because X is extensive, higher order contributions to its fluctuations can be expected to be relatively negligible, and its probability distribution can be expected to be Gaussian.

The focus of this section is on an isolated system. The value of the non-conserved variable can evolve in time, $\dot{X}^{0}(t) \neq 0$, where t is time, the superscript dot signifies the time derivative, and the superscript 0 signifies the adiabatic (Hamiltonian) evolution. In this isolated system, $\bar{X} = 0$. An initial nonzero fluctuation, $X_0 \neq 0$, most likely evolves adiabatically back to zero, $\overline{X^0}(t \mid X_0) \to 0$, $t \to \infty$. (The evolution is not deterministic even in this isolated system because X_0 is a macrostate and the initial microstate conditioned on this macrostate is statistically distributed.) What we expect, and it remains to prove, is that there is a regime in which the rate of regression is constant and proportional to the initial fluctuation, $\overline{\dot{X}^{0}} = \tilde{\Lambda} X_0$.

Further, the formalism is designed to accommodate reservoirs that can exchange material related to X in such a way that the most likely value is nonzero, $\bar{X}(x_{\mathrm{r}}) \neq 0$, where x_{r} is a reservoir parameter. In this case one expects, and again it remains to prove, that the adiabatic rate of change of X in the presence of the reservoir is the same as if the nonzero value had arisen from a spontaneous fluctuation of the isolated system, $\overline{\dot{X}^0} = \tilde{\Lambda}\bar{X}(x_{\mathrm{r}})$. The arrangement of the reservoir induces a thermodynamic non-equilibrium system in a steady state, and this idea is one aspect of Onsager's regression hypothesis (Onsager 1931).

For the present no reservoirs are included so that the results pertain to an equilibrium isolated system. The time correlation matrix for X is defined to be

$$Q(\tau) \equiv k_{\mathrm{B}}^{-1}\langle X(\tau)\, X(0)\rangle.$$

(3.4 3)

The averand is a dyadic matrix; some authors write it as $X(\tau)\, X(0)^{\mathrm{T}}$, where the superscript T means transpose. The average is over the configurations of the isolated system. The auto- or self-correlation functions for the elements of X correspond to the diagonal elements of this matrix, and the cross-correlation functions correspond to the off-diagonal elements. Note the unconventional appearance of Boltzmann's constant in the definition of the time correlation matrix, which simplifies many of the following results.

An equilibrium system is homogeneous in time, which means that

$$\langle X(\tau)\, X(0)\rangle = \langle X(0)\, X(-\tau)\rangle, \quad \text{or} \quad Q(\tau) = Q(-\tau)^{\mathrm{T}}.$$

(3.4 4)

The condition of macroscopic reversibility is manifest as

$$\langle X(\tau)\, X(0)\rangle = \left\langle X(-\tau)^{\dagger}\, X(0)^{\dagger}\right\rangle$$
$$= \langle X(-\tau)\, X(0)\rangle, \quad \text{or} \quad Q(\tau) = Q(-\tau).$$

(3.4 5)

This result only holds for the present pure parity case. Together these two results say that the pure parity time correlation matrix is symmetric, $Q(\tau) = Q(\tau)^{\mathrm{T}}$, and even in the time interval.

The entropy of the isolated system may be expanded to second order about $X = 0$,

$$S(X) = S(0) + \frac{1}{2}S'' \colon XX.$$

(3.4 6)

The linear term vanishes because $X = 0$ is the most likely state, which is to say the state of maximum entropy, $S'(0) = 0$. One can discard the constant term $S(0)$ because we are only interested in the variation with X. The Taylor expansion can be terminated at the quadratic term, because higher order terms scale with successive higher powers of the inverse of the square root of the system size. The quadratic term is intensive on average, since $\langle XX\rangle \sim \mathcal{O}(V)$.

The second derivative matrix is

$$S'' \equiv \frac{\partial^2 S(X)}{\partial X \partial X}\Big|_{X=0}.$$

(3.4 7)

This can be called the fluctuation matrix, and it is obviously symmetric $S''^{\mathrm{T}} = S''$, since $\partial^2 S/\partial X_\alpha \partial X_\beta = \partial^2 S/\partial X_\beta \partial X_\alpha$. Since the entropy is a maximum at $X = 0$, the fluctuation matrix must be negative definite.

The probability distribution for the fluctuation is the exponential of the entropy. By the above expansion the exponent is a quadratic form, which is to say that the distribution is Gaussian

$$\wp(X) = \frac{1}{W}\,\mathrm{e}^{S'':XX/2k_{\mathrm{B}}}, \quad W = \sqrt{\mathrm{Det}\{-2\pi k_{\mathrm{B}} S''^{-1}\}}.$$

(3.48)

The negative of the inverse of the fluctuation matrix is the static correlation function,

$$Q_0 \equiv Q(0) = -S''^{-1}.$$

(3.49)

This is readily derived from its definition,

$$\begin{aligned}
Q(0) &= k_{\mathrm{B}}^{-1}\langle X(0)\,X(0)\rangle \\
&= \frac{1}{k_{\mathrm{B}} W}\int \mathrm{d}X\,XX\mathrm{e}^{S'':XX/2k_{\mathrm{B}}} \\
&= \frac{1}{k_{\mathrm{B}} W}\int \mathrm{d}X\,Xk_{\mathrm{B}}S''^{-1}\frac{\partial}{\partial X}\mathrm{e}^{S'':XX/2k_{\mathrm{B}}} \\
&= \frac{1}{W}\left\{\left[XS''^{-1}\mathrm{e}^{S'':XX/2k_{\mathrm{B}}}\right]_{-\infty}^{\infty} - S''^{-1}\int \mathrm{d}X\,\mathrm{e}^{S'':XX/2k_{\mathrm{B}}}\right\} \\
&= -S''^{-1}.
\end{aligned}$$

(3.50)

In the integration by parts, the integrated part is zero because extreme fluctuations at the limits of integration have vanishing probability.

In the same fluctuation spirit we shall expand the second entropy of the isolated system to quadratic order,

$$S^{(2)}(X', X \mid \tau) = \frac{1}{2}A(\tau) : XX + B(\tau) : X'X + \frac{1}{2}A'(\tau) : X'X'.$$

(3.51)

This is for the unconditional transition $X \xrightarrow{\tau} X'$. Here X is the current macrostate and X' is the destination macrostate. If $\tau > 0$ then X' is the future macrostate, and if $\tau < 0$ then X' is the past macrostate. The coefficient matrices are the second derivatives of the second entropy evaluated in the most likely state, $X = X' = 0$,

$$A(\tau) \equiv \frac{\partial^2 S^{(2)}(0, X \mid \tau)}{\partial X \partial X}\Big|_{X=0}, \quad A'(\tau) \equiv \frac{\partial^2 S^{(2)}(X', 0 \mid \tau)}{\partial X' \partial X'}\Big|_{X'=0},$$

and

$$B(\tau) \equiv \frac{\partial^2 S^{(2)}(X', X \mid \tau)}{\partial X \partial X'}\Big|_{X'=X=0}.$$

(3.52)

Clearly the first two matrices are symmetric, $A(\tau) = A(\tau)^{\mathrm{T}}$ and $A\prime(\tau) = A\prime(\tau)^{\mathrm{T}}$. The fluctuation second entropy has maximum value $S^{(2)}(0,0 \mid \tau) = 0$.

For the present equilibrium isolated system one has time homogeneity, which implies $S^{(2)}(X\prime, X \mid \tau) = S^{(2)}(X, X\prime \mid -\tau)$. This yields

$$A(-\tau) = A\prime(\tau), \quad \text{and} \quad B(\tau) = B(-\tau)^{\mathrm{T}}.$$

(3.53)

Macroscopic reversibility says that the second entropy must have the symmetry

$$S^{(2)}(X\prime, X \mid \tau) = S^{(2)}(X\prime^{\,\dagger}, X^{\dagger} \mid -\tau) = S^{(2)}(X\prime, X \mid -\tau).$$

(3.54)

The final equality, which only holds in the present pure parity case, can be confirmed by inspection of the quadratic form. It follows that

$$A(\tau) = A\prime(\tau) \quad \text{and} \quad B(\tau) = B(\tau)^{\mathrm{T}}.$$

(3.55)

Maximizing the second entropy with respect to $X\prime$ yields the most likely destination state given the present state,

$$\frac{\partial S^{(2)}(X\prime, X \mid \tau)}{\partial X\prime}\Big|_{X\prime = \overline{X\prime}} = 0.$$

(3.56)

Performing the derivative on the quadratic form yields the most likely destination explicitly,

$$\overline{X\prime} \equiv \bar{X}(\tau \mid X) = -A(\tau)^{-1} B(\tau) X.$$

(3.57)

The destination may be in the future, $\tau > 0$, or in the past, $\tau < 0$. If in the future, this result answers the question 'where will the system most likely go to?'. If in the past it answers the question 'where did the system most likely come from?'. In either case this result says that the most likely destination fluctuation is linearly proportional to the current fluctuation. This is a direct consequence of the Gaussian form for the fluctuations. Because the fluctuation form for the first or normal entropy is quadratic in the fluctuation, $S(X) = S\prime\prime: XX/2$, its gradient is linear in the fluctuation, $\nabla S(X) \equiv S\prime(X) = S\prime\prime X$. The gradient in the first entropy can be called the thermodynamic force. The fact that the most likely destination is linearly proportional to the fluctuation X is the same as saying that it is linearly proportional to the thermodynamic force. Hence the present result for the most likely destination provides the justification for an aspect of the regression hypothesis (Onsager 1931), namely that the rate of return to equilibrium is linearly proportional to the gradient in the entropy.

It is useful to rewrite the second entropy as a function of the departure from the most likely destination state,

$$S^{(2)}(X\prime, X \mid \tau) = \frac{1}{2} A(\tau) : \left[X\prime + A(\tau)^{-1} B(\tau) X\right]^2$$
$$+ \frac{1}{2}\left[A(\tau) - B(\tau)A(\tau)^{-1}B(\tau)\right] : XX.$$

Here and often below the notation of a vector squared is short-hand for the corresponding dyadic matrix.

Above the reduction condition on the second entropy was given, equation (3.38). It says that the second entropy for the transition to the most likely destination state equals the first entropy of the current state,

$$S^{(2)}(\overline{X\prime}, X \mid \tau) = S(X).$$

Physically, the reduction condition arises from the sharply peaked nature of the probability distribution, which means that fluctuations are relatively negligible. In the above form for the second entropy, the first term vanishes when $X\prime = \bar{X}(\tau \mid X)$, leaving only the final term that is quadratic in X. Hence by the reduction condition, the coefficient of this must be equal to the first entropy fluctuation matrix,

$$A(\tau) - B(\tau)A(\tau)^{-1}B(\tau) = S\prime\prime.$$

This is an exact relation between the two fluctuation transition matrices, the third matrix already having been eliminated, $A\prime(\tau) = A(\tau)$. Because this holds for all time intervals, it will prove to be the key equation in determining the coefficients of a small τ expansion of the fluctuation transition matrices.

The transition probability is the exponential of the second entropy,

$$\wp(X\prime, X \mid \tau) = \frac{1}{W^{(2)}} e^{S^{(2)}(X\prime, X \mid \tau)/k_{\mathrm{B}}}.$$

The exponent is a quadratic form, which means that the transition probability is a Gaussian, by design. Using the second entropy rearranged as above, and also the reduction condition, the normalization factor is readily seen to be

$$W^{(2)} = \sqrt{\mathrm{Det}\left\{-2\pi k_{\mathrm{B}} A(\tau)^{-1}\right\}} \sqrt{\mathrm{Det}\{-2\pi k_{\mathrm{B}} S\prime\prime^{-1}\}}.$$

For a Gaussian distribution, the average value equals the most likely value. This allows the easy evaluation of the time correlation matrix, and gives its relationship with the transition fluctuation matrices,

$$\begin{aligned}
Q(\tau) &= k_{\mathrm{B}}^{-1}\langle X\prime X\rangle \\
&= k_{\mathrm{B}}^{-1}\left\langle \overline{X\prime} X\right\rangle \\
&= -A(\tau)^{-1}B(\tau)k_{\mathrm{B}}^{-1}\langle XX\rangle \\
&= A(\tau)^{-1}B(\tau)S\prime\prime^{-1}.
\end{aligned}$$

Using these, the most likely destination, equation (3.57), can be rewritten in terms of the time correlation function,

$$\overline{X\prime} = -A(\tau)^{-1}B(\tau)X = -Q(\tau)S\prime\prime X.$$

The final two factors give the gradient in the first entropy, $S''X = S'(X) \equiv \nabla S(X)$. This is the thermodynamic force that drives the fluctuation back to its equilibrium value. This result is a general form of the regression hypothesis that holds for arbitrary time intervals τ. If one multiplies both sides of this by X^{T} on the right and takes the average then one obtains the identity,

$$\left\langle [\nabla S(X)] X^{\mathrm{T}} \right\rangle = -k_{\mathrm{B}} \mathrm{I}.$$

(3.65)

One can eliminate the two fluctuation transition matrices in favor of the time correlation matrix. From the reduction condition, equation (3.60), and the time correlation function expression equation (3.63), one obtains

$$A(\tau)^{-1} S'' = \mathrm{I} - A(\tau)^{-1} B(\tau) A(\tau)^{-1} B(\tau) = \mathrm{I} - Q(\tau) S'' Q(\tau) S''.$$

(3.66)

Hence

$$A(\tau) = [\mathrm{I} - S'' Q(\tau) S'' Q(\tau)]^{-1} S''.$$

(3.67)

This is evidently symmetric, as it must be. Taking the transpose of the time correlation expression equation (3.63), and using the result of macroscopic reversibility, $B(\tau)^{\mathrm{T}} = B(\tau)$, one also obtains

$$B(\tau) = [\mathrm{I} - S'' Q(\tau) S'' Q(\tau)]^{-1} S'' Q(\tau) S'' = A(\tau) Q(\tau) S''.$$

(3.68)

Hence the time correlation function matrix completely determines the fluctuation transition matrices. It follows that it also determines the second entropy and the transition probability.

3.2.2 Exponential Markovian decay

So far the analysis describes exactly fluctuations, with the coefficients having been related to the time correlation function. The general result for the most likely destination state, equation (3.64), is a form of the regression hypothesis, since it showed that this was linearly proportional to the current state. That result was valid for arbitrary time intervals. Now the aim is to elucidate the time dependence of the time correlation function and of the fluctuation transition matrices, which form the proportionality constant for the most likely destination.

The time correlation function can of course be obtained quantitatively and in molecular detail using statistical mechanics for any particular system. The aim of thermodynamics, however, is to abstract from any particular system and to obtain a general description of matter in terms of a few macroscopic parameters that can be measured, which aim is pursued here.

The starting point is Markovian systems, which in many cases are a good model of real world behavior. It will now be shown that the time correlation functions in these systems are exponentially decaying. Following that derivation, the small time limit will be analyzed, in which limit it will be shown that both Markovian and non-Markovian time correlation functions decay linearly. This limit turns out to be sufficient to develop the formalism of non-equilibrium thermodynamics of steady-state systems.

For the transition $X \xrightarrow{\tau} X'$, for sufficiently long-time intervals the current and the destination macrostate are uncorrelated. This means that the transition probability becomes the product of the probability of each fluctuation independently, and the second

entropy becomes the sum of the first entropies, $S^{(2)}(X\prime, X \mid \tau) \to S(X\prime) + S(X)$, $\tau \to \infty$. This means that the transition fluctuation matrices must behave as

$$A(\tau) \to S\prime\prime, \quad B(\tau) \to 0, \quad \tau \to \infty.$$

<div align="right">(3.69)</div>

These limiting expressions can also be derived from equations (3.67) and (3.68) using the fact that the time correlation function vanishes in the long-time limit, $Q(\tau) \to 0$, $\tau \to \infty$.

For small time intervals, the most likely destination state must be the same as the current state, $\bar{X}(\tau \mid X) \to X$, $\tau \to 0$. This must be so since it takes time to depart the current macrostate. Inserting this into equation (3.57) yields

$$A(\tau)^{-1} B(\tau) \to -\mathrm{I}, \quad \tau \to 0.$$

<div align="right">(3.70)</div>

Again, I is the identity matrix.

The present analysis is for fluctuations all with the same time parity. In this case equations (3.53) and (3.55) show that the fluctuation transition matrices are even functions of time, $A(\tau) = A(-\tau)$, and $B(\tau) = B(-\tau)$. From this one can see that the time correlation function is also an even function of time, $Q(\tau) = Q(-\tau)$, which in fact was shown in equations (3.44) and (3.45).

The results for $\tau \to \infty$ are formally exact. However this is not the limit that is required for the formal thermodynamic analysis of steady-state non-equilibrium systems. For these one has to ascertain the applicable time regime, and the qualitative behavior of the time correlation function therein. This can be done by looking at a specific system, and reasoning inductively that the behavior is universally applicable.

The regression of a fluctuation, $\bar{X}(\tau \mid X)$, is shown in figure 3.2. The particular macrostate illustrated is the first energy moment, $X \equiv E_1$, which is a non-conserved extensive variable that is the simplest measure of the non-uniformity of the energy distribution in the system. The negative values shown indicate that initially there is excess energy in the negative half of the system.
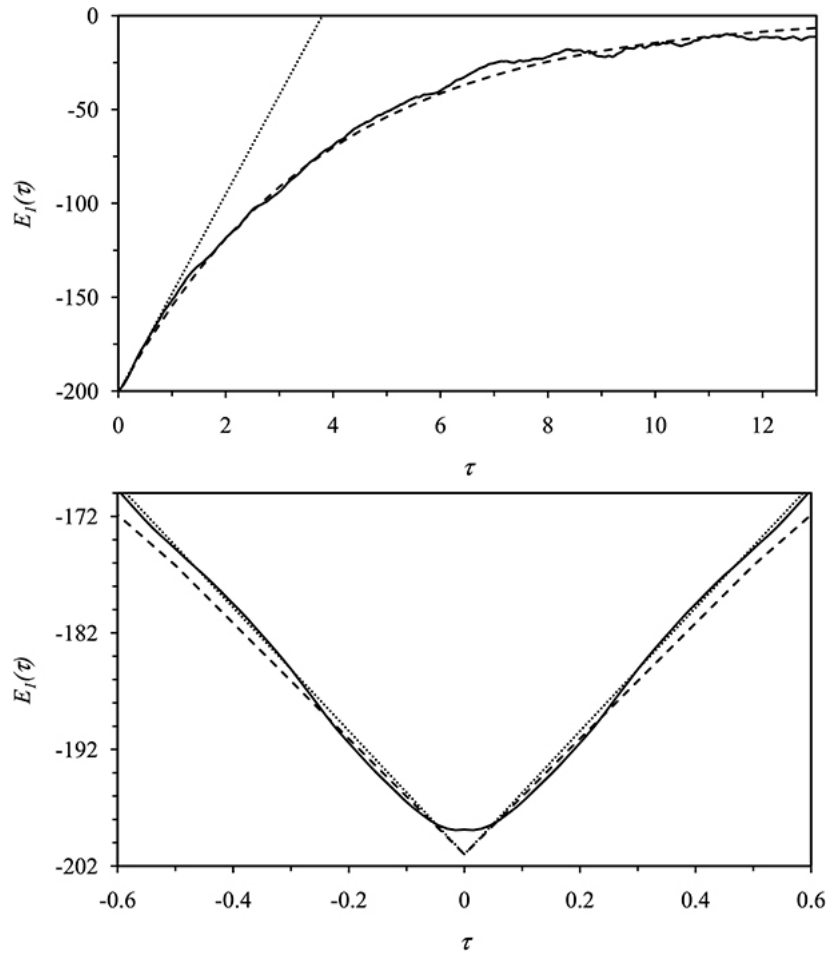
Figure 3.2. Adiabatic decay of the first energy moment (dimensionless units). The solid curve is the average over dynamically disordered initial states of molecular dynamics trajectories of a Lennard-Jones fluid. The dotted line is fitted, $E_1(\tau) = a + b \,|\, \tau \,|$, and the dashed curve is an exponential, $E_1(\tau) = a\mathrm{e}^{|\tau|b/a}$. The lower chart magnifies the data at small times. Data from Attard (2005a).

The behavior shown in figure 3.2 typify the regression of all fluctuations of the same time parity. In this case the regression is an even function of time. There are three more or less distinct time regimes. In the molecular or infinitesimal time regime, for the data of figure 3.2 this is $|\,\tau\,| \lesssim 0.1$, the fluctuation is smooth about $\tau = 0$, as can be seen in the lower part of the figure. Hence it has a Taylor expansion in even powers of τ, $\bar{X}(\tau \,|\, X) \sim X + c(X)\tau^2 + d(X)\tau^4 + \cdots.$

The small time regime is $0.1 \lesssim |\,\tau\,| \lesssim 1$ in figure 3.2, and it is identified by linear or steady decay of the fluctuation. Because of the pure parity, the regression must be an even function of time, which means that there must be non-analytic terms in its expansion in this regime, $\bar{X}(\tau \,|\, X) \sim a(X) + b(X) \,|\, \tau \,|$, with $a(X) \approx X$. Terms of order $\mathcal{O}(\tau^2, |\,\tau\,|^3)$ have been neglected here. This non-analytic expansion is the only expansion that is both linear in τ and even in τ. It cannot apply in the molecular regime, where the expansion

must be analytic, as the lower part of figure 3.2 shows. The non-analytic terms in the small-τ expansion sum an infinity of terms in the molecular regime Taylor expansion.

The long-time regime is here $\mid \tau \mid \gtrsim 1$. In this particular system the fluctuation regresses exponentially to zero, as the fit in figure 3.2 confirms. However such exponential behavior is not the case for all fluctuations, and sometimes the regression can be power law or even non-monotonic.

The molecular regime can be thought of as an inertial regime where the regression of the fluctuation comes up to speed from a standing start. The extent and nature of this regime varies with the type of fluctuation and the details of the intermolecular potential. More universal is the small time regime in which the rate of change of the fluctuation is constant. The actual rate of change is proportional to the magnitude of the initial fluctuation, with the proportionality constant depending upon the specific system. The latter are related to the hydrodynamic transport coefficients, as will shortly be clarified. This is the reason why the short time regime is the relevant regime for steady-state non-equilibrium thermodynamics. Since the magnitude of the fluctuation decreases with increasing time, so must the driving force and the rate of change. This sets an upper limit on the small time regime which is defined by a constant rate of change. That is, in the small time regime, the change in magnitude of the fluctuation must be relatively negligible, $\mid X(\tau) - X(0) \mid \ll \mid X(0) \mid$.

The exponential decay evident in figure 3.2 is an indicator of a Markov system. Such a system is defined by the present state solely determining the future state irrespective previous history of the system. In a Markov system the transition probability for three states at times $t_1 < t_2 < t_3$ factorizes

$$\wp(X_3, X_2, X_1 \mid t_{32}; t_{21}) = \wp(X_3 \mid t_{32}, X_2; t_{21}, X_1)\wp(X_2, X_1 \mid t_{21})$$
$$\approx \wp(X_3 \mid t_{32}, X_2)\wp(X_2, X_1 \mid t_{21}).$$

(3.71)

The first equality is the formally exact expression for the conditional probability: the probability of X_3 depends upon the preceding two states. The final equality is true if the intermediate state X_2 removes any influence of the state X_1 on the final state X_3. One can say that the final state has no memory of the earlier state. This is the Markov approximation

Markovian systems have a greatly simplified form of regression. The most likely transition $X_1 \to X_3$ in time t_{31} according to equation (3.64) is

$$\bar{X}_3 = -Q(t_{31})S\prime\prime X_1.$$

(3.72)

Similar to the analysis of Fox and Uhlenbeck (1970), one can divide the time interval with an intermediate time, $t_1 < t_2 < t_3$, so that the total transition is the product of a pair of successive transitions,

$$\bar{X}_3 = -Q(t_{32})S\prime\prime\bar{X}_2 = Q(t_{32})S\prime\prime Q(t_{21})S\prime\prime X_1.$$

(3.73)

The first equality invokes the Markov approximation: the transition $X_2 \to X_3$ is not influenced directly by the prior state X_1. For these two expressions to be equal one must have

$$-Q(t_{31})S\prime\prime = Q(t_{32})S\prime\prime Q(t_{21})S\prime\prime,$$

(3.74)

which implies that

$$Q(\tau) = -\mathrm{e}^{|\tau|\Lambda S \prime\prime/2} S \prime\prime^{-1}.$$

(3.7
5)

Here Λ is a constant matrix, and exponentiation is defined by the power law expansion. This is known as Doob's theorem, and it applies to any random process that is both Gaussian and Markov (Doob 1942). The absolute value of the time interval appears here because the time correlation function must be an even function of time in the present case of pure parity.

This general analysis proves the exponentially decay of the time correlation functions in a Markov systems. The decay rate is determined by Λ, which will turn out to be related to the transport coefficients of the system.

For small enough time intervals the exponential can be linearized,

$$Q(\tau) = -S\prime\prime^{-1} - \frac{|\tau|}{2}\Lambda, \quad |\tau|\,\mathrm{Det}\{\Lambda S\prime\prime\} \ll 1.$$

(3.7
6)

A more precise delineation of the regime of validity of this linearization would invoke the eigenvalue of $\Lambda S\prime\prime$ that is smallest in magnitude.

3.2.3 Small time expansion

The preceding result that the time correlation function of a Markov system was linear in $|\tau|$ on short time scales cannot hold on infinitesimal time scales. Instead in that regime the Taylor expansion, which is analytic, must involve only even powers of the time interval,

$$Q(\tau) = -S\prime\prime^{-1} + \frac{1}{2}\tau^2 \ddot{Q}(0) + \cdots, \quad \tau \to 0.$$

(3.7
7)

This and the preceding two equations give different expressions for the time dependence of the time correlation function. The expressions apply respectively in each of the three time regimes identified in figure 3.2.

The physical origin of these three time regimes is readily understood. Initially, $\tau = 0$, the system is positionally ordered, $X(0) \neq 0$, and dynamically disordered, $\dot{X}(0) = 0$. Over molecular time scales, say τ_{relax}, the system becomes dynamically ordered, $\dot{X}(\tau_{\mathrm{relax}}) \neq 0$. Even though dynamic order is a lower entropy state than dynamic disorder, we shall soon show that dynamic order that leads to increasing position disorder has greater second entropy than dynamic disorder.

The amount of dynamic order induced after this molecular time interval is proportional to the extent of the fluctuation, $\dot{X}(\tau) \propto X(\tau)$, $\tau > \tau_{\mathrm{relax}}$. This result is manifest in the exponential decay of $\bar{X}(\tau \mid X)$ in a Markov system. However the result is more general than Markov systems, both on symmetry grounds (the vector X is the only vector present, and so it must be the one that sets the direction of the vector $\dot{X}$, which is the so-called Curie principle), and on the grounds that for fluctuations higher order terms in X can be neglected.

For time intervals longer than the infinitesimal time τ_{relax}, but not too much longer, the changes in the fluctuation from its initial value are negligible, $X(\tau) \approx X(0)$, and so one has $\dot{X}(\tau) \propto X(0)$. Integrating this over this so-called short time regime leads to equation (3.76), where the absolute value of the time interval is required from the symmetry of the present pure parity system. We shall call the time intervals where this is valid short time intervals.

From the nature of the above argument, the short time expansion equation (3.76), is likely to be universally applicable. It is not restricted to the Markov systems that were used to derive it.

The following results apply to the small time regime. Expansions in powers of $|\tau|$ will now be obtained for the second entropy fluctuation matrices.

One can rearrange the fluctuation form of the second entropy as

$$
\begin{aligned}
S^{(2)}(X\prime, X \mid \tau) &= \frac{1}{2}A(\tau) : X^2 + B(\tau) : X\prime X + \frac{1}{2}A(\tau) : X\prime^2 \\
&= \frac{-1}{2}B(\tau) : [X\prime - X]^2 + \frac{1}{2}[A(\tau) + B(\tau)] : [X\prime^2 + X^2].
\end{aligned}
$$

(3.78)

This is for a pure parity system. This uses the inside-out convention for the double scalar product, $A : XY \equiv Y \cdot AX$, and writes the symmetric dyadic matrix as the square of a vector, $X^2 \equiv XX$.

As mentioned above, in the limit $\tau \to 0$ the destination state must be the same as the current state, $\overline{X\prime} = X$. Hence the transition probability must become a delta-function, $\wp(X\prime, X \mid \tau) \to \delta(X\prime - X)\wp(X)$, $\tau \to 0$. In view of this, the first term in the second equality of the second entropy above must diverge,

$$
B(\tau) = \frac{1}{|\tau|}\Lambda^{-1} + B + \mathcal{O}(\tau), \quad |\tau| \to 0.
$$

(3.79)

The constant matrix Λ is written to be consistent with the Markov analysis, as will soon be made clear explicitly. Since the second term in the second equality of the second entropy must remain finite, the leading term must cancel so that the expansion for $A(\tau)$ must be of the form

$$
A(\tau) = \frac{-1}{|\tau|}\Lambda^{-1} + A + \mathcal{O}(\tau), \quad |\tau| \to 0.
$$

(3.80)

To leading-order these two expansions give $A(\tau)^{-1}B(\tau) = -\mathrm{I}$, $|\tau| \to 0$, which result was already given as equation (3.70). By design, in the present case of pure parity, only terms even in time appear in these expansions.

Because the second entropy gives the transition probability, the quadratic form must be negative, which means that $A(\tau)$ must be negative definite, which means that Λ must be positive definite. Since $A(\tau)$ is symmetric, so is Λ.

Inserting these two expansions into the reduction condition, equation (3.60), yields

$$
\begin{aligned}
S\prime\prime &= A(\tau) - B(\tau)A(\tau)^{-1}B(\tau) \\
&= \frac{-1}{|\tau|}\Lambda^{-1} + A - \left[\frac{1}{|\tau|}\Lambda^{-1} + B\right]\left[\frac{-1}{|\tau|}\Lambda^{-1} + A\right]^{-1}\left[\frac{1}{|\tau|}\Lambda^{-1} + B\right] \\
&= 2[A + B] + \mathcal{O}(\tau).
\end{aligned}
$$

(3.81)

Only the leading order term, $\mathcal{O}(|\tau|^0)$ has been retained here, linear and higher order terms having been neglected. One sees from this that the sum of the two constant coefficient matrices in the small time expansion of the transition fluctuation matrices is

determined to equal half the first entropy fluctuation matrix. This will turn out to be a remarkably powerful result.

For a Markov system one has the stronger result, $B = 0$ and $A = S\prime\prime/2$. This is not required for what follows.

With this result one can give the explicit proof of the regression hypothesis. Given the current state X, the most likely destination state of a transition, equation (3.57), for short time intervals τ, is

$$
\begin{aligned}
\bar{X}(\tau \mid X) &= - A(\tau)^{-1} B(\tau) X \\
&= - \left[\frac{-1}{|\tau|} \Lambda^{-1} + A \right]^{-1} \left[\frac{1}{|\tau|} \Lambda^{-1} + B \right] X \\
&= [\mathrm{I} + |\tau| \Lambda (A + B)] X \\
&= X + \frac{|\tau|}{2} \Lambda S\prime\prime X + \mathcal{O}(\tau^2).
\end{aligned}
$$

(3.82)

This says that the most likely change or regression in a fluctuation is linear in time and linearly proportional to the thermodynamic force $\nabla S(X) \equiv S\prime(X) = S\prime\prime X$. The regression is symmetric in the future and in the past. This holds for systems of pure parity and is the general form of the regression hypothesis.

Multiplying both sides of this equation by $k_{\mathrm{B}}^{-1} X$ on the right, the average yields the time correlation function, $Q(\tau) = -S\prime\prime^{-1} - |\tau| \Lambda/2$. This is just the Markov result in the small time limit, equation (3.76).

For such a Markov system, one can simply piece together this result for a small time interval to obtain the result for larger time intervals, which is exponential decay of the time correlation function. For a non-Markov system such a procedure would be invalid, because successive time intervals would be correlated with each other, whereas the present second entropy expression neglects all conditions prior to the current state.

In the small time limit the second entropy is explicitly

$$
\begin{aligned}
S^{(2)}(X\prime, X \mid \tau) &= \frac{-1}{2|\tau|} \Lambda^{-1} : \left[X\prime - \bar{X}(\tau \mid X) \right]^2 + \frac{1}{2} S\prime\prime : X^2 \\
&= \frac{-1}{2|\tau|} \Lambda^{-1} : [X\prime - X]^2 + \frac{1}{2} S : X\prime X - \frac{|\tau|}{8} X \cdot S\prime\prime \Lambda S\prime\prime X \\
&= \frac{-1}{2|\tau|} \Lambda^{-1} : [X\prime - X]^2 + \frac{1}{4} S\prime\prime : \left[X\prime^2 + X^2 \right] - \frac{|\tau|}{8} X \cdot S\prime\prime \Lambda S\prime\prime X \\
&\quad + \mathcal{O}(\tau^2).
\end{aligned}
$$

(3.83)

This neglects $-B : [X\prime - X]^2/2$ and other terms of order τ^2. By design, the first equality satisfies the reduction condition exactly. That is, the maximum value of the second entropy with respect to the destination state given the current state equals the first entropy of the current state,

$$
\overline{S^{(2)}}(\tau \mid X) \equiv S^{(2)}(\overline{X\prime}, X \mid \tau) = \frac{1}{2} S\prime\prime : X^2.
$$

(3.84)

However, the final equality in equation ([3.83](#)) is more approximate in the sense that it satisfies the reduction condition to $\mathcal{O}(\tau)$.

3.2.4 Stochastic dissipative equations of motion

The difference between the actual destination and the most likely destination is the stochastic part of the transition,

$$X\prime - \bar{X}(\tau \mid X) = \tilde{R}.$$
(3.85)

Hence the transition in full is

$$X(t+\tau) = X(t) + \frac{\mid\tau\mid}{2}\Lambda S\prime\prime X(t) + \tilde{R}(t) + \mathcal{O}(\tau^2).$$
(3.86)

This is the stochastic dissipative equation of motion.

The second entropy in the small time limit given by the first equality in equation ([3.83](#)) is

$$S^{(2)}(X\prime, X \mid \tau) = S(X) - \frac{1}{2\mid\tau\mid}\Lambda^{-1} : \left[X\prime - \bar{X}(\tau \mid X)\right]^2.$$
(3.87)

Since $\wp(X\prime, X \mid \tau) = \wp(X\prime \mid X, \tau)\wp(X)$, and $\wp(X) = e^{S(X)/k_B}/Z^{(1)}$, in the small time limit the conditional transition probability is just the Gaussian

$$
\begin{aligned}
\wp(X\prime \mid X, \tau) &= \frac{1}{Z\prime}e^{-\Lambda^{-1}:\left[X\prime - \bar{X}(\tau\mid X)\right]^2/2k_B\mid\tau\mid} \\
&= \frac{1}{Z\prime}e^{-\Lambda^{-1}:\tilde{R}\tilde{R}/2k_B\mid\tau\mid} \\
&= \wp(\tilde{R}),
\end{aligned}
$$
(3.88)

where $Z\prime \equiv \left[\mathrm{Det}\left\{2\pi k_B \mid \tau \mid \Lambda\right\}\right]^{1/2}$. Since $\tilde{R} = X\prime - \bar{X}(\tau \mid X)$, the conditional probability distribution for $X\prime$ is equal to the probability distribution for $\tilde{R}$ because at constant X, $\mathrm{d}X\prime = \mathrm{d}\tilde{R}$. From this it follows that the stochastic change on average vanishes, $\left\langle\tilde{R}\right\rangle = 0$, and it has variance

$$\left\langle\tilde{R}\tilde{R}\right\rangle = \mid\tau\mid k_B\Lambda.$$
(3.89)

This last result is the fluctuation–dissipation theorem.

3.2.5 Regression theorem and the reciprocal relations

The small time expansion is all that is required to treat steady-state, thermodynamic, non-equilibrium systems, as will shortly be made clear. It also gives several related results that have been long known, albeit by different methods.

It has just been established that in the small time limit to linear order in τ, given the current state X, the most likely destination state is

$$\overline{X\prime} \equiv \bar{X}(\tau \mid X) = -Q(\tau)S\prime\prime X = X + \frac{1}{2} \mid \tau \mid \Lambda S\prime\prime X.$$

(3.9
0)

One may define the coarse velocity as the average velocity over the interval. Its most
likely value is

$$\overset{\circ}{X}(X) \equiv \frac{\overline{X\prime} - X}{\tau} = \frac{\hat{\tau}}{2}\Lambda S\prime\prime X.$$

(3.9
1)

Here and throughout the sign of the time interval is $\hat{\tau} \equiv \mathrm{sign}(\tau) = \tau/\mid \tau \mid$. This says
that the coarse velocity is proportional to the thermodynamic force,
$S\prime(X) \equiv \partial S(X)/\partial X = S\prime\prime X$.

This result is essentially the Onsager regression hypothesis: the most likely flux is
linearly proportional to the thermodynamic driving force (Onsager 1931). Of course the
present derivation makes it a theorem rather than an hypothesis. Flux is here the coarse

velocity $\overset{\circ}{X}$, as will become clearer in the explicit treatment of heat flow in section 3.4. The
'internal' thermodynamic driving force is here the gradient of the sub-system entropy. This
will be shown below to equal the 'external' thermodynamic driving force that is imposed by
the reservoirs.

An interesting feature of the present result is that the coarse velocity is independent of
the length of the time interval over which it is measured, which one would expect of any
velocity, but it does depend upon the sign of τ. The most likely flux reverses sign for the
future, $\hat{\tau} > 0$, compared to the past, $\hat{\tau} < 0$. This is an obvious consequence of a
fluctuation in an equilibrium system: the system will most likely return to equilibrium, and
the system most likely came from equilibrium.

The transport matrix Λ is the coefficient matrix for the leading order term in the small
time expansion of $A(\tau)$ and $B(\tau)$, equations (3.79) and (3.80). Since both of these are
symmetric the transport matrix itself must be symmetric

$$\Lambda = \Lambda^{\mathrm{T}}.$$

(3.9
2)

This says that the off-diagonal transport coefficients, which couple different forces and
fluxes, are equal. This result is just the reciprocal relations, which Onsager (1931) originally
explained in terms of macroscopic reversibility.

There are a number of comparatively well-known examples of reciprocal relations. One
simple example is multi-component isothermal diffusion. The regression theorem in this
case is equivalent to Fick's law, since it says that the mass fluxes are proportional to the
chemical potential gradients. There are two relevant diffusion constants for species α and
β. One gives the proportionality of the flux α to the chemical potential gradient β, and the
other gives the obverse, the proportionality of the flux β to the chemical potential gradient
α. Because the transport matrix is symmetric, these are equal to each other.

Another example is the thermoelectric effect, which couples electric and heat currents.
The Seebeck effect applies to an electric current driven by a temperature gradient, and the
Peltier effect applies to a heat flux driven by a voltage difference, or electro-chemical
gradient. The symmetry of the transport matrix Λ ensures that the respective
proportionality constants for these two effects are equal to each other.

Probably the most widely used example of a reciprocal relation, albeit one in which the
origin of the phenomenon is rarely acknowledged, is chemical reaction rates. Even
undergraduate students learn that the rate coefficient for the forward reaction is equal to
that of the reverse reaction. This result in fact can only be proven by the symmetry of the
transport matrix. This at that time unexplained example of a reciprocal relation was one of

the original motivations for Onsager's (1931) work for which he was subsequently awarded a Nobel prize.

Onsager's original derivation of the reciprocal relations consisted of combining the regression hypothesis (that the rate of regression of a fluctuation was proportional to the fluctuation itself) with macroscopic reversibility. In the present case the regression hypothesis has been derived as a theorem. Macroscopic reversibility was used in the present derivation to show that the cross fluctuation transition matrix is symmetric, $B(\tau) = B(\tau)^{\mathrm{T}}$, and that the two 'pure' fluctuation transition matrices are equal, $A(\tau) = A\prime(\tau)$. Equivalently, the time correlation function is symmetric, $Q(\tau) = Q(\tau)^{\mathrm{T}}$.

These symmetries, as well as the symmetry of the transport matrix, only hold for the pure parity case. Mixed parity is treated in section 3.3.

3.2.6 Green–Kubo expression

The Green–Kubo relations give the hydrodynamic transport coefficients as time integrals of the time correlation function of an equilibrium system (Green 1954, Kubo 1966, Kubo *et al* 1978, Keizer 1987, Zwanzig 2001). The above result for the small time expansion of the most likely destination equation (3.90), is equivalent to the Green–Kubo expressions, as is now shown. Since Onsager (1931) gave the relationship between the time correlation function and the transport coefficient in his paper on the reciprocal relations, it may be observed that he had given the Green–Kubo expressions some 20 years before either Green or Kubo.

The result for the most likely destination, equation (3.90), may be rewritten in Green–Kubo form as

$$
\begin{aligned}
\Lambda &= \frac{-2}{|\tau|}\left[Q(\tau) + S\prime\prime^{-1}\right] \\
&= \frac{-2}{k_{\mathrm{B}}\,|\tau|}\left[\langle X(\tau)X(0)\rangle - \langle X(0)X(0)\rangle\right] \\
&= \frac{-2}{k_{\mathrm{B}}\,|\tau|}\int_0^{\tau}\mathrm{d}t\prime\,\left\langle \dot{X}(t\prime)X(0)\right\rangle.
\end{aligned}
$$

(3.93)

For small but non-infinitesimal τ, where the present expansion is valid, the right-hand side must be independent of τ. In practice this defines the regime of validity for the small τ expansion. Typically simulations show the right-hand side has a broad peak as a function of τ, the maximum of which is generally taken as 'the' transport coefficient. The average is an equilibrium average, and it is typically taken for an isothermal rather than an isolated system. One can use integration by parts and various symmetry arguments to rewrite this is in several mathematically (but not necessarily computationally) equivalent ways.

3.2.7 Physical meaning of the second entropy

In order to provide a physical interpretation of the second entropy, it is useful to rewrite it in terms of the flux (i.e. the coarse velocity, $\overset{\circ}{X} \equiv (X\prime - X)/\tau$) and the transport matrix. Using equation (3.90), the departure from the most likely destination is

$$
X\prime - \overline{X\prime} = \tau\overset{\circ}{X} - \frac{|\tau|}{2}\Lambda S\prime\prime X + \mathcal{O}(\tau^2).
$$

(3.94)

Inserting this into the expression for the second entropy, the first equality in equation (3.83), to leading-order in τ one has

$$S^{(2)}(X\prime, X \mid \tau) = S(X) + \frac{1}{2}\left[\frac{-\Lambda^{-1}}{\mid \tau \mid} + \mathcal{O}(\tau^0)\right] : \left[X\prime - \bar{X}\prime\right]^2$$

$$= S(X) - \frac{\mid \tau \mid}{2}\Lambda^{-1} : \left[\overset{\circ}{X} - \overline{\overset{\circ}{X}}\right]^2 + \mathcal{O}(\tau^2)$$

$$= S(X) - \frac{\mid \tau \mid}{2}\Lambda^{-1} : \overset{\circ}{X}{}^2 + \frac{\tau}{2}\overset{\circ}{X} \cdot S\prime\prime X - \frac{\mid \tau \mid}{8}X \cdot S\prime\prime\Lambda S\prime\prime X + \mathcal{O}(\tau^2).$$

$$(3.9$$
$$5)$$

This expression writes the second entropy explicitly in terms of the coarse velocity; as a quadratic form, it gives the entropy due to the fluctuations in flux. By construction the flux, $\overset{\circ}{X}$, does not depend upon the length of the time interval, and as such it is an intensive function of time. In contrast, the second entropy as a linear function of the time interval is extensive in time.

3.2.7.1 Dynamic order versus structural disorder
Writing the second entropy in terms of the fluctuation in the flux allows for a straightforward physical interpretation. The four terms on the right-hand side of the final equality each contribute to the understanding of it.

   The first term is the ordinary entropy $S(X) = S\prime\prime : X^2/2$. Once the optimum flux $\overline{\overset{\circ}{X}}(X)$ has been established, this is all that remains of the second entropy. Given the optimum flux, maximizing $S(X)$ yields the optimum static structure $\bar{X}$. For the case of an isolated system this vanishes, but when exchange with a reservoir is allowed, the reservoir first entropy added to this term gives an optimum structure that is nonzero, $\bar{X}(x_{\mathrm{r}}) \neq 0$. This case will be analyzed below. The appearance of the first entropy is a direct consequence of the reduction condition, equation (3.38). On the basis of this discussion one can say that the first or ordinary entropy determines the static structure of a non-equilibrium system.
   Turning to the second term, one sees that it is quadratic in the flux. Since the transport matrix must be positive definite (cf the discussion following equation (3.80)), this term is negative and acts to reduce the second entropy. This is the term that prevents the flux becoming infinitely large, which is a crucial difference between the second entropy approach and the approaches based upon the principle of extreme dissipation (cf the discussion in section 3.1.2 and following equation (3.96)). The reason that the optimum flux must have an upper limit is that it represents a dynamically ordered state, which is a state of low entropy. Of course the challenge, and the point of the second entropy, is to balance quantitatively the unfavorable aspects of dynamic order with the favorable aspect of reducing the structural order over time.
   This brings us to the third term, which is proportional to the thermodynamic force, $\nabla S(X) = S\prime(X) = S\prime\prime X$. In fact it is proportional to the scalar product of the flux and the force, $\tau \overset{\circ}{X} \cdot \nabla S(X)/2$. In the future, $\tau > 0$, this term is positive when the flux is in the same direction as the force, which is the direction that increases the first entropy. (The case $\tau < 0$, and the meaning of the appearance of terms with τ and $\mid \tau \mid$ in the second entropy is discussed along with the irreversibility of the fluctuation below.) Because $S\prime\prime$ is negative definite, the direction of increasing first entropy is opposite to the direction of the fluctuation X, which is to say that the first entropy increases when the flux reduces X (going forward in time). This term is half the entropy dissipated over the time interval. It is this second term that drives the flux to be nonzero and that balances the entropic cost of the dynamic order that is the flux.

The final term does not appear to have a straightforward physical interpretation. Obviously it is negative and depends on the static structure, not the flux. Mathematically the reduction condition demands it so that the second and third terms can be canceled when the flux is optimum.

One can see that the physical basis for the variational principle embodied by the second entropy is rather simple, namely that the optimum flux arises from two competing effects. The (static) order of the system decreases over time as entropy is produced by the flux being aligned with the thermodynamic gradient. This term is favorable, linear in the flux, and dominates at small fluxes. Competing against this is the dynamic order of the system, which increases as the magnitude of the flux increases. This term unfavorable, quadratic in the flux, and dominates at large fluxes.

3.2.7.2 Dissipation

The variational nature of the second entropy stands in stark contrast to the dissipation, the extremization of which, it will be recalled from section 3.1.2, has been advocated by some authorities as the principle for determining the optimum non-equilibrium state. In general the rate of entropy production is a function of both the fluctuation and the flux,

$$\dot{S}(\dot{X}, X) = \dot{X} \cdot \frac{\partial S(X)}{\partial X} = \dot{X} \cdot S'' X.$$

(3.96)

In this formally exact expression, the fluctuation X and the flux $\dot{X}$ are independent variables.

According to the second entropy, for a given fluctuation X, the optimal rate of entropy production is therefore

$$\overset{\circ}{\dot{S}}(X) \equiv \dot{S}(\overset{\circ}{\dot{X}}, X) = \frac{\hat{\tau}}{2} X \cdot S'' \Lambda S'' X.$$

(3.97)

The optimal flux that appears here was determined by maximizing the second entropy, not by extremizing the dissipation. Since the transport matrix Λ is positive definite, the optimal dissipation is evidently positive. (The flux is generally measured going forward in time, $\hat{\tau} > 0$.)

The rate of entropy production, equation (3.96), is neither maximized nor minimized by the optimum flux. At a glance one can see from equation (3.96), taking a one-component system and $X > 0$ to make the point, that the maximum dissipation occurs at $\dot{X} = -\infty$, and the minimum dissipation occurs at $\dot{X} = +\infty$. A function that is linear in the flux has no non-trivial extrema; it can be a maximum or a minimum only at the boundaries of its range. It is the quadratic form of the second entropy that enables it to be used as a variational principle, which is in stark contrast to the dissipation. As was remarked in section 3.1.2, this is fundamentally why the dissipation does not give the variational principle for non-equilibrium systems.

3.2.7.3 Irreversible trajectory

As mentioned, the second entropy contains terms that depend on τ and on $|\tau|$. (Actually terms dependent on τ only occur in the mixed parity case; see section 3.3.) The former distinguishes between the future and the past, whereas the latter does not. It is these latter terms that make the trajectory irreversible.

Irreversibility means that if one goes backward from the most likely destination, most likely one does not regain the original state. If X is the original fluctuation then this reversal

operation corresponds to

$$\bar{X}(-\tau \mid \bar{X}(\tau \mid X)) = \left[\mathrm{I} + \frac{\mid \tau \mid}{2}\Lambda S'' \right]\left[\mathrm{I} + \frac{\mid \tau \mid}{2}\Lambda S'' \right] X$$
$$= X + \mid \tau \mid \Lambda S'' X + \mathcal{O}(\tau^2).$$

One sees that reversing the time step after the first step gives more of the same.

Consider the sequence of transitions $X_1 \xrightarrow{\tau} X_2 \xrightarrow{-\tau} X_3$. The initial state is X_1, the most likely intermediate state is $X_2 \equiv \bar{X}(\tau \mid X_1)$, and the most likely final state is $X_3 \equiv \bar{X}(-\tau \mid X_2)$. This says that in the future, $\tau > 0$ the system will move from X_1 toward equilibrium, so that $S(X_2) > S(X_1)$. However, the most likely prior state for X_2 is X_3 not X_1, since the former is even closer to equilibrium, $S(X_3) > S(X_2) > S(X_1)$. Reversibility and irreversibility is further discussed in section 5.4.2.

3.2.8 Third entropy

Although only the second entropy is necessary to deal with steady-state, thermodynamic, non-equilibrium systems, it is of interest to discuss the third entropy because this shows how non-Markovian behavior can arise. Since Gaussian fluctuations of all orders are the sum of pair-wise products, the third entropy in its most general form is

$$S^{(3)}(X_3, X_2, X_1 \mid t_{32}, t_{21}) = \frac{1}{2}C : X_3^2 + \frac{1}{2}C' : X_1^2 + \frac{1}{2}D : X_2^2$$
$$+ E : X_2 X_3 + E' : X_2 X_1 + F : X_1 X_3.$$

For simplicity the dependence on the time intervals of the fluctuation transition matrices has been suppressed. In the symmetric case that $t_{32} = t_{21}$, the primed and unprimed matrices are equal. By rearranging this in terms of $(X_3 - X_2)^2$ and $(X_2 - X_1)^2$, in the limit $t_{31} \to 0$ one would conclude that to leading order the fluctuation transition matrices C, D, and E would go like $\mid t_{21} \mid^{-1}$. The fluctuation transition matrix F contains the non-Markovian behavior. It goes like $\mathcal{O}(\mid t_{21} \mid^0)$ to leading order.

This matrix is zero in a Markovian system, $F = 0$. In this case the third entropy is just the sum of the consecutive second entropies,

$$S^{(3)}(X_3, X_2, X_1 \mid t_{32}, t_{21}) = S^{(2)}(X_3, X_2 \mid t_{32}) + S^{(2)}(X_2, X_1 \mid t_{21}) - S(X_2).$$

The final term cancels the first entropy that is counted twice for the intermediate state.

3.2.9 Beyond fluctuations

The above analysis was for fluctuations in an isolated system, and for these the quadratic forms for the first and second entropy are valid. One can expect that the fluctuations in an extensive variable are relatively small. For heat flow below, in section 3.4.4 we shall treat the case of energy exchange with heat reservoirs. This raises the general possibility of how to treat the problem when the relevant variable X is not relatively small.

Suppose that the reservoirs impose a nonzero most likely value, $\bar{X}(x_\mathrm{r})$. The fluctuation about this may be denoted $\delta X = X - \bar{X}(x_\mathrm{r})$. This fluctuation may be considered small, and the first and second entropy may be expanded to quadratic order in it. All the above theory now holds with $X \Rightarrow \delta X$. The first entropy fluctuation matrix S'' is evaluated at

$X = \bar{X}(x_{\mathrm{r}})$ rather than at $X = 0$, and similarly for the second entropy fluctuation transition matrices $A(\tau)$ and $B(\tau)$. The time correlation function is define in terms of this fluctuation,

$$Q(\tau) = k_{\mathrm{B}}^{-1} \langle \delta X(\tau)\, \delta X(0) \rangle_{x_{\mathrm{r}}}.$$

(3.1
01)

3.3 Mixed parity fluctuations

3.3.1 Second entropy

Up to now fluctuations of pure time parity have been treated, which is to say either $X^{\dagger} = X$, or else $X^{\dagger} = -X$. (Recall that the dagger denotes the macrostate with the same molecular positions but with the molecular velocities reversed.) Now our attention turns to the mixed parity case, where the components of X each have pure parity, but they do not all necessarily have the same parity. This is not a serious restriction because any mixed parity macrostate can be decomposed into its pure parity components, $X_{\alpha}^{\pm} = [X_{\alpha} \pm X_{\alpha}^{\dagger}]/2$, and these individually can be included in X.

The diagonal parity matrix ε is defined with elements $\varepsilon_{ij} = \pm\delta_{ij}$ such that

$$X^{\dagger} = \varepsilon X,$$

(3.1
02)

Obviously $\varepsilon^2 = \mathrm{I}$, and $X_i^{\dagger} = \varepsilon_{ii} X_i$.

In general the time correlation function is the equilibrium average of the dyadic matrix of the fluctuations, $Q(\tau) \equiv k_{\mathrm{B}}^{-1} \langle X(\tau) X(0) \rangle$. The time homogeneity of an equilibrium system holds as well for mixed parity systems as it does for pure parity systems,

$$\langle X(\tau) X(0) \rangle = \langle X(0) X(-\tau) \rangle, \quad \text{or} \quad Q(\tau) = Q(-\tau)^{\mathrm{T}}.$$

(3.1
03)

However, macroscopic reversibility in the mixed parity case reads,

$$\langle X(\tau) X(0) \rangle = \Big\langle X(-\tau)^{\dagger} X(0)^{\dagger} \Big\rangle, \quad \text{or} \quad Q(\tau) = \varepsilon Q(\tau)^{\mathrm{T}} \varepsilon.$$

(3.1
04)

The first entropy in fluctuation approximation is unchanged,

$$S(X) = \frac{1}{2} S\!\prime\!\prime \!: XX,$$

(3.1
05)

again with the fluctuation matrix related to the time correlation matrix,

$$S\!\prime\!\prime \equiv \frac{\partial^2 S(X)}{\partial X \partial X}\Big|_{X=0} = -k_{\mathrm{B}} \langle XX \rangle^{-1} = -Q(0)^{-1}.$$

(3.1
06)

Macroscopic reversibility at $\tau = 0$ reads $Q(0) = \varepsilon Q(0)\varepsilon$, which means that

$$\varepsilon S\!\prime\!\prime \varepsilon = S\!\prime\!\prime, \quad \text{or} \quad \varepsilon S\!\prime\!\prime = S\!\prime\!\prime \varepsilon.$$

(3.1
07)

Macroscopic reversibility at $\tau = 0$, means the coupling between variables of different time parity instantaneously vanishes,

$$\langle X_i(0)X_j(0)\rangle = 0, \quad \text{if } \varepsilon_{ii} \neq \varepsilon_{jj}. \tag{3.108}$$

From this it follows that $\{S\prime\prime^{-1}\}_{ij} = 0$ if $\varepsilon_{ii}\varepsilon_{jj} = -1$. This result says that if all variables of the same parity are grouped together then the time correlation matrix is 'block' diagonal at $\tau = 0$, as is the fluctuation matrix. This result does not imply that variables of different parity are uncorrelated for all times.

The quadratic form for the second entropy in this mixed parity case can most generally be written as

$$S^{(2)}(X\prime, X \mid \tau) = \frac{1}{2}X \cdot A(\tau)X + X \cdot B(\tau)X\prime + \frac{1}{2}X\prime{\cdot}C(\tau)X\prime. \tag{3.109}$$

Here $X \equiv X(0)$ is the current state, and $X\prime \equiv X(\tau)$ is the destination state. The matrices $A(\tau)$ and $C(\tau)$ are second derivative matrices of the same variable,

$$A(\tau) \equiv \frac{\partial^2 S^{(2)}(0, X \mid \tau)}{\partial X \partial X}\Big|_{X=0}, \quad \text{and} \quad C(\tau) \equiv \frac{\partial^2 S^{(2)}(X\prime, 0 \mid \tau)}{\partial X\prime \partial X\prime}\Big|_{X\prime=0}. \tag{3.110}$$

These are clearly symmetric, $A(\tau) = A(\tau)^{\mathrm{T}}$ and $C(\tau) = C(\tau)^{\mathrm{T}}$. The matrix $B(\tau)$ is the cross second derivative,

$$B(\tau) \equiv \frac{\partial^2 S^{(2)}(X\prime, X \mid \tau)}{\partial X \partial X\prime}\Big|_{X\prime=X=0}. \tag{3.111}$$

The second entropy must be negative $S^{(2)}(X\prime, X \mid \tau) \leqslant S^{(2)}(0, 0 \mid \tau) = 0$. This follows because fluctuations are unfavorable, and constant contributions have been set to zero. This means that the matrices $A(\tau)$ and $C(\tau)$ must be negative definite, and the double matrix formed from the three coefficient matrices must also be negative definite.

The second entropy has similar symmetries to the time correlation function. From time homogeneity, $S^{(2)}(X\prime, X \mid \tau) = S^{(2)}(X, X\prime \mid -\tau)$, it follows that

$$A(-\tau) = C(\tau), \quad \text{and} \quad B(\tau) = B(-\tau)^{\mathrm{T}}. \tag{3.112}$$

Macroscopic reversibility applied to the second entropy reads $S^{(2)}(X\prime, X \mid \tau) = S^{(2)}(X\prime^{\dagger}, X^{\dagger} \mid -\tau) = S^{(2)}(\varepsilon X\prime, \varepsilon X \mid -\tau)$. Hence the fluctuation transition matrices have the symmetries

$$\varepsilon A(\tau)\varepsilon = C(\tau), \quad \text{and} \quad \varepsilon B(\tau)\varepsilon = B(\tau)^{\mathrm{T}}. \tag{3.113}$$

The most likely destination state given the present state maximizes the second entropy over $X\prime$. This gives

$$\overline{X\prime} \equiv \bar{X}(\tau \mid X) = -C(\tau)^{-1}B(\tau)^{\mathrm{T}}X.$$

Similarly, the most likely initial state that would lead to the destination $X\prime$ maximizes the second entropy over X,

$$\bar{X} \equiv \bar{X}(-\tau \mid X\prime) = -A(\tau)^{-1}B(\tau)X\prime.$$

This can also be written $\bar{X}(-\tau \mid X\prime) = -C(-\tau)^{-1}B(-\tau)^{\mathrm{T}}X\prime.$

It is useful to rewrite the second entropy in terms of the departure from the most likely destination state,

$$S^{(2)}(X\prime, X \mid \tau) = \frac{1}{2}\left[X\prime + C(\tau)^{-1}B(\tau)^{\mathrm{T}}X\right]^{\mathrm{T}}C(\tau)\left[X\prime + C(\tau)^{-1}B(\tau)^{\mathrm{T}}X\right]$$
$$+ \frac{1}{2}X^{\mathrm{T}}\left[A(\tau) - B(\tau)C(\tau)^{-1}B(\tau)^{\mathrm{T}}\right]X.$$

Alternatively, it can be written in terms of the departure from the most likely initial state,

$$S^{(2)}(X\prime, X \mid \tau) = \frac{1}{2}\left[X + A(\tau)^{-1}B(\tau)X\prime\right]^{\mathrm{T}}A(\tau)\left[X + A(\tau)^{-1}B(\tau)X\prime\right]$$
$$+ \frac{1}{2}X\prime^{\mathrm{T}}\left[C(\tau) - B(\tau)^{\mathrm{T}}A(\tau)^{-1}B(\tau)\right]X\prime.$$

The reduction condition, equation (3.38), for the present mixed parity case reads

$$S^{(2)}(\bar{X}\prime, X \mid \tau) = S(X), \text{ or } S^{(2)}(X\prime, \bar{X} \mid \tau) = S(X\prime).$$

It follows from the preceding expressions that

$$A(\tau) - B(\tau)C(\tau)^{-1}B(\tau)^{\mathrm{T}} = S\prime\prime,$$

and

$$C(\tau) - B(\tau)^{\mathrm{T}}A(\tau)^{-1}B(\tau) = S\prime\prime.$$

These two expressions are equivalent to each other, since one can be turned into the other by multiplying before and after by the parity matrix, and applying macroscopic reversibility, equation (3.113). This result of the reduction condition may be compared to the result in the pure parity case, equation (3.60), in which case $A(\tau) = C(\tau)$ and $B(\tau) = B(\tau)^{\mathrm{T}}$.

The transition fluctuation matrices can be related to the time correlation matrix by invoking the fact that for Gaussian statistics, the average value and the most likely value are the same. Hence the time correlation function is given by

$$Q(\tau) = k_{\mathrm{B}}^{-1}\langle X'X\rangle$$
$$= k_{\mathrm{B}}^{-1}\left\langle \overline{X'}X\right\rangle$$
$$= -C(\tau)^{-1}B(\tau)^{\mathrm{T}}k_{\mathrm{B}}^{-1}\langle XX\rangle$$
$$= C(\tau)^{-1}B(\tau)^{\mathrm{T}}S''^{-1}.$$

(3.121)

Similarly, $Q(\tau) = S''^{-1}B(\tau)^{\mathrm{T}}A(\tau)^{-1}$.

Using these and the reduction condition, equation (3.119), the transition fluctuation matrices can each be expressed of the time correlation matrix and the entropy fluctuation matrix. One can readily show that

$$A(\tau) = \left[\mathrm{I} - S''Q(\tau)^{\mathrm{T}}S''Q(\tau)\right]^{-1}S''.$$

(3.122)

This reduces to the pure parity result, equation (3.67), when the time correlation function matrix is symmetric.

Inserting this into the second form for the time correlation function, one obtains for the cross fluctuation matrix

$$B(\tau) = \left[\mathrm{I} - S''Q(\tau)^{\mathrm{T}}S''Q(\tau)\right]^{-1}S''Q(\tau)^{\mathrm{T}}S'' = A(\tau)Q(\tau)^{\mathrm{T}}S''.$$

(3.123)

As in the pure parity case, the time correlation function completely determines the fluctuation transition matrices, the second entropy, and the transition probability. In particular, the most likely destination, equation (3.114), may be written

$$\bar{X}(\tau \mid X) = -Q(\tau)S''X.$$

(3.124)

This is formally unchanged from the result in the case of pure parity, equation (3.64) (although of course $Q(\tau)$ is not now symmetric). The term $S''X = S'(X) \equiv \nabla S(X)$ remains the thermodynamic driving force toward the equilibrium state. Again, as in the pure parity case, multiplying by X^{T} on the right, the subsequent average yields an identity, since $\langle [\nabla S(X)]X\rangle = -k_{\mathrm{B}}\mathrm{I}$.

3.3.2 Small time expansion

The difference between the small time expansion for the previous pure parity systems and the present mixed parity system, is that the time correlation function is no longer an even function of time. This means that now both τ and $\mid \tau \mid$ appear in the expansion. Because of this, we shall need the sign of the time interval, $\hat{\tau} \equiv \mathrm{sign}\ \tau \equiv \tau / \mid \tau \mid$.

The pure parity expansions (3.79) and (3.80) carry over directly to the present mixed parity case by including additional terms,

$$A(\tau) = C(-\tau) = \frac{A_{-1}}{\mid \tau \mid} + \frac{A'_{-1}}{\tau} + A_0 + A'_0\hat{\tau} + \mathcal{O}(\tau),$$

(3.125)

and

$$B(\tau) = \frac{B_{-1}}{|\tau|} + \frac{B'_{-1}}{\tau} + B_0 + B'_0\hat{\tau} + \mathcal{O}(\tau).$$

Since $A(\tau) = A(\tau)^{\mathrm{T}}$, all the coefficient matrices that appear in its expansion are symmetric. Since $B(-\tau) = B(\tau)^{\mathrm{T}}$, in its expansion the unprimed coefficient matrices are symmetric, and the primed coefficient matrices are antisymmetric.

As in the pure parity case, equation (3.70), since $\bar{X}(\tau \mid X) \to X$, $\tau \to 0$ still holds, one must still have that $A(\tau)^{-1}B(\tau) \to -\mathrm{I}$, $\tau \to 0$. Hence the leading-order coefficients satisfy

$$B_{-1} + \hat{\tau}B'_{-1} = -A_{-1} - \hat{\tau}A'_{-1}.$$

Now A'_{-1} is symmetric, whereas B'_{-1} is antisymmetric. Since the two parts of this equation must separately be equal, it follows that $A'_{-1} = B'_{-1} = 0$. With this the expansions become

$$A(\tau) = \frac{-\Lambda^{-1}}{|\tau|} + A_0 + A'_0\hat{\tau} + \mathcal{O}(\tau),$$

and

$$B(\tau) = \frac{\Lambda^{-1}}{|\tau|} + B_0 + B'_0\hat{\tau} + \mathcal{O}(\tau).$$

As in the pure parity case, because $A(\tau)$ is symmetric and negative definite, the transport matrix must be symmetric and positive definite.

The two remaining symmetric coefficient matrices A_0 and B_0, can be related to each other by invoking the reduction condition. Expanding this to zeroth order in the time interval one has

$$
\begin{aligned}
S'' &= A(\tau) - B(\tau)C(\tau)^{-1}B(\tau)^{\mathrm{T}} \\
&= \left[-\Lambda^{-1}|\tau|^{-1} + A_0 + A'_0\hat{\tau}\right] - \left[\Lambda^{-1}|\tau|^{-1} + B_0 + B'_0\hat{\tau}\right] \\
&\quad \times \left[-\Lambda^{-1}|\tau|^{-1} + A_0 - A'_0\hat{\tau}\right]^{-1}\left[\Lambda^{-1}|\tau|^{-1} + B_0 - B'_0\hat{\tau}\right] \\
&= \left[-\Lambda^{-1}|\tau|^{-1} + A_0 + A'_0\hat{\tau}\right] - \left[\Lambda^{-1}|\tau|^{-1} + B_0 + B'_0\hat{\tau}\right] \\
&\quad \times \left[-\mathrm{I} + |\tau|\Lambda A_0 - \tau\Lambda A'_0\right]^{-1}\left[\mathrm{I} + |\tau|\Lambda B_0 - \tau\Lambda B'_0\right] \\
&= \left[-\Lambda^{-1}|\tau|^{-1} + A_0 + A'_0\hat{\tau}\right] - \left[\Lambda^{-1}|\tau|^{-1} + B_0 + B'_0\hat{\tau}\right] \\
&\quad \times \left[-\mathrm{I} + |\tau|\Lambda(A_0 + B_0) - \tau\Lambda(A'_0 + B'_0) + \mathcal{O}(\tau^2)\right] \\
&= \left[-\Lambda^{-1}|\tau|^{-1} + A_0 + A'_0\hat{\tau}\right] + \left[\Lambda^{-1}|\tau|^{-1} + A_0 + 2B_0 - A'_0\hat{\tau} + \mathcal{O}(\tau)\right] \\
&= 2(A_0 + B_0) + \mathcal{O}(\tau).
\end{aligned}
$$

The time correlation function may also be expanded for small times,

$$Q(\tau) = C(\tau)^{-1} B(\tau)^{\mathrm{T}} S\!/\!/^{-1}$$
$$= \left[-\Lambda^{-1} \mid \tau \mid^{-1} + A_0 - A_0' \hat{\tau} \right]^{-1} \left[\Lambda^{-1} \mid \tau \mid^{-1} + B_0 - B_0' \hat{\tau} \right] S\!/\!/^{-1}$$
$$= [\mathrm{I} - \mid \tau \mid \Lambda A_0 + \tau \Lambda A_0']^{-1} [-\mathrm{I} - \mid \tau \mid \Lambda B_0 + \tau \Lambda B_0'] S\!/\!/^{-1}$$
$$= [-\mathrm{I} - \mid \tau \mid \Lambda (A_0 + B_0) + \tau \Lambda (A_0' + B_0')] S\!/\!/^{-1}$$
$$= -S\!/\!/^{-1} - \frac{\mid \tau \mid}{2} \Lambda + \tau \Theta + \mathcal{O}(\tau^2).$$

<div align="right">(3.1 31)</div>

The final equality defines Θ, the coefficient matrix of τ. Since $Q(-\tau) = Q(\tau)^{\mathrm{T}}$, this must be antisymmetric, which yields

$$\Theta \equiv \Lambda (A_0' + B_0') S\!/\!/^{-1} = -S\!/\!/^{-1} (A_0' - B_0') \Lambda.$$

<div align="right">(3.1 32)</div>

The fluctuation transition coefficient matrices that couple to $\hat{\tau}$ can be written in terms of the antisymmetric transport matrix,

$$A_0' = \frac{1}{2} \left[\Lambda^{-1} \Theta S\!/\!/ - S\!/\!/ \Theta \Lambda^{-1} \right], \quad \text{and} \quad B_0' = \frac{1}{2} \left[\Lambda^{-1} \Theta S\!/\!/ + S\!/\!/ \Theta \Lambda^{-1} \right].$$

<div align="right">(3.1 33)</div>

In view of equation (3.131), one can *define* 'the' transport matrix in the mixed parity case as $L(\hat{\tau}) \equiv \Lambda - 2\hat{\tau}\Theta$. This is defined in this way to retain the same proportionality between flux and thermodynamic force as in the pure parity case, as will shortly be made explicit. The transport matrix $L(\hat{\tau})$ consists of a symmetric and an antisymmetric matrix. These give an irreversible and a reversible contribution to the trajectory, respectively. The antisymmetric matrix changes sign if the transport matrix is used to predict the past rather than the more usual case of predicting the future. This is a real physical effect.

The time correlation matrix has the symmetry $Q(-\tau) = \varepsilon Q(\tau) \varepsilon$, or, in component form, $Q_{ij}(-\tau) = \varepsilon_{ii} \varepsilon_{jj} Q_{ij}(\tau)$. From the small time expansion, equation (3.131), it follows that the symmetric part of the transport matrix has elements that vanish for states of different parity,

$$\Lambda_{ij} = 0 \quad \text{if } \varepsilon_{ii} \varepsilon_{jj} = -1.$$

<div align="right">(3.1 34)</div>

We have already seen the equivalent result that $\{ S\!/\!/^{-1} \}_{ij} = 0$ if $\varepsilon_{ii} \varepsilon_{jj} = -1$. It also follows that the antisymmetric part of the transport matrix has elements that vanish for states of the same parity,

$$\Theta_{ij} = 0 \quad \text{if } \varepsilon_{ii} \varepsilon_{jj} = 1.$$

<div align="right">(3.1 35)</div>

In view of this behavior, it is often convenient to group the components of X of the same parity together. In such a representation the various coefficient matrices subdivide into four square blocks, with those linking components of like parity on the main diagonal, and those linking unlike parity components on the off-diagonal. One sees that in this representation, the matrices Λ, $S\!/\!/^{-1}$ (and hence $S\!/\!/$) are symmetric and block diagonal, and the matrix Θ is antisymmetric and block adiagonal.

Inserting the above expansions and definitions into equation (3.124), one obtains the optimum flux as

$$\overset{\circ}{\overline{X}} = \frac{\hat{\tau}}{2}[\Lambda - 2\hat{\tau}\Theta]S\prime\prime X = \frac{\hat{\tau}}{2}L(\hat{\tau})S\prime(X).$$

As in the pure parity case, the most likely flux is linearly proportional to the thermodynamic driving force. The term in brackets is the origin of the definition of 'the' transport matrix used above. In the present case there are both reversible and irreversible contributions to the most likely flux, whereas in the pure parity case the contributions were only irreversible.

The second entropy may be rewritten in terms of the flux. Invoking the departure of $X\prime$ from its most likely value and the reduction condition one has

$$S^{(2)}(X\prime, X \mid \tau) = S(X) + \frac{1}{2}C(\tau) : \left[X\prime - \overline{X\prime}\right]^2$$

$$= S(X) + \frac{1}{2}\left[\frac{-\Lambda^{-1}}{\mid\tau\mid} + \mathcal{O}(\tau^0)\right] : \left[X\prime - X - \frac{\mid\tau\mid}{2}L(\hat{\tau})S\prime\prime X + \mathcal{O}(\tau^2)\right]^2$$

$$= S(X) - \frac{\mid\tau\mid}{2}\Lambda^{-1} : \left[\overset{\circ}{X} - \frac{\hat{\tau}}{2}L(\hat{\tau})S\prime\prime X\right]^2 + \mathcal{O}(\tau^2)$$

$$= S(X) - \frac{\mid\tau\mid}{2}\Lambda^{-1} : \overset{\circ}{X}^2 + \frac{\tau}{2}\overset{\circ}{X} \cdot S\prime\prime X - \mid\tau\mid \overset{\circ}{X} \cdot \Lambda^{-1}\Theta S\prime\prime X$$

$$- \frac{\mid\tau\mid}{2}\Lambda^{-1} : [L(\hat{\tau})S\prime\prime X]^2 + \mathcal{O}(\tau^2).$$

As in the pure parity case, section 3.2.7, there is a term quadratic in the flux, which is negative and therefore unfavorable. There are now two terms linear in the flux, which also couple to the thermodynamic gradient, again as in the pure parity case. The second of these terms includes the antisymmetric transport matrix Θ, which does not appear in the pure parity case.

The antisymmetric part of the transport matrix Θ gives a reversible contribution to the most likely evolution of the fluctuation. It will be recalled from the pure parity case that the symmetric transport matrix Λ gave an irreversible contribution. For the present mixed parity case one has,

$$\bar{X}(\tau \mid X) - \bar{X}(-\tau \mid X) = - [Q(\tau) - Q(-\tau)]S\prime\prime X$$

$$= - 2\tau\Theta S\prime\prime X + \mathcal{O}(\tau^2).$$

This would be zero for the pure parity case. This is analogous to the result that one would obtain for ordinary mechanical evolution such as that given by Hamilton's equations, which is why it is called a reversible contribution.

It should be noted that although Θ affects the most likely evolution of the fluctuation in the present mixed parity case, it has no effect on the optimum rate of entropy production. This follows from the symmetric nature of the entropy fluctuation matrix, since one has

$$\overset{\circ}{\bar{S}}(X) = \frac{1}{\tau}\left[\bar{X}(\tau \mid X) - X\right] \cdot S''X$$

$$= X \cdot \left[\frac{\hat{\tau}}{2} S''\Lambda + S''\Theta\right] S''X$$

$$= \frac{\hat{\tau}}{2} X \cdot S''\Lambda S''X + \mathcal{O}(\tau).$$

(3.139)

The contribution from Θ vanishes because it is an antisymmetric matrix, and the double scalar product is symmetric. Hence the optimum dissipation is formally the same as for a pure parity system, equation (3.97). One might have predicted this because by the Second Law of Thermodynamics the entropy production must be irreversible, which is to say it must be proportional to the sign of the time interval $\hat{\tau}$, as this result shows.

3.4 Steady heat flow

The canonical non-equilibrium system is that of steady heat flow. This is now analyzed in detail. Analogous methods can be applied to other types of thermodynamic fluxes (driven diffusion, volume changes, electrical currents etc). The remaining generic type not treated here is a mechanical non-equilibrium system, which typically has an applied potential that varies with time. Both thermodynamic and mechanical non-equilibrium systems will be treated at the level of statistical mechanics in chapter 6.

3.4.1 First energy and first temperature

3.4.1.1 Canonical equilibrium system

Heat is a form of energy, and so the latter is the focus of this chapter. It will be recalled from chapter 2 that the derivative of the isolated system entropy with respect to energy is essentially the temperature,

$$\frac{1}{T} = \frac{\partial S}{\partial E}.$$

(3.140)

Because the number and volume of the sub-system will be held constant throughout, these will not be shown and we can simply write $S(E)$.

The temperature is an intensive variable and as such it is also given by the energy density derivative of the entropy density, equation (2.15),

$$\frac{1}{T} = \frac{\partial \sigma}{\partial \varepsilon},$$

(3.141)

where $\sigma \equiv S/V$ and $\varepsilon = E/V$.

The canonical equilibrium system consists of a sub-system able to exchange energy with a reservoir of temperature T. The constrained total entropy is

$$S_{\text{total}}(E \mid T) = S(E) - \frac{E}{T}.$$

(3.142)

This is maximized by the equilibrium energy $\bar{E}(T)$,

$$\frac{\partial S_{\text{total}}(E \mid T)}{\partial E}\Big|_{\bar{E}} = 0 \Leftrightarrow T_{\text{s}}(\bar{E}) = T,$$

(3.1
43)

at which point the temperature of the sub-system equals that of the reservoir. The constrained thermodynamic potential is minus the temperature times the constrained total entropy,

$$F(E \mid T) \equiv -TS_{\text{total}}(E \mid T) = E - TS(E).$$

(3.1
44)

The Helmholtz free energy is the minimum value of this,

$$\bar{F}(T) = F(\bar{E}(T) \mid T) = \bar{E}(T) - TS(\bar{E}(T)).$$

(3.1
45)

3.4.1.2 Canonical non-equilibrium system

The canonical non-equilibrium system consists of a sub-system sandwiched between, and in thermal contact with, two heat reservoirs of different temperatures (figure 3.3). The heat flux J, which is uniform in the optimum non-equilibrium state, is the rate of energy change of a reservoir per unit cross-sectional area.
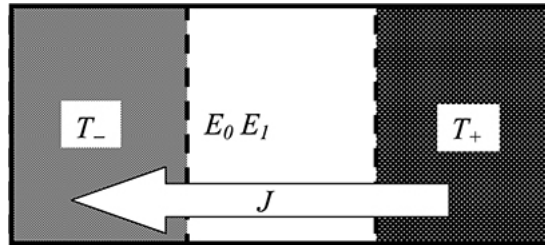
Figure 3.3. The canonical non-equilibrium system consists of two heat reservoirs of temperatures $T_+ > T_-$, in thermal contact with a sub-system with zeroth E_0 and first E_1 energy moments. The uniform heat flux is J.

The treatment of the canonical non-equilibrium system first requires a detailed analysis of the static and dynamic properties of the sub-system in isolation. This is where we begin, and reservoirs will not be introduced until section 3.4.4.

For the analysis of equilibrium thermodynamics, chapter 2, the thermodynamic limit was invoked. This says that the reservoir is infinitely larger than the sub-system, and that the boundary region is infinitely smaller than the sub-system. The first of these ensures that exchange with the sub-system does not change the field variables of the reservoir, and the second of these ensures that the precise boundary conditions have no effect on the properties of the sub-system, so that the latter may be analyzed as isolated.

For the analysis of non-equilibrium thermodynamics, the same thermodynamic limit has to be invoked for both reservoirs and both boundaries. For this reason the boundary conditions parallel to the flux need not be specified as they have negligible effect on the sub-system. An additional condition is imposed, namely that the gradient in the reservoir field variables, essentially their difference divided by the width of the sub-system, is small. This means that all quantities can be expanded in the gradient and all terms beyond the first order can be neglected.

It is also necessary to insist that the conductivity of the reservoirs be much greater than that of the sub-system. This means that the temperature within each reservoir is equal to that at its boundary with the sub-system. Or at least the width of any region of temperature non-uniformity in the reservoirs is small compared to the width of the sub-system.

The reservoirs impose a temperature gradient on the sub-system. Because of the local properties of the entropy density, there will be a corresponding gradient in energy density. One can specify the energy density $\varepsilon(\mathbf{r})$ as a macrostate of the constrained system. Choosing the z-axis to be aligned with the gradient, one need only consider spatial variations in this direction, $\varepsilon(z)$.

The energy density may be expanded to first order (since the gradient is vanishingly small, all higher order contributions can be neglected),

$$\varepsilon(z) = \varepsilon_0 + \varepsilon_1 z.$$

(3.146)

The macrostate is now specified by two parameters, $\{\varepsilon_0, \varepsilon_1\}$.

For a sub-system of width L and cross-sectional area A, the total energy is just the zeroth energy moment,

$$E_0 = A \int_{-L/2}^{L/2} \mathrm{d}z\, \varepsilon(z) = AL\varepsilon_0.$$

(3.147)

The first energy moment is

$$E_1 = A \int_{-L/2}^{L/2} \mathrm{d}z\, \varepsilon(z)z = \frac{AL^3}{12}\varepsilon_1.$$

(3.148)

With these the macrostate of the isolated sub-system can be specified in a global sense by the zeroth and first energy moments and one can write

$$S(E_0, E_1) = A \int_{-L/2}^{L/2} \mathrm{d}z\, \sigma(\varepsilon(z)).$$

(3.149)

Note that it does not matter whether or not the optimum number density is spatially varying as this is a second order effect that can be neglected.

It is important to note that both of these energy moments are extensive with the cross-sectional area of the sub-system, A. In chapter 2 where the theory for equilibrium thermodynamics was developed, it was pointed out that extensivity played a key role, so that entropy was best formulated in terms of extensive variables. An explicit benefit of this is that in the thermodynamic limit the Taylor expansion of the reservoir entropy could be terminated at the linear term. A second benefit was that the fluctuations in the extensive exchangeable variables were relatively negligible. These same benefits can be exploited in the present formulation of non-equilibrium thermodynamic systems. In this case the thermodynamic limit of relevance is $A \to \infty$. Since the width L is fixed, this also corresponds to the volume of the sub-system becoming infinite.

The local reciprocal temperature of the sub-system may also be expanded to linear order

$$\frac{\partial \sigma(\varepsilon(z))}{\partial \varepsilon(z)} = \frac{1}{T(z)}$$

$$= \frac{1}{T(0)} + z\frac{\mathrm{d}T(z)^{-1}}{\mathrm{d}z}\Big|_{z=0} + \mathcal{O}(T\prime\prime, (T\prime)^2)$$

$$\equiv \frac{1}{T_0} + z\frac{1}{T_1}.$$

(3.150)

This defines the zeroth temperature T_0 as the temperature at the mid-plane, and the first temperature T_1 as the reciprocal of the gradient of the reciprocal temperature at the mid-plane, $T_1^{-1} = -T^{-2}\nabla T$. These are properties of the sub-system, which will sometimes below be emphasized with the sub-script s.

A change in the energy density is $\Delta\varepsilon(z) = \Delta\varepsilon_0 + z\Delta\varepsilon_1$. This changes the isolated sub-system entropy in the specified macrostate by

$$\Delta S(E_0, E_1) = A\int_{-L/2}^{L/2} \mathrm{d}z\, \frac{\partial \sigma(\varepsilon(z))}{\partial \varepsilon(z)} \Delta\varepsilon(z)$$

$$= A\int_{-L/2}^{L/2} \mathrm{d}z\, \left[\frac{1}{T_0} + z\frac{1}{T_1}\right][\Delta\varepsilon_0 + z\Delta\varepsilon_1]$$

$$= AL\frac{\Delta\varepsilon_0}{T_0} + \frac{AL^3}{12}\frac{\Delta\varepsilon_1}{T_1}$$

$$= \frac{\Delta E_0}{T_0} + \frac{\Delta E_1}{T_1}.$$

(3.151)

Hence one can conclude that the zeroth temperature is the thermodynamic conjugate of the zeroth energy moment,

$$\frac{\partial S(E_0, E_1)}{\partial E_0} = \frac{1}{T_0},$$

(3.152)

and that the first temperature is the conjugate of the first energy moment,

$$\frac{\partial S(E_0, E_1)}{\partial E_1} = \frac{1}{T_1}.$$

(3.153)

This last one says in brief that the first energy moment is conjugate to the temperature gradient, a result that will play a key role in the analysis of heat flow. These give the temperatures as functions of the moments, $T_0(E_0, E_1)$ and $T_1(E_0, E_1)$. We shall assume without proof that these relationships are invertible, $E_0(T_0, T_1)$ and $E_1(T_0, T_1)$.

It is somewhat idiosyncratic to analyze heat flow in terms of the zeroth and first temperatures. The relationship with conventional results is that the zeroth temperature corresponds to the temperature, $T_0 \Rightarrow T$, and the first temperature to the temperature gradient, $T_1^{-1} \Rightarrow -T^{-2}\nabla T$. Like the present treatment, the conventional treatment assumes that the temperature gradient is small so that terms beyond first order can be neglected.

3.4.2 Second entropy

The second entropy formalism, section 3.1.4, was based on extensive non-conserved variables of the isolated system that most likely vanished, $\bar{X} = 0$. For the present case of heat flow, the corresponding variable is the first energy moment, E_1, since this is not conserved, and it is extensive with the system cross-sectional area A. Further, it is most likely zero, since the entropy is a maximum when the energy is uniformly distributed throughout the system. As we have just seen, the field variable conjugate to the first energy moment is the first temperature, T_1. From now on the zeroth energy moment, E_0, and also the zeroth temperature, T_0, will be suppressed and not treated explicitly.

For heat flow we can fix the direction (e.g. parallel to the z-axis), and take E_1 to be a scalar. This means that X is a scalar because it is the only fluctuating variable that we shall consider. Hence all of the matrices that were defined in the general first and second entropy analysis, section 3.1.4, become scalars here.

Accordingly, the part of the first entropy of the isolated system relevant to heat flow depends upon the first energy moment, with $\bar{E}_1 = 0$ being the macrostate of maximum entropy. Invoking the fluctuation approximation the first entropy is

$$S(E_1) = \frac{1}{2} S \prime\prime E_1^2.$$

(3.154)

As just mentioned, the fluctuation coefficient is a scalar,

$$S \prime\prime \equiv \left. \frac{\partial^2 S(E_1)}{\partial E_1^2} \right|_{E_1=0} = \frac{-k_{\mathrm{B}}}{\left\langle E_1(0)^2 \right\rangle}.$$

(3.155)

The averages here and below are canonical equilibrium ones. By the uniqueness of the thermodynamic state, these are the same as in an equilibrium isolated system at the energy $E_0 = \bar{E}_0(T_0)$.

The thermodynamic force is the inverse first temperature,

$$\frac{1}{T_1} \equiv \frac{\partial S(E_1)}{\partial E_1} = S \prime\prime E_1.$$

(3.156)

One can also write the thermodynamic force as $\nabla S(E_1) = S\prime(E_1) = S \prime\prime E_1$.

For the transition $E_1 \xrightarrow{\tau} E_1'$, the second entropy is

$$S^{(2)}(E_1', E_1 \mid \tau) = \frac{A(\tau)}{2} \left[E_1'^{\,2} + E_1{}^2 \right] + B(\tau) E_1' E_1.$$

(3.157)

(Obviously as a scalar X has pure parity; in this case it is even.) From the expansions of the fluctuation transition coefficients for short times, equations (3.79) and (3.80), in the present scalar, even parity case one has

$$A(\tau) = \frac{-1}{|\tau| \Lambda} + A + \mathcal{O}(\tau), \text{ and } B(\tau) = \frac{1}{|\tau| \Lambda} + B + \mathcal{O}(\tau),$$

(3.158)

with $A + B = S \prime\prime / 2$ from the reduction condition, equation (3.81). Inserting the expansion into the most likely destination gives

$$\overline{E_1'} = \frac{-B(\tau)}{A(\tau)} E_1$$

$$= E_1 + \frac{\mid \tau \mid \Lambda}{2} S'' E_1 + \mathcal{O}(\tau^2).$$

Writing the second entropy in terms of the departure from this most likely destination, in the small time limit one has

$$S^{(2)}(E_1', E_1 \mid \tau) = \frac{1}{2} S'' E_1^2 - \frac{1}{2\Lambda \mid \tau \mid} \left[E_1' - E_1 - \frac{\mid \tau \mid \Lambda}{2} S'' E_1 \right]^2 + \mathcal{O}(\tau^2)$$

$$= \frac{1}{2} S'' E_1^2 - \frac{\mid \tau \mid}{2\Lambda} \left[\mathring{E}_1 - \frac{\hat{\tau}\Lambda}{2} S'' E_1 \right]^2$$

$$= \frac{1}{2} S'' E_1^2 - \frac{\mid \tau \mid}{2\Lambda} (\mathring{E}_1)^2 + \frac{\tau}{2} \mathring{E}_1 S'' E_1 - \frac{\mid \tau \mid \Lambda}{8} (S'' E_1)^2,$$

where $\hat{\tau} = \text{sign}(\tau)$. This is the particular form of equation (3.95) for heat flow. The first term on the right-hand side is the first entropy, and so this satisfies the reduction condition for the most likely destination.

The coarse velocity, $\mathring{E}_1 \equiv [E_1' - E_1]/\tau$, which is the average rate of change over the time interval, has most likely value $\hat{\tau}\Lambda S'' E_1/2$. The rate of change of the first energy moment is directly related to the energy flux in the isolated system. Because the system is isolated, the dynamics are adiabatic, which fact is signified by appending the superscript 0.

The energy flux, J_E^0, is defined as the amount of energy crossing a plane, per unit area, per unit time. Because we are presently considering the energy flux alone, the number flux is necessarily assumed zero, and so this corresponds to the conductive energy flux.

To be definite, we assume that the flux is in the z-direction. We also assume that it is uniform, $J_E^0(\mathbf{r}) = J_E^0$, which can be expected to be the average or most likely flux. This means that the only change in energy density occurs at the z-boundaries of the isolated system; the same amount of energy per unit time flows across each and every plane perpendicular to the z-axis throughout the system. Because we take the thermodynamic limit, we need not be overly concerned with the nature of the boundary regions beyond quantifying the rate of energy change in them.

Such a uniform flux removes energy at a rate $A J_E^0$ from the boundary region at $z = -L/2$, and it adds energy at the same rate to the boundary region at $z = L/2$. The rate of change of the first energy moment due to this is

$$\mathring{E}_1 = \frac{-L}{2}(-A J_E^0) + \frac{L}{2}(A J_E^0) = V J_E^0$$

Here $V = AL$ is the volume of the isolated sub-system.

We are interested in time intervals τ that are longer than molecular times, so that following a fluctuation the rate of change of energy moment can accelerate from zero to a steady value. We also want non-uniformities in the energy profile to decay away (or to average to zero), which can be expected since in general short wavelength inhomogeneities decay more quickly than long wavelength ones, and the first moment is the non-uniformity with the longest wavelength. However, we also require the time interval

to be sufficiently short so that the change in energy moment is relatively negligible, and so that the short time expansion can be terminated at the linear term. The data in figure 3.2 show the regression of the first energy moment in a particular system, and it can be seen that there exist a range of values for the time interval τ in which both requirements are satisfied.

Under these circumstance, we expect a steady decay of the first energy moment, in which case the coarse velocity and the instantaneous velocity are equal. Equating the most likely energy flux to the most likely coarse rate of change of energy moment per unit volume, one has

$$
\begin{aligned}
\bar{J}^0_{\mathrm{E}} &= \frac{\hat{\tau}\Lambda}{2V}S\prime\prime E_1 \\
&= \frac{\hat{\tau}\Lambda}{2V}\frac{1}{T_1} \\
&= \frac{-\hat{\tau}\Lambda}{2VT^2}\nabla T.
\end{aligned}
$$

(3.162)

Hydrodynamics almost invariably refers to the future state. In this case $\hat{\tau}=1$, and this result says that the most likely energy flux is linearly proportional to the temperature gradient, and in the opposite direction (i.e. energy flows from hot to cold). This is Fourier's law, and one can identify the conventional thermal conductivity as $\lambda \equiv \Lambda/2VT^2$.

3.4.3 Green–Kubo thermal conductivity

Green–Kubo theory relates the hydrodynamic transport coefficients to an integral of the time correlation function. The general approach was derived in section 3.2.6 from the regression theorem. For the present problem of heat transport, the thermal conductivity is related to the transport coefficient by $\lambda \equiv \Lambda/2VT^2$. One can rearrange the most likely regression of the first energy moment, equation (3.159), in several ways to give Λ and hence λ. If one multiplies by $E_1(t)$ and averages then one obtains

$$
\begin{aligned}
\lambda(\tau) &= \frac{-1}{Vk_{\mathrm{B}}T^2\mid\tau\mid}\left\langle\left[\bar{E}_1(t+\tau\mid E_1(t))-E_1(t)\right]E_1(t)\right\rangle \\
&= \frac{-1}{Vk_{\mathrm{B}}T^2\mid\tau\mid}\left\langle[E_1(t+\tau)-E_1(t)]E_1(t)\right\rangle.
\end{aligned}
$$

(3.163)

Terms $\mathcal{O}(\tau)$ have been neglected here. In the second equality, the actual value replaces the most likely value, which is permissible for Gaussian statistics.

In principle, $\lambda(\tau)$ ought to be independent of the time interval. Multiplying both sides by τ, differentiating with respect to τ, and ignoring $\mathrm{d}\lambda(\tau)/\mathrm{d}\tau$, yields

$$
\lambda(\tau) = \frac{-\hat{\tau}}{Vk_{\mathrm{B}}T^2}\left\langle\dot{E}_1(t+\tau)E_1(t)\right\rangle.
$$

(3.164)

The averages that appear here are equilibrium averages. One can evaluate them for a canonical equilibrium system. The procedure is to generate points in phase space, $\mathbf{\Gamma}_0$, according to the Maxwell–Boltzmann distribution at the zeroth temperature, $\wp_{\mathrm{MB}}(\mathbf{\Gamma}_0\mid T_0)$. Then the adiabatic trajectory (i.e. Hamiltonian trajectory for an isolated system) is

generated, $\mathbf{\Gamma}^0(\tau \mid \mathbf{\Gamma}_0)$. The first moments, $E_1(0) \equiv E_1(\mathbf{\Gamma}_0)$ and $E_1(\tau) \equiv E_1(\mathbf{\Gamma}^0(\tau \mid \mathbf{\Gamma}_0))$ are then calculated and multiplied together. This is then averaged over the initial points. Equivalently, the change in first moment $E_1(\tau) - E_1(0)$ can be calculated by integrating its rate of change $\dot{E}_1(t)$ on the adiabatic trajectory over the time interval. Specifically, equation (3.163) can be written

$$\lambda(\tau) = \frac{-1}{V k_{\mathrm{B}} T^2 \mid \tau \mid} \int d\mathbf{\Gamma}_0 \, \wp_{\mathrm{MB}}(\mathbf{\Gamma}_0 \mid T_0) \int_0^{\tau} dt \, \dot{E}_1(\mathbf{\Gamma}^0(t \mid \mathbf{\Gamma}_0)) E_1(\mathbf{\Gamma}_0).$$

(3.165)

To the extent that this or equation (3.164) is independent of τ, then the short time regime can be said to exist and the theory is applicable.

The two formulae above are tested in a simulation of a Lennard-Jones fluid in figure 3.4. It is evident that $\lambda(\tau)$ converges more quickly to the steady-state value λ using the terminal velocity of the first energy moment, equation (3.164), than it does using the coarse velocity, equation (3.163). This makes sense because the latter includes the conductivity of the inertial phase, which is low or zero, and the relative contribution of this to the overall conductivity decays as τ^{-1}, which is quite slow. The instantaneous terminal velocity, after sufficient time, gives the conductivity of the system in the steady state.

Figure 3.4. Thermal conductivity from Green–Kubo expressions for a Lennard-Jones fluid at $\rho = 0.8$ and $T_0 = 2$ (dimensionless units). The simulations use the coarse velocity, Equation (3.163) (crosses), or the instantaneous velocity, equation (3.164) (circles). The solid line is a guide to the eye. Data from Attard (2005b).
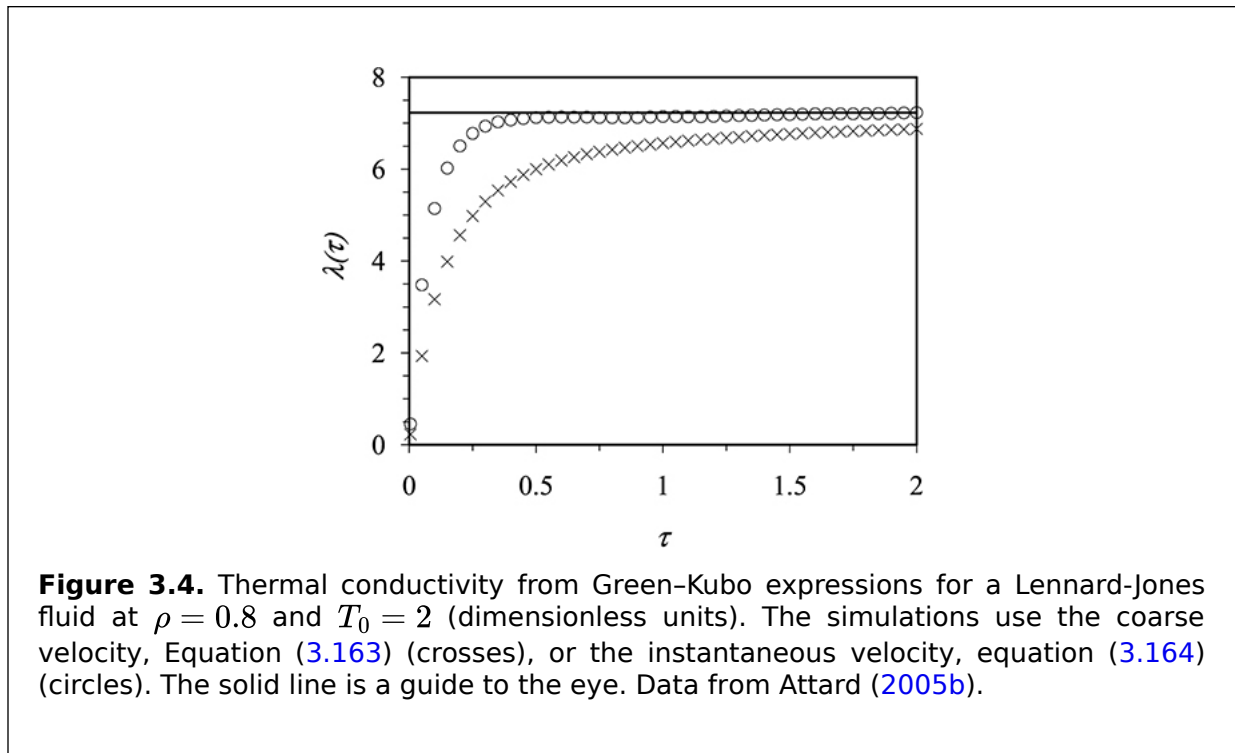
3.4.4 Reservoirs of different temperatures

The analysis of the static and dynamic structure of the isolated sub-system for heat flow is now complete. It has been shown that the first energy moment of the sub-system and its rate of change were the extensive variables essential to the treatment of heat flow. Now the analysis is extended to include energy exchange with heat reservoirs. Two reservoirs of different temperatures are used, sandwiching the sub-system as in figure 3.3. It will be shown that the reservoirs induce a nonzero first energy moment in the sub-system. It will also be shown that the energy flux through the sub-system due to the reservoirs is equal to

the rate of decay of the energy moment as if it had arisen from a spontaneous fluctuation in an isolated sub-system. Finally, it will be shown how the two apparently contradictory requirements of constant energy moment and constant rate of decay of the energy moment are enforced by the reservoirs to give a steady-state system.

3.4.4.1 First entropy

With reference to figure 3.3, the reservoir and sub-system boundaries are at $z = \pm L/2$, and the temperature of each respective reservoir is $T_{r\pm}$. Let $\Delta E_{r\pm}$ be the change in energy of the respective reservoir due to exchange with the sub-system at the boundary. It follows that the total change in the reservoir entropy is

$$
\begin{aligned}
\Delta S_r &= \frac{\Delta E_{r+}}{T_{r+}} + \frac{\Delta E_{r-}}{T_{r-}} \\
&= \frac{\Delta E_{r0}}{T_{r0}} + \frac{\Delta E_{r1}}{T_{r1}}.
\end{aligned}
$$

(3.166)

The second equality is exact and no approximation has been invoked here. The change in the zeroth and first energy moments of the reservoirs are defined as

$$
\Delta E_{r0} \equiv \Delta E_{r+} + \Delta E_{r-}, \quad \text{and} \quad \Delta E_{r1} \equiv \frac{L}{2}\Delta E_{r+} - \frac{L}{2}\Delta E_{r-}.
$$

(3.167)

The zeroth and first reservoir temperatures are defined as

$$
\frac{1}{T_{r0}} \equiv \frac{1}{2}\left[\frac{1}{T_{r+}} + \frac{1}{T_{r-}}\right], \quad \text{and} \quad \frac{1}{T_{r1}} \equiv \frac{1}{L}\left[\frac{1}{T_{r+}} - \frac{1}{T_{r-}}\right].
$$

(3.168)

These definitions are obviously analogous to those used for the isolated sub-system. It is implicitly assumed that the energy exchange takes place at the boundary and that the reservoir temperature $T_{r\pm}$ applies here because any temperature gradients within a reservoir are negligible.

Let $\Delta^r E_{s\pm}$ be the reservoir-induced change in the sub-system energy at each respective boundary. This change is due to exchange with the reservoir, and so from energy conservation one must have $\Delta^r E_{s\pm} = -\Delta E_{r\pm}$. This can be written in terms of the change in zeroth and first moment,

$$
\Delta^r E_{s0} = -\Delta E_{r0}, \quad \text{and} \quad \Delta^r E_{s1} = -\Delta E_{r1}.
$$

(3.169)

Obviously the energy exchange is assumed to occur at each boundary over a region with width small compared to the length of the sub-system.

In addition to this reservoir-induced change, the energy moments also change adiabatically. Adiabatic evolution means that of the isolated sub-system, as determined by Hamilton's equations of motion, without any direct influence of the reservoir (or artificial thermostats, or dissipative or stochastic forces). The adiabatic change can also be called the internal change because it is determined solely by the sub-system forces and it would occur if the sub-system were isolated from the reservoirs. The total change in the sub-system energy moments is

$$
\Delta E_{s0} = \Delta^0 E_{s0} + \Delta^r E_{s0}, \quad \text{and} \quad \Delta E_{s1} = \Delta^0 E_{s1} + \Delta^r E_{s1}.
$$

The superscript 0 denotes the adiabatic change, and the superscript r denotes the reservoir-induced change. The total energy of an isolated system is conserved, which means that the adiabatic change of the zeroth energy moment vanishes, $\Delta^0 E_{s0} = 0$.

In the canonical equilibrium case the reservoir energy was written $E_{\mathrm{r}} = E_{\mathrm{total}} - E_{\mathrm{s}}$ and the reservoir entropy was written $S_{\mathrm{r}} = -E_{\mathrm{s}}/T_{\mathrm{r}}$, the constant independent of the sub-system being discarded. We shall use similar notions for the present canonical non-equilibrium case, with two adjustments. First, the reservoirs exchange energy over time, and so we have to keep this in mind, at least notionally. Second, the change in reservoir energy cannot be equated to the entire change in the sub-system energy, but only the reservoir-induced part of it. In view of these considerations, and of the definitions of moments and temperatures above, the reservoir entropy can be written as

$$S_{\mathrm{r}}(t) = \frac{E_{\mathrm{r}0}(t)}{T_{\mathrm{r}0}} + \frac{E_{\mathrm{r}1}(t)}{T_{\mathrm{r}1}}.$$

The time-dependent moments defined here are meant to be only the part due to exchanges with the sub-system. Although we expect a monotonic energy flux from the hot reservoir to the cold via the sub-system, we assume that the reservoirs are so large that over the time of any measurement the reservoir temperatures are constant. These moments are defined such that their changes are the changes in reservoir moment given above,

$$\Delta E_{\mathrm{r}0}(t) = \Delta E_{\mathrm{r}0} = -\Delta^{\mathrm{r}} E_{\mathrm{s}0},$$

and

$$\Delta E_{\mathrm{r}1}(t) = \Delta E_{\mathrm{r}1} = -\Delta^{\mathrm{r}} E_{\mathrm{s}1}.$$

With these, the constrained entropy of the total system may be written as

$$S_{\mathrm{total}}(E_{\mathrm{s}0}, E_{\mathrm{s}1} \mid T_{\mathrm{r}0}, T_{\mathrm{r}1}) = S_{\mathrm{s}}(E_{\mathrm{s}0}, E_{\mathrm{s}1}) + \frac{E_{\mathrm{r}0}(t)}{T_{\mathrm{r}0}} + \frac{E_{\mathrm{r}1}(t)}{T_{\mathrm{r}1}}.$$

As above, we shall write the sub-system entropy as a quadratic form in the first energy moment. This is valid provided that the temperature gradient between the reservoirs is small.

The total entropy here has two constraints: the zeroth and first energy moments. Because the zeroth moment is a conserved variable, its treatment does not materially differ from the canonical equilibrium case. To find its optimum value one can maximize the total entropy with respect to $E_{\mathrm{s}0}$, which, is the same as differentiating with respect to $\Delta^{\mathrm{r}} E_{\mathrm{s}0}$. The first energy moment is not conserved, and it can change by either $\Delta^0 E_{\mathrm{s}1}$ or by $\Delta^{\mathrm{r}} E_{\mathrm{s}1}$. In maximizing the constrained total entropy with respect to $E_{\mathrm{s}1}$, one has to specify which of these two possible changes is meant. Because the reservoir first energy moment is $E_{\mathrm{r}1}(t)$, it is clear that the reservoir-induced change is the required one. The justification for this is that in this structural part of the problem the reservoir is the determining factor. It is only in the dynamic part of the problem that the adiabatic change becomes an essential element, as will be seen.

In view of this discussion, the derivatives of the constrained total entropy with respect to the reservoir-induced change in sub-system moments are

$$\frac{\partial S_{\text{total}}}{\partial^{\text{r}} E_{\text{s0}}} = \frac{1}{T_{\text{s0}}} - \frac{1}{T_{\text{r0}}},$$

(3.175)

and

$$\frac{\partial S_{\text{total}}}{\partial^{\text{r}} E_{\text{s1}}} = \frac{1}{T_{\text{s1}}} - \frac{1}{T_{\text{r1}}}.$$

(3.176)

The reservoir terms here follow from the results above, $\Delta E_{\text{r0}}(t) = -\Delta^{\text{r}} E_{\text{s0}}$, and $\Delta E_{\text{r1}}(t) = -\Delta^{\text{r}} E_{\text{s1}}$.

This formally gives the result that intuition expected: the constrained total entropy is a maximum with respect to reservoir-induced changes of energy when the zeroth temperature of the sub-system equals that of the reservoir, and similarly for the first temperature,

$$\bar{T}_{\text{s0}} = T_{\text{r0}}, \quad \text{and} \quad \bar{T}_{\text{s1}} = T_{\text{r1}}.$$

(3.177)

Of course if one unpacks the definitions, these simply say that the sub-system temperatures at the boundaries most likely equal those of the reservoirs. This result gives the most likely static structure for this steady-state thermodynamic non-equilibrium system.

In obtaining this static structure the adiabatic change, Δ^0 was not required or used. It will now be shown that maximizing the second entropy with respect to this adiabatic change is what gives the most likely heat flux.

3.4.4.2 Second entropy

We now turn to the energy flux, and for this we need the second entropy. Since the zeroth energy moment is a conserved variable with no adiabatic change, $\Delta^0 E_{\text{s0}} = 0$, we can suppress both the zeroth energy moment and zeroth temperature. The equality of the zeroth temperature of the sub-system and the reservoir, $\bar{T}_{\text{s0}} = T_{\text{r0}}$, determines the most likely zeroth energy moment $\bar{E}_{\text{s0}} = E_{\text{s0}}(\bar{T}_{\text{s0}})$.

In terms of notation, above Δ^{r} denoted the mathematical variation in a constraint. Such a variation was used to optimize the first entropy. In this section Δ_τ^0 and Δ_τ^{r} will be used to denote the adiabatic and reservoir-induced physical change over the time interval τ. Dividing by τ these give the adiabatic and reservoir-induced energy fluxes. In the formalism Δ_τ^0 and Δ_τ^{r} are constrained variables with respect to which the second entropy will be maximized.

Above, the second entropy for an isolated system was given in terms of fluxes in general, equation (3.95), and in particular for heat flow, equation (3.160). These were in the small time limit, valid to $\mathcal{O}(\tau)$. The four terms in the expressions have obvious physical interpretations. First is the first or ordinary entropy of the sub-system, which is quadratic in the first energy moment. Second is the term quadratic in the flux, which acts to limit the flux. Third is the rate of entropy production, which is linear in the flux. Fourth is a term quadratic in the first entropy gradient which is required by the reduction condition. By flux in equation (3.160) is meant $\overset{\circ}{E}_1$, the coarse rate of change of the first energy moment, which for an isolated system is the internal or adiabatic rate of change.

The present case differs from the earlier one by the addition of reservoirs. Hence one has to modify equation (3.160) in several respects. First, the first entropy of the sub-system has to be replaced by that of the total system. Second, the term quadratic in the internal

flux remains unchanged, because this remains the cost of dynamic order with or without reservoirs. Third, to the internal rate of entropy production is added the rate of entropy production due to the reservoirs. The latter consists of the dissipation of the reservoir entropy, plus the reservoir-induced dissipation of the sub-system entropy. Fourth, the term independent of the flux is determined by the reduction condition for the total system, namely that the maximum value of the second entropy (i.e. its value in the state of optimum flux), equals the total first entropy.

These give the second entropy for the case of heat reservoirs as

$$
\begin{aligned}
S_{\text{total}}^{(2)} & (\Delta_\tau^0 E_{\text{s}1}, \Delta_\tau^{\text{r}} E_{\text{s}1}, E_{\text{s}1} \mid T_{\text{r}1}, \tau) \\
= & \frac{1}{2} S_{\text{s}} \prime\prime E_{\text{s}1}^2 + \frac{E_{\text{r}1}(t)}{T_{\text{r}1}} - \frac{1}{2\Lambda \mid \tau \mid}[\Delta_\tau^0 E_{\text{s}1}]^2 + \frac{1}{2}\Delta_\tau^0 E_{\text{s}1} S_{\text{s}}\prime\prime E_{\text{s}1} \\
& + \frac{1}{2}\Delta_\tau^{\text{r}} E_{\text{s}1}\left[S_{\text{s}}\prime\prime E_{\text{s}1} - \frac{1}{T_{\text{r}1}}\right] + \frac{\mid \tau \mid \Lambda}{8}\left[S_{\text{s}}\prime\prime^2 E_{\text{s}1}^2 - \frac{2S_{\text{s}}\prime\prime E_{\text{s}1}}{T_{\text{r}1}}\right].
\end{aligned}
$$

(3.178)

This is valid in the small time limit, with terms $\mathcal{O}(\tau^2)$ neglected. The thermodynamic force of the sub-system is $S_{\text{s}}\prime\prime E_{\text{s}1} = T_{\text{s}1}^{-1}$. The various terms here can be matched with the required modifications of equation (3.160) that were just discussed. The fourth term here is the internal rate of entropy production, and is the same as the third term in equation (3.160). The fifth term here is the reservoir-induced production of total first entropy, which is the sum of that of the sub-system and that of the reservoir. The sixth term is determined by the reduction condition, the fact that it must comprise terms quadratic and linear in the first energy moment, and the need for its derivative to vanish at the optimum moment, as is discussed shortly.

It will be noted that this is linear in the reservoir-induced change in moment, $\Delta_\tau^{\text{r}} E_{\text{s}1}$. Hence the second entropy cannot be maximized with respect to the latter. Physically, since this is boundary driven flow, and since the conductivity of the reservoir is much greater than that of the sub-system, any change in sub-system moment from its optimum value can be immediately counteracted by flow from the reservoir. Hence one expects

$$
\overline{\Delta_\tau^{\text{r}} E_{\text{s}1}} = -\Delta_\tau^0 E_{\text{s}1}.
$$

(3.179)

We shall discuss this further shortly, and below we shall rewrite the second entropy imposing this condition explicitly.

Differentiating the second entropy with respect the internal flux gives

$$
\frac{\partial S_{\text{total}}^{(2)}}{\partial \Delta_\tau^0 E_{\text{s}1}} = \frac{-1}{\Lambda \mid \tau \mid}\Delta_\tau^0 E_{\text{s}1} + \frac{1}{2} S_{\text{s}}\prime\prime E_{\text{s}1}.
$$

(3.180)

In this $\Delta_\tau^{\text{r}} E_{\text{s}1}$ has been held constant; see below for the contrary case. Setting this to zero yields the most likely internal flux,

$$
\overline{\overset{\circ}{E_1}^0} \equiv \frac{\overline{\Delta_\tau^0 E_{\text{s}1}}}{\tau} = \frac{\hat{\tau}\Lambda}{2}\frac{1}{T_{\text{s}1}}.
$$

(3.181)

This is identical to that obtained for the isolated system, equation (3.162), as might have been expected because it arises from the adiabatic optimization of the flux. Again in

essence, it gives Fourier's law, that the sub-system flux is proportional to the sub-system temperature gradient.

Differentiating with respect to the reservoir-induced change in moment gives

$$
\frac{\partial S^{(2)}_{\text{total}}}{\partial^{\text{r}} E_{\text{s1}}} = S_{\text{s}} '' E_{\text{s1}} - \frac{1}{T_{\text{r1}}} + \frac{1}{2} S_{\text{s}} '' \left[\Delta^0_\tau E_{\text{s1}} + \Delta^{\text{r}}_\tau E_{\text{s1}} \right]
$$
$$
+ \frac{|\tau| \Lambda}{8} \left[2 S_{\text{s}} ''^2 E_{\text{s1}} - \frac{2 S_{\text{s}} ''}{T_{\text{r1}}} \right].
$$

(3.182)

Setting this to zero determines $\bar{E}_{\text{s1}}$, the optimum sub-system moment at the start of the transition. For this purpose one need only consider the reservoir-induced derivative, $\partial^{\text{r}} E_{\text{s1}}$, assuming that adiabatic changes are negligible in comparison. This is equivalent to assuming that the thermal conductivity of the reservoirs is much greater than that of the sub-system. This is a necessary assumption for the reservoir formalism of non-equilibrium thermodynamics, akin to the thermodynamic limit in defining what a reservoir is.

From the results of maximizing the first entropy, equation (3.177), the optimum first moment is given by

$$
S_{\text{s}} '' \bar{E}_{\text{s1}} \equiv \frac{1}{\bar{T}_{\text{s1}}} = \frac{1}{T_{\text{r1}}}.
$$

(3.183)

In order for the preceding derivative to vanish at this point, and hence for the second entropy to be consistent with the first entropy, one must have

$$
\Delta^{\text{r}}_\tau E_{\text{s1}} = -\Delta^0_\tau E_{\text{s1}}.
$$

(3.184)

This is the result anticipated above, namely that the reservoir-induced change in moment must cancel exactly the adiabatic change in moment.

One can insert this result into the expression for the second entropy and eliminate the reservoir-induced change in moment,

$$
S^{(2)}_{\text{total}}(\Delta^0_\tau E_{\text{s1}}, E_{\text{s1}} \mid T_{\text{r1}}, \tau)
$$
$$
= \frac{1}{2} S_{\text{s}} '' E^2_{\text{s1}} + \frac{E_{\text{r1}}(t)}{T_{\text{r1}}} - \frac{1}{2\Lambda |\tau|} \left[\Delta^0_\tau E_{\text{s1}} \right]^2 + \frac{1}{2 T_{\text{r1}}} \Delta^0_\tau E_{\text{s1}} - \frac{|\tau| \Lambda}{8 T^2_{\text{r1}}}.
$$

(3.185)

The final term has also been changed in order to satisfy the reduction condition. With this expression, the derivative with respect to the internal flux now gives

$$
\frac{\partial S^{(2)}_{\text{total}}}{\partial \Delta^0_\tau E_{\text{s1}}} = \frac{-1}{\Lambda |\tau|} \Delta^0_\tau E_{\text{s1}} + \frac{1}{2 T_{\text{r1}}},
$$

(3.186)

which vanishes when

$$
\overline{\overset{\circ}{E}_1{}^0} \equiv \frac{\overline{\Delta^0_\tau E_{\text{s1}}}}{\tau} = \frac{\hat{\tau} \Lambda}{2} \frac{1}{T_{\text{r1}}}.
$$

(3.187)

This is practically equivalent to the original result, equation (3.181), since $\bar{T}_{\text{s1}} = T_{\text{r1}}$.

The second entropy evaluated for the most likely adiabatic change is

$$S_{\text{total}}^{(2)}(\overline{\Delta_\tau^0 E_{\text{s1}}}, E_{\text{s1}} \mid T_{\text{r1}}, \tau)$$

$$= S_{\text{total}}(E_{\text{s1}}) - \frac{1}{2\Lambda \mid \tau \mid} \left[\frac{\mid \tau \mid \Lambda}{2T_{\text{r1}}} \right]^2 + \frac{1}{2T_{\text{r1}}} \frac{\mid \tau \mid \Lambda}{2T_{\text{r1}}} - \frac{\mid \tau \mid \Lambda}{8T_{\text{r1}}^2}$$

$$= S_{\text{total}}(E_{\text{s1}}).$$

(3.188)

By design, this satisfies the reduction condition exactly.

The original second entropy expression, equation (3.178), evaluated at $\overline{\Delta_\tau^0 E_{\text{s1}}} = \mid \tau \mid \Lambda S_\text{s}\prime\prime E_{\text{s1}}/2$ and $\Delta^{\text{r}} E_{\text{s1}} = -\overline{\Delta_\tau^0 E_{\text{s1}}}$, is

$$S_{\text{total}}^{(2)}(\overline{\Delta_\tau^0 E_{\text{s1}}}, \Delta_\tau^{\text{r}} E_{\text{s1}}, E_{\text{s1}} \mid T_{\text{r1}}, \tau)$$

$$= S_{\text{total}}(E_{\text{s1}}) - \frac{1}{2\Lambda \mid \tau \mid} \left[\frac{\mid \tau \mid}{2} \Lambda S_\text{s}\prime\prime E_{\text{s1}} \right]^2 + \frac{\mid \tau \mid}{4} \Lambda S_\text{s}\prime\prime E_{\text{s1}} S\prime\prime E_{\text{s1}}$$

$$- \frac{\mid \tau \mid}{4} \Lambda S_\text{s}\prime\prime E_{\text{s1}} \left[S_\text{s}\prime\prime E_{\text{s1}} - \frac{1}{T_{\text{r1}}} \right] + \frac{\mid \tau \mid \Lambda}{8} \left[S_\text{s}\prime\prime^2 E_{\text{s1}}^2 - \frac{2 S_\text{s}\prime\prime E_{\text{s1}}}{T_{\text{r1}}} \right]$$

$$= S_{\text{total}}(E_{\text{s1}}).$$

(3.189)

By design, this also satisfies the reduction condition exactly.

3.4.4.3 Cancelation of adiabatic flux

Let us further discuss the condition $\Delta_\tau^{\text{r}} E_{\text{s1}} = -\Delta_\tau^0 E_{\text{s1}}$. This says that the adiabatic change in the sub-system energy moment is exactly canceled by the reservoir-induced change. Once the sub-system has reached the optimum static structure induced by the reservoirs $T_{\text{s1}}(\bar{E}_{\text{s1}}) = T_{\text{r1}}$, then most likely there is no further change over time in that structure, $\overline{\Delta_\tau E_{\text{s1}}} = 0$.

This cancelation between the internally and externally induced changes in first energy moment has the effect of holding it constant. Hence there is no longer an inertial period over molecular time scales in which the flux comes up to speed. Nor is there decay over longer time scales as the fluctuation and hence driving force are reduced. In this steady-state system there is no distinction between the coarse velocity $\overset{\circ}{\dot{E}_1}{}^0$ and the instantaneous velocity $\dot{E}_1{}^0$.

The total change in the sub-system first energy moment over the time interval is of course the sum of the adiabatic change and the reservoir-induced change, which are individually nonzero. By energy conservation, the reservoir-induced change in the sub-system first energy moment is equal and opposite to the change in the reservoir first energy moment,

$$\overline{\Delta_\tau^0 E_{\text{s1}}} = -\overline{\Delta_\tau^{\text{r}} E_{\text{s1}}} = \overline{\Delta_\tau E_{\text{r1}}}.$$

(3.190)

Although the steady state is defined as the state in which the sub-system is macroscopically constant in time, $\overline{\Delta_\tau E_{\text{s1}}} = 0$, this result says that there remains a steady energy flow between the reservoirs. In fact the energy discrepancy manifest in their temperature difference is steadily diminishing. (However, the reservoirs are so large that

this effect is immeasurably small.) In a steady-state non-equilibrium system, the entropy of the system steadily increases while the entropy (i.e. structure and dynamics) of the sub-system is constant.

The important quantitative meaning of this last result is that given the sub-system temperature gradient equal to that between the reservoirs, then the adiabatic energy flow in the sub-system is equal to the energy flow between the reservoirs. Since the adiabatic energy flow is equal to the rate of regression of the fluctuation of the isolated system (with the same energy moment), equation (3.181), this proves another aspect of Onsager's regression hypothesis: the flux in an isolated system is the same as the flux in the same system between two reservoirs if the spontaneous structural fluctuation in the isolated system is the same as that induced by the reservoirs.

3.5 Variational hydrodynamics

This treatment of heat flow can be extended to hydrodynamic fluxes in general. The second entropy provides a variational principle for these fluxes, and it gives in the optimum state the conventional hydrodynamic equations such as the Navier–Stokes and energy equations. The derivation is a little lengthy (Attard 2012 ch 5) and in this section only the final variational principle is given. This is then applied to the case of steady heat flow in slab geometry, where the results reduce to those derived above from the behavior of the first energy moment.

The present variational formulation of hydrodynamics treats the material fluxes as constrained or fluctuating quantities that can take values independent of the thermodynamic driving forces. A similar fluctuation approach was pursued by Landau and Lifshitz (1957, 1959). Other researchers have extended and applied this approach to a range of different hydrodynamic phenomena (Fox and Uhlenbeck 1970, Keizer 1987, Schöpf and Rehberg 1994, Ortiz de Zárate and Sengers 2006). Such variational approaches to hydrodynamics possibly have computational and numerical advantages over conventional methods of evolving the system with the usual hydrodynamic equations. The present approach is directly based on the second entropy, and as such it makes direct the connection between non-equilibrium thermodynamics and hydrodynamic phenomena.

3.5.1 General result

For a system consisting of a sub-system and a reservoir, the total dissipation in the steady state is (Attard 2012)

$$
\dot{S}_{\text{total}} = \int d\mathbf{r} \left\{ \mathbf{J}_{\text{E}}^0 \cdot \nabla \frac{1}{T} - \frac{1}{T} \underline{\underline{\Pi}} : \nabla \mathbf{v} + \frac{1}{T} \sum_\alpha \dot{\xi}_\alpha A_\alpha \right.
$$
$$
\left. - \sum_k \mathbf{J}_{\text{N},k}^0 \cdot \left[\frac{\nabla \psi_k}{T} + \nabla \frac{\mu_k}{T} \right] \right\}.
$$

(3.191)

This is valid for arbitrary fluxes and fields; everything in the integrand is a function of position and time. This assumes that if energy can cross the boundary, then the temperature of the sub-system equals that of the reservoir at every point on the boundary, and similarly for any thermodynamic field variable conjugate to any flux across the boundary (Attard 2012, section 5.2). This is for the steady state; in the contrary case the boundary integrals are nonzero.

Here $T(\mathbf{r}, t)$ is the temperature, $\mathbf{v}(\mathbf{r}, t)$ is the local barycentric or center of mass velocity, μ_k is the chemical potential of species k, and $\psi_k(\mathbf{r}, t)$ is the external potential. For chemical reaction α, $A_\alpha = \sum_k \mu_k \nu_{\alpha k}$ is the chemical affinity, $\nu_{\alpha k}$ is the stoichiometric

coefficient. These are the specified field variables of the sub-system that give the thermodynamic driving forces. For the constrained fluxes, $\dot{\xi}_\alpha$ is the rate of reaction, and $\mathbf{J}_E(\mathbf{r},t)$ and $\mathbf{J}_N(\mathbf{r},t)$ are the energy and number fluxes, respectively. The superscript 0 on the latter denotes the diffusive part of the flux, which is the flux in excess of the convective flux carried by the local barycentric velocity. The diffusive part of the momentum flux, $\underline{\underline{\Pi}}(\mathbf{r},t)$, is also called the viscous pressure tensor.

In general, the second entropy consists of four terms: the first entropy of the structure, the term quadratic in the fluxes, the term bilinear in the fluxes and the thermodynamic forces that represent half the rate of total entropy production, and the term independent of the fluxes that ensures that the reduction condition is satisfied by the optimum fluxes. Since the rate of dissipation is a scalar, because of the Curie symmetry principle (de Groot and Mazur 1984, section VI.2) for an isotropic system there can be no coupling between scalars, vectors, and traceless second rank tensors, nor between vector components in different directions. Hence the three types of fluxes decouple and the second entropy is

$$S^{(2)} = S_{\text{tot}} + |\tau| \left[S_0^{(2)} + S_1^{(2)} + S_2^{(2)} \right].$$

(3.192)

The total first entropy is

$$S_{\text{tot}} = \int d\mathbf{r}\, \sigma_0(\varepsilon^{\text{int}}(\mathbf{r},t), \underline{n}(\mathbf{r},t)) + \oint d\mathbf{r}\, \frac{E_r(\mathbf{r},t)}{T_r(\mathbf{r})}.$$

(3.193)

This the sum of the sub-system entropy and the reservoir entropy, with the latter being the boundary integral form of the reservoir terms in equation (3.174), assuming that only energy is exchangeable. The first entropy density, σ_0, is a function of the local species number density, $\underline{n}(\mathbf{r},t)$, and the internal entropy density, $\varepsilon^{\text{int}}(\mathbf{r},t) = \varepsilon(\mathbf{r},t) - \underline{n}(\mathbf{r},t)\cdot\underline{\psi}(\mathbf{r},t) - \rho(\mathbf{r},t)v(\mathbf{r},t)^2/2$, where ρ is the local total mass density, and $v^2 = \mathbf{v}\cdot\mathbf{v}$ is the square of the local barycentric velocity. When the total first entropy is optimized with respect to the internal energy density, the sub-system temperature on the boundary equals the reservoir temperature.

The remaining three terms in the second entropy come from the independently coupled scalar, vector, and traceless second rank tensor fluxes. The viscous pressure tensor is decomposed into its scalar and traceless parts, $\underline{\underline{\Pi}} = \pi\underline{\underline{I}} + \underline{\underline{\Pi}}^*$, with $\pi = \text{TR}[\underline{\underline{\Pi}}]/3$. Similarly, the shear rate tensor may be decomposed as $\nabla\mathbf{v} = [\nabla\cdot\mathbf{v}/3]\underline{\underline{I}} + [\nabla\mathbf{v}]^*$. Since the viscous pressure tensor is symmetric, one has the result that $\underline{\underline{\Pi}}:\nabla\mathbf{v} = \pi\nabla\cdot\mathbf{v} + \underline{\underline{\Pi}}^*:[\nabla\mathbf{v}]^{\text{sym.}*}$.

The second entropy for the scalar fluxes is a mixed parity case,

$$S_0^{(2)} = \int d\mathbf{r}\, \frac{-1}{2}\Lambda_0^{-1} : \{\pi, \dot{\underline{\xi}}\}^2 - \frac{1}{2}\Lambda_0^{-1} : \{\bar{\pi}, \dot{\bar{\underline{\xi}}}\}^2$$

$$+ \frac{\hat{\tau}}{2T}\{\pi, \dot{\underline{\xi}}\}\cdot\Lambda_0^{-1}L_0(\hat{\tau})\{-\nabla\cdot\mathbf{v}, \underline{A}\}.$$

(3.194)

One sees here and below the usual terms: a term quadratic in the flux, a term bilinear in the flux and thermodynamic force, and a term independent of the flux designed to satisfy the reduction condition for the optimum fluxes.

The second entropy for the vector fluxes is a pure parity case,

$$S_1^{(2)} = \int d\mathbf{r} \; \frac{-1}{2} \Lambda_1^{-1} : \{\mathbf{J}_{\mathrm{E}}^0, \underline{\mathbf{J}}_{\mathrm{N}}^0\}^2 - \frac{1}{2} \Lambda_1^{-1} : \{\bar{\mathbf{J}}_{\mathrm{E}}^0, \underline{\bar{\mathbf{J}}}_{\mathrm{N}}^0\}^2$$

$$+ \frac{\hat{\tau}}{2} \{\mathbf{J}_{\mathrm{E}}^0, \underline{\mathbf{J}}_{\mathrm{N}}^0\} \cdot \left\{ \nabla \frac{1}{T}, \left[\frac{-1}{T} \nabla \underline{\psi} - \nabla \frac{\mu}{T} \right] \right\}.$$

In an isotropic system, there is no coupling between different Cartesian components, and so Λ_1 is composed of three identical blocks on the main diagonal.

The second entropy for the traceless part of the viscous pressure tensor (i.e. diffusive momentum flux tensor) is also a pure parity case,

$$S_2^{(2)} = \int d\mathbf{r} \; \frac{-1}{2} \Lambda_2^{-1} : \underline{\underline{\mathbf{\Pi}}}^* \underline{\underline{\mathbf{\Pi}}}^* - \frac{1}{2} \Lambda_2^{-1} : \overline{\underline{\underline{\mathbf{\Pi}}}}^* \overline{\underline{\underline{\mathbf{\Pi}}}}^* - \frac{\hat{\tau}}{2T} \underline{\underline{\mathbf{\Pi}}}^* : [\nabla \mathbf{v}]^{\mathrm{sym},*}.$$

In an isotropic system there is a single transport constant, $\Lambda_2^{-1} : \underline{\underline{\mathbf{\Pi}}}^* \underline{\underline{\mathbf{\Pi}}}^* = \lambda_2^{-1} \underline{\underline{\mathbf{\Pi}}}^* : \underline{\underline{\mathbf{\Pi}}}^*$.
This completes the variational formulation of hydrodynamics in the most general case.

Maximization of this most general form for the second entropy with respect to the diffusive fluxes gives the standard equations of hydrodynamics (Attard 2012, ch 5). This allows the transport constants that appear here, the various Λ, to be identified with the conventional hydrodynamic transport coefficients.

3.5.2 Conductive heat flow

For the case of conductive heat flow, the only nonzero flux is that of energy. In the slab geometry invoked earlier in this chapter, this related to the rate of change of the first energy moment of the sub-system, $\mathbf{J}_{\mathrm{E}}^0 = \dot{E}_1^0 \hat{\mathbf{z}}/V$.

The total first entropy is as given earlier, equation (3.174),

$$S_{\mathrm{total}}(E_{\mathrm{s}0}, E_{\mathrm{s}1} \mid T_{\mathrm{r}0}, T_{\mathrm{r}1}) = S_{\mathrm{s}}(E_{\mathrm{s}0}, E_{\mathrm{s}1}) + \frac{E_{\mathrm{r}0}(t)}{T_{\mathrm{r}0}} + \frac{E_{\mathrm{r}1}(t)}{T_{\mathrm{r}1}}.$$

As usual, maximizing this with respect to the first energy moment of the sub-system shows that the sub-system first temperature in the optimum state is equal to that of the reservoir, $\bar{T}_{\mathrm{s}1} = T_{\mathrm{r}1}$. In the remainder of the second entropy only the term that couples the vector fluxes is nonzero, and it is

$$S_1^{(2)} = \int d\mathbf{r} \left\{ \frac{-1}{2} \Lambda_1^{-1} : \mathbf{J}_{\mathrm{E}}^0 \mathbf{J}_{\mathrm{E}}^0 - \frac{1}{2} \Lambda_1^{-1} : \bar{\mathbf{J}}_{\mathrm{E}}^0 \bar{\mathbf{J}}_{\mathrm{E}}^0 + \frac{\hat{\tau}}{2} \mathbf{J}_{\mathrm{E}}^0 \cdot \nabla \frac{1}{T} \right\}$$

$$= A \int_{-L/2}^{L/2} dz \left\{ \frac{-1}{2\Lambda_1 V^2} (\dot{E}_1^0)^2 - \frac{1}{2\Lambda_1 V^2} (\overline{\dot{E}_1^0})^2 + \frac{\hat{\tau}}{2V} \dot{E}_1^0 \frac{1}{T_{\mathrm{s}1}} \right\}$$

$$= \frac{-1}{2\Lambda_1 V} (\dot{E}_1^0)^2 - \frac{1}{2\Lambda_1 V} (\overline{\dot{E}_1^0})^2 + \frac{\hat{\tau}}{2} \dot{E}_1^0 \frac{1}{T_{\mathrm{s}1}}.$$

Combined with the expression for the total first entropy, this is identical to the expression for the second entropy earlier derived directly for the transitions in the first energy moment, equation (3.185). That expression enforced the condition that the reservoir-induced change in first energy moment was equal and opposite to the adiabatic

change, which in physical terms ensures temperature equality between the sub-system and the reservoir on each boundary.

In the present notation, the optimum heat flux is given by $\partial S^{(2)}/\partial \dot{E}_1^0|_{\overline{\dot{E}_1^0}} = 0$, which gives

$$\overline{\dot{E}_1^0} = \frac{\hat{\tau}V\Lambda_1}{2}\frac{1}{T_{s1}}, \text{ or } \quad \bar{\mathbf{J}}_{\mathrm{E}}^0 = \frac{-\hat{\tau}\Lambda_1}{2T^2}\nabla T.$$

(3.199)

Comparing this to Fourier's law, one can identify the thermal conductivity in terms of the present transport coefficient as $\lambda = \Lambda_1/2T^2$.

3.6 Stochastic, dissipative hydrodynamic equations

An important question in hydrodynamics is how to include noise in the evolution equations. Whilst it is understood that many systems require such a stochastic contribution to initiate or to modify a transition, there is little consensus on the way to characterize such a random element, or how to include it in the evolution equations, or what its probability distribution is.

In actual fact this problem was addressed and solved by Landau and Lifshitz (1957, 1959). Although most conventional hydrodynamicists do not recognize the fluctuating hydrodynamics approach of Landau and Lifshitz, a number of more enlightened workers have variously applied and developed it (Fox and Uhlenbeck 1970, Keizer 1987, van Beijeren and Cohen 1988, Schöpf and Rehberg 1994, Ortiz de Zárate and Sengers 2006). The present second entropy approach to non-equilibrium systems provides a basis for, and a way of systematizing, fluctuating hydrodynamics.

The stochastic dissipative hydrodynamic equations in general begin with the conventional equations of hydrodynamics, as given, in any standard book on hydrodynamics. To be definite, here is followed the notation used in section 5.4 of Attard (2012). The conventional equations give the most likely evolution, to which must be added the stochastic contributions. The latter are just the Gaussian distributed diffusive fluxes, as given by the exponential of the three contributions to the second entropy, equations (3.194), (3.195), and (3.196).

For simplicity, I shall illustrate the idea in the Boussinesq approximation, which is well established (Reid and Harris 1958, Busse 1967, Bodenschatz *et al* 2000, Attard 2012). The conventional hydrodynamic equations for convective flow in Boussinesq approximation comprise the incompressibility equation,

$$0 = \nabla \cdot \mathbf{v}(\mathbf{r}, t),$$

(3.200)

the Navier–Stokes equation (infinite Prandtl number, instantaneous velocity relaxation),

$$0 = \mathcal{R}\nabla_{\parallel}^2 T(\mathbf{r}, t) + \nabla^2\nabla^2 v_z(\mathbf{r}, t).$$

(3.201)

and the energy equation is

$$\frac{\partial T(\mathbf{r}, t)}{\partial t} = v_z(\mathbf{r}, t) - \mathbf{v}(\mathbf{r}, t) \cdot \nabla T(\mathbf{r}, t) + \nabla^2 T(\mathbf{r}, t).$$

(3.202)

These are in dimensionless form with $\mathcal{R}$ being the Rayleigh number. The temperature here is the departure from that in the conducting state.

For the stochastic, dissipative version of these, one simply returns the energy equation to its exact form in terms of the diffusive energy flux explicitly, and writes the latter as the sum of its most likely and stochastic parts,

$$\mathbf{J}_{\mathrm{E}}^0(\mathbf{r},t) = \overline{\mathbf{J}_{\mathrm{E}}^0}(\mathbf{r},t) + \frac{1}{\Delta_t}\tilde{\mathbf{R}}(\mathbf{r},t)$$

$$= -\lambda \nabla T(\mathbf{r},t) + \frac{1}{\Delta_t}\tilde{\mathbf{R}}(\mathbf{r},t).$$

(3.203)

This is written in SI units, Δ_t is the time step, and the heat flux and temperature are the departure from conduction. With this the stochastic, dissipative energy equation is (dimensionless)

$$\frac{\partial T(\mathbf{r},t)}{\partial t} = v_z(\mathbf{r},t) - \mathbf{v}(\mathbf{r},t)\cdot\nabla T(\mathbf{r},t) + \nabla^2 T(\mathbf{r},t) - \frac{1}{\Delta_t^*}\tilde{\mathbf{R}}^*(\mathbf{r},t).$$

(3.204)

The probability distribution of the stochastic heat flux increment is given by the exponential of the vector contribution to the second entropy, equation (3.195),

$$S_1^{(2)}/k_{\mathrm{B}} = \frac{-1}{4T_{00}^2 \lambda k_{\mathrm{B}} \mid \Delta_t \mid} \int d\mathbf{r}\, \tilde{\mathbf{R}}(\mathbf{r},t)\cdot\tilde{\mathbf{R}}(\mathbf{r},t)$$

$$= \frac{-c_{\mathrm{p}}^*}{4T_{00}^{*2}\mid \Delta_t^* \mid} \int d\mathbf{r}^*\, \tilde{\mathbf{R}}^*(\mathbf{r},t)\cdot\tilde{\mathbf{R}}^*(\mathbf{r},t),$$

(3.205)

Here the dimensionless heat capacity is $c_{\mathrm{p}}^* \equiv c_{\mathrm{p}}L_z^3/k_{\mathrm{B}}$, the unit of time is $\tau \equiv L_z^2 c_{\mathrm{p}}/\lambda$, and the dimensionless stochastic heat increment is $\tilde{\mathbf{R}}^* = \tilde{\mathbf{R}}/(-\Delta_{\mathrm{T}})c_{\mathrm{p}}L_z$, with L_z being the height of the convective cell, T_{00} being the mid-plane temperature, and $\Delta_T < 0$ being the temperature difference. The stochastic heat flux increment is independently distributed at each time step.

It can be mentioned that this is equivalent in fluctuating hydrodynamics to determining the variance of the stochastic heat flux from the fluctuation–dissipation theorem (Landau and Lifshitz 1957, 1959, Fox and Uhlenbeck 1970, Keizer 1987, van Beijeren and Cohen 1988, Schöpf and Rehberg 1994, Ortiz de Zárate and Sengers 2006).

3.7 Non-equilibrium pattern formation

In this chapter it has been argued that the second entropy, which is the entropy of transitions, is the appropriate entropy for non-equilibrium systems. The first entropy, which is essentially the entropy of structure or of state, is not so much irrelevant or non-existent for non-equilibrium systems, but rather it is insufficient to answer the important questions that one can ask about a non-equilibrium system.

An important and interesting example that underscores this point occurs in the phenomena of transitions between non-equilibrium states, or, relatedly, the formation of patterns in non-equilibrium systems. (The present author deprecates calling these non-equilibrium phases, primarily because their thermodynamic behavior is qualitatively different to that of equilibrium phases, especially in the nature of their stability and the transitions between them.) The general principle for non-equilibrium systems is that it is the transitions not the states that are the basic elements. For the present discussion we

select a subset of all non-equilibrium systems, namely steady-state non-equilibrium systems, in which this general principle may be clearly elucidated. In steady-state systems the macroscopic state of the sub-system is constant in time, while at the same time entropy is produced continually in the reservoir or environment.

A possibly confusing point about steady-state systems is that the 'state' of the sub-system appears constant in time, and therefore it is tempting to ask: what is the optimum state? In fact, because entropy is continually produced by the reservoir, such a question is meaningless. The correct question is: what is the optimum current state given the initial state?, or, what is the optimal transition from a given initial state? Such questions can be answered even if the final state, once established, is steady and constant in time.

A common and relatively simple example is the convective rolls that can occur in a fluid transporting heat. The challenge is to articulate the principle that determines the optimum shape and arrangement of such convective rolls for a given temperature gradient. This is an open question long-standing in hydrodynamics where it is called the wave number selection problem (see e.g. Bodenschatz *et al* (2000), and references therein).

In an equilibrium system, in which both the sub-system and the reservoir are macroscopically constant in time, the optimum state (or structure, or pattern) is the one with maximum total first entropy, (equivalently, minimum free energy). Of course there can often be barriers between the different states of the subs-system, and so sometimes more than one state may be observed. Nevertheless, *the* equilibrium state is well-defined as the state of maximum total entropy, and other possible states with lower total entropy are described as meta-stable states and, by definition, they occur with a lower likelihood.

One cannot apply the same principle to a state or pattern of a non-equilibrium system, even of a steady-state non-equilibrium system, because entropy is continually produced. Hence it would make little sense to maximize the sub-system first entropy (both because it is only part of the total entropy, and because it would eventually be swamped by the amount of entropy produced by the reservoirs). Nor would it be meaningful to maximize the total entropy, because this monotonically increases with time. One cannot salvage the situation by maximizing (or extremizing) the rate of entropy production, because there is no law in thermodynamics that says that the rate of dissipation should be a maximum (see section 3.1.2).

As argued above, it is the second entropy rather than the first entropy that is useful for non-equilibrium systems. Instead of maximizing the first entropy, one should maximize the second entropy. Since the latter is the entropy of transitions, the question that should be posed is *not* what is the optimum non-equilibrium state?, *but rather* what is the optimum transition? Or, equivalently, given the current state, what is the optimum state after a given time?

Here I give two quantitative examples from convective heat flow that answer this question. The first concerns the conductive-convective transition, and the second concerns the cross-roll transition.

I compute the optimum transitions comparing the deterministic and stochastic versions of the hydrodynamic equations in the Boussinesq approximation. I also compare a semi-analytic theory, developed in section 3.7.2, that gives the mode with the fastest initial growth rate from conduction. A numerical version of this method, combined with white noise added to the initial state, is applied in section 3.7.4 to the cross-roll transition.

3.7.1 Stochastic method

Numerical results for the conduction–convection transition are given shortly. In this case the convective state is modeled as ideal straight rolls parallel to the *y*-axis, or *y*-rolls. Following Busse (1967), a discrete Fourier expansion for the fields is used, with the temperature being expanded as

$$T(\mathbf{r}, t) = \sum_{n=1}^{N} T_{n0}^{\mathrm{s}} \ \sin \ \pi_n z$$

$$+ \sum_{n=1}^{N} \sum_{j=j_{\min}}^{j_{\max}} \left\{ T_{nj}^{\mathrm{s}} \ \sin \ \pi_n z \ \cos \ j a_1 x + T_{nj}^{\mathrm{c}} \ \cos \ \pi_n' z \ \cos \ j a_1 x \right\},$$

(3.2
06)

where $\pi_n = 2n\pi$ and $\pi_n' = (2n-1)\pi$. Boussinesq symmetry is not enforced, but emerges spontaneously in the course of the transition. A similar expansion holds for the z-component of velocity (but without the $j = 0$ term). The x-component of velocity interchanges sines and cosines.

For the deterministic evolution, the small wave number $a_1 \approx 0.1$ is replaced by the wave number of the specified fundamental mode $a_x = 2.5\text{–}6$, and only about six harmonics are used.

The (dimensionless) stochastic heat increment has expansion

$$\tilde{R}_z(\mathbf{r}, t) \ = \ \sum_{n=1}^{N} R_{z,n0}^{\mathrm{c}} \ \cos \ \pi_n z$$

$$+ \sum_{n=1}^{N} \sum_{j=j_{\min}}^{j_{\max}} \left\{ R_{z,nj}^{\mathrm{c}} \ \cos \ \pi_n z \ \cos \ j a_1 x + R_{z,nj}^{\mathrm{s}} \ \sin \ \pi_n' z \ \cos \ j a_1 x \right\},$$

(3.2
07)

and

$$\tilde{R}_x(\mathbf{r}, t) = \sum_{n=1}^{N} \sum_{j=j_{\min}}^{j_{\max}} \left\{ R_{x,nj}^{\mathrm{s}} \ \sin \ \pi_n z \ \sin \ j a_1 x + R_{x,nj}^{\mathrm{c}} \ \cos \ \pi_n' z \ \sin \ j a_1 x \right\}.$$

(3.2
08)

The (dimensionless) second entropy is

$$S_1^{(2)}/k_{\mathrm{B}} = \frac{-c_{\mathrm{p}}^*}{4 T_{00}^{*2} \mid \Delta_t^* \mid} \frac{L_x^* L_y^*}{4} \left[\sum_{n=1}^{N} 2 (R_{x,n0}^{\mathrm{c}})^2 \right.$$

$$\left. + \sum_{n=1}^{N} \sum_{j=j_{\min}}^{j_{\max}} \left\{ (R_{x,nj}^{\mathrm{s}})^2 + (R_{z,nj}^{\mathrm{c}})^2 + (R_{x,nj}^{\mathrm{c}})^2 + (R_{z,nj}^{\mathrm{s}})^2 \right\} \right].$$

(3.2
09)

Based on the experimental configuration of Busse and Whitehead (1971), $L_x = L_y = 30\text{–}80$ cm, $L_z = 0.5\text{–}1$ cm, and $a_1 \approx 2\pi/L_x \approx 0.1\text{–}0.2$ cm^{-1}. It was found that $N = 6$, $j_{\min} = 1\text{–}25$, $j_{\max} = 175$ and $\Delta_t^* = 10^{-4}\text{-}10^{-3}$ yielded reliable results. With these parameters, the standard deviation in the (dimensionless) stochastic heat increment coefficients is typically $\mathcal{O}(10^{-12})$.

Note that in addition to the stochastic contribution to the rate of change of temperature (and the other fields), there is also a stochastic contribution to the initial state at the commencement of the transition. This is governed by the exponential of the static part of the total first entropy. The fluctuation in the internal energy contribution of the static part of the total entropy is (Attard 2012 section 6.2)

$$\Delta S_{\text{tot,st}}^{(1),\text{int}} = \int_V \text{d}\mathbf{r}\, \sigma_{\text{tot,st}}^{\text{int}}(\mathbf{r}; \Delta T)$$

$$= \int_V \text{d}\mathbf{r}\, \frac{-c_\text{p}}{2T_{00}^2} \Delta T(\mathbf{r})^2.$$

(3.2 10)

This gives independent Gaussian distributions for the Fourier coefficients for the fluctuation about the initial state.

The gravitational contribution is

$$\sigma_{\text{tot,st}}^{\text{g}}(\mathbf{r}; \Delta T) = \frac{\alpha m n_{00} g z}{T_{00}} \left\{ 1 - \frac{z \Delta_T}{L_z T_{00}} \right\} \Delta T(\mathbf{r}),$$

(3.2 11)

and the kinetic energy contribution is

$$\sigma_{\text{tot,st}}^{\text{KE}}(\mathbf{r}; \Delta\mathbf{v}) = \frac{-m n_{00}}{2T_{00}} \left\{ 1 - \frac{z \Delta_T}{L_z T_{00}} \right\} \Delta\mathbf{v}(\mathbf{r}) \cdot \Delta\mathbf{v}(\mathbf{r}).$$

(3.2 12)

Consistent with the Boussinesq approximation, these have here been neglected.

The stochastic algorithm used in Attard (2012, section 6.6) solved the deterministic hydrodynamic equations with white noise added to the initial state. The power added by the noise was $\mathcal{O}(10^5)$ times the power that would have been added by the present Gaussian distribution. Nevertheless, as will be seen, it does not appear to have affected the result for the optimum wave number following the conduction–convection or the cross-roll transition.

3.7.2 Linear perturbation growth rate

A plausible alternative approach to find the optimum transition from conduction to convection is based on the mode with the fastest initial growth rate. A small amplitude, periodic, convective perturbation is imposed on the conductive state and its rate of growth or decay is determined. (Linear perturbation theory more generally in convection is reviewed by Bodenschatz *et al* (2000).) It is plausible that the wave number of the perturbation with the largest positive growth rate is the wave number of the most likely final state for the conduction–convection transition.

The temperature and velocity perturbations are written with a time-dependent amplitude

$$T(\mathbf{r}, t) = A(t)T(\mathbf{r}), \quad \text{and} \quad v_z(\mathbf{r}, t) = A(t)v_z(\mathbf{r}),$$

(3.2 13)

so that the energy equation becomes

$$\frac{\dot{A}(t)}{A(t)} = \frac{v_z(\mathbf{r})}{T(\mathbf{r})} + \frac{1}{T(\mathbf{r})} \nabla^2 T(\mathbf{r}) = \xi.$$

(3.2 14)

Here the non-linear term has been neglected because the perturbation can be take to be infinitesimal in its initial growth phase. The two sides of the first equality must individually be constant because they depend upon different independent variables. It follows that $A(t) = A_0 e^{\xi t}$, and ξ is the growth rate of the perturbation.

The perturbation is taken to be periodic with wave number *a*,

$$\nabla_\parallel^2 T(\mathbf{r}) = -a^2 T(\mathbf{r}), \quad \text{and} \quad \nabla_\parallel^2 v_z(\mathbf{r}) = -a^2 v_z(\mathbf{r}).$$

With this the Navier–Stokes equation becomes

$$0 = -a^2 \mathcal{R} T(\mathbf{r}) + [\partial_z^2 - a^2]^2 v_z(\mathbf{r}).$$

Putting this into the energy equation and rearranging gives

$$0 = a^2 \mathcal{R} v_z(\mathbf{r}) - \xi[\partial_z^2 - a^2]^2 v_z(\mathbf{r}) + [\partial_z^2 - a^2]^3 v_z(\mathbf{r}).$$

This is the same as the result given by Reid and Harris (1958) except for the inclusion of the growth rate term, $\xi \neq 0$. The solution is evidently

$$v_z(\mathbf{r}) = \{A_1 \, \cosh \, p_1 z + A_2 \, \cosh \, p_2 z + A_3 \, \cosh \, p_3 z\} f(x, y),$$

where f is periodic with wave number a, $\nabla_\parallel^2 f = -a^2 f$. The p_j are the roots of the cubic equation

$$0 = a^2 \mathcal{R} - \xi[p^2 - a^2]^2 + [p^2 - a^2]^3.$$

The three boundary conditions at $z = \pm 1/2$ are

$$v_z = \partial_z v_z = [\partial_z^2 - a^2]^2 v_z = 0.$$

The first says that the normal component of the velocity must vanish at the upper and lower rigid surfaces. The second comes from the fact that the horizontal velocity vanishes at the surfaces (stick boundary conditions) and the compressibility equation. The third comes from the vanishing of the temperature perturbation at the surfaces and the Navier–Stokes equation. Since the boundary conditions are homogeneous, a non-trivial solution exists only if the determinant of the linear system of equations vanishes. The characteristic equation for this is

$$\begin{vmatrix} C_1 & C_2 & C_3 \\ p_1 S_1 & p_2 S_2 & p_3 S_3 \\ [p_1^2 - a^2]^2 C_1 & [p_2^2 - a^2]^2 C_2 & [p_3^2 - a^2]^2 C_3 \end{vmatrix} = 0.$$

Here $C_j \equiv \cosh \, p_j/2$ and $S_j \equiv \sinh \, p_j/2$, $j = 1, 2, 3$. Given that the roots $p_j(\mathcal{R}, a, \xi)$ have been found, this is a transcendental equation that determines the rate of growth of the mode, $\xi(\mathcal{R}, a)$. It proved relatively straightforward to solve this equation numerically. It is a reasonable proposition —absent contrary evidence in specific cases— that the preferred final wave number $\bar{a}(\mathcal{R})$ for any transition is the one that has the largest positive initial growth rate.

3.7.3 Wave number selection: conduction to convection

Figure 3.5 shows the most likely wave number of straight roll convection following a transition from conduction. The circles are the stochastic, dissipative hydrodynamics equations, section 3.6, the triangles are deterministic hydrodynamic equations with added

initial white noise (Attard 2012, section 6.6), and the solid curve is the wave number of the initially fastest growing mode given by the linear perturbation theory, section 3.7.2. It can be seen that the mode with the initial fastest growth rate obtained from the perturbation analysis agrees with both the stochastic results. As mentioned above, the fluctuations in the initial temperature are given by equation (3.210), which is Gaussian white noise. There is no fundamental justification for instead adding uniform white noise to the initial state and then using conventional deterministic hydrodynamics equations for the transition, which is the method used in section 6.6 of Attard (2012). Moreover, the total power that was added as white noise turns out to be several orders of magnitude larger than what would have been given by equation (3.210). Nevertheless, the white noise does not appear to have done much damage, and both stochastic methods in figure 3.5 show that the initially fastest growing mode from conduction most likely establishes itself as the final steady state. The non-linear coupling in the hydrodynamic equations that comes into play as the amplitudes of the modes increases during the transition is, with rare exceptions, insufficient to change the dominant mode once it has begun to be established.

Figure 3.5. Wave number of the spontaneous conduction–convection transition as a function of Rayleigh number. The circles result from the stochastic, dissipative hydrodynamics equations, section 3.6, the triangles result from the conventional deterministic hydrodynamics equations with uniform white noise initially added to the temperature (Attard 2012 section 6.6), and the solid curve is the linear perturbation theory based on the largest initial growth rate, section 3.7.2. Also shown is the wave number of maximum heat flux (maximum Nusselt number, dashed curve), and the wave number of maximum sub-system first entropy (dotted curve) (Attard 2012, section 6.6).

   The results in figure 3.5 show a wave number that increases with increasing Rayleigh number, at first rapidly, and then, beyond about $\mathcal{R} \approx 10^4$, slowly. In contrast, the experimental measurements of Willis *et al* (1972) for a large Prandtl number silicone oil show a wave number for convection that strongly decreases with increasing Rayleigh number, in apparent contradiction of the present results. One possible reason for the discrepancy is that the present model of perfect straight rolls is too great an idealization of the experiments, and that the defects, curved rolls, and finite domains included in the experimental averages need to be accounted for. Another possibility, more germane to the present point, is that the experimental protocol of Willis *et al* corresponds to a convection–

convection transition that is induced by a change in Rayleigh number, $a_x \xrightarrow{\Delta \mathcal{R}} a'_x$, whereas the present results are for the conduction–convection transition at fixed Rayleigh number. Arguably, the contradiction between the two results actually confirms the point being made: there is no such thing as an optimum non-equilibrium state, but there is an optimum transition from a given non-equilibrium state.

Also shown in the figure are the wave number of maximum sub-system first entropy and of maximum first entropy production (equivalently maximum heat flow, also known as the Nusselt number). These are one-time properties that are a function only of the convecting state, not of the transition. It can be seen that neither the first entropy nor the rate of first entropy production determine the wave number of the convecting state selected by the transition.

It should be mentioned that at each Rayleigh number the range of wave numbers that yield stable steady-state solutions to the hydrodynamic equations can be quite broad. For example, at $\mathcal{R} = 10\,000$, the present perturbation theory yields a positive growth rate in the range $0.75 \leqslant \alpha \leqslant 8.97$, and at $\mathcal{R} = 6000$ the range is $0.99 \leqslant \alpha \leqslant 7.54$, which agrees with the region of stability estimated by Reid and Harris (1958).

In the present case of the conduction–convection transition, one can conclude that the initial growth rate is responsible for wave number selection of the non-equilibrium pattern. This is a combined property of the initial state (the calculations are a perturbation carried out in the conducting state) and of the final state (the growth rate is of a trial wave number for the final steady state). It is the entropy of the transition, the second entropy, not the entropy of the final state, that is maximized by the optimum non-equilibrium pattern.

3.7.4 Wave number selection: cross-roll transition

The principle being argued here—that the objects of non-equilibrium systems are the transitions not the states—is exemplified by the calculated and measured data shown in figure 3.6. In this case the transition is between two linear straight roll convective states that are mutually orthogonal.
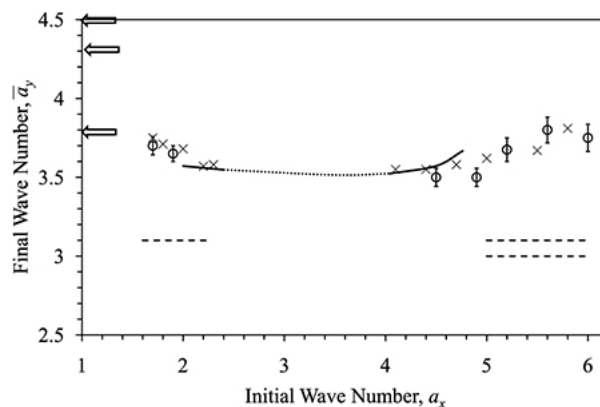
Figure 3.6. Cross-roll transition at a Rayleigh number $\mathcal{R} = 10^4$. The circles are the most likely final wave number for the full transition, and the crosses are the wave number a_y with initial fastest growth rate, both from the hydrodynamics equations in Boussinesq approximation (see text) and initially in a stable straight roll state with wave number a_x. The curve is a linear perturbation theory (Busse 1967) giving the wave number with largest initial growth rate (solid portion positive, dotted portion negative) (Busse and Whitehead 1971). The horizontal dashed lines approximate the range of measured laboratory data (Busse and Whitehead 1971). The horizontal

arrows on the vertical axis denote for straight rolls at $\mathcal{R} = 10^4$ the calculated wave number of maximum heat flow ($a_y = 3.77$), maximum sub-system entropy ($a_y = 4.53$), and the most likely wave number following a spontaneous transition from the conducting state ($\bar{a}_y = 4.25$).

In the experiments the system was initially given a wave number by a shadow mask and intense light source, and the transition to an orthogonal state upon removal of the light source was observed (Busse and Whitehead 1971). Although the final state was generally dominated by straight rolls, systems with defects, curved rolls, and finite domains were also used in the reported results.

The circles represent the full cross-roll transition, which were obtained using typically N = 6 vertical z-modes, Q = 6 horizontal x-modes (harmonics of a_x), and $j_{\mathrm{max}} = 150$ horizontal y-modes (harmonics of $a_1 = 0.2$, with $j_{\mathrm{min}} = 1$). Both the conventional deterministic hydrodynamic equations and the stochastic, dissipative hydrodynamic equations of section 3.6 were used. The initial conditions were either no added noise, Gaussian white noise (section 3.6), or uniform white noise that was several orders of magnitude larger (Attard 2012, section 6.6). No systematic differences between the various cases were found.

The initial growth rate was also estimated by again solving both the deterministic and the stochastic hydrodynamic equations in Boussinesq approximation, with or without added white noise. Again no systematic differences between the various cases were found. A particular fundamental mode a_y and five harmonics were specified, and the exponential growth rate during the initial, linear part of the transition was obtained. This was repeated with different specified fundamental modes to find the one with the largest growth rate.

The calculations for the initial growth rate are in principle identical to the linear perturbation theory of Busse (1967), notwithstanding slightly different computational parameters. The latter essentially solves the deterministic hydrodynamic equations in Boussinesq approximation using a Fourier expansion with $N + Q = 6$–12 modes, and finds the wave number of the perturbation to the initial straight roll steady state with the largest initial growth rate. If this is negative the initial state is stable and no transition is expected to occur; if it is positive the initial state is unstable and a cross-roll transition is expected to occur. In the latter case one expects that at least the initial growth will be dominated by the transverse straight roll state with that wave number. Numerical results from this perturbation approach are given by Busse and Whitehead (1971) and are shown as the curves in figure 3.6.

It can be seen in figure 3.6 that the two calculations of the mode with the initial fastest growth rate are in good agreement with each other. They are also in relatively good agreement with the most likely final wave number for the full transition. In some ways this agreement is surprising because in the full transitions it was often observed that, due to the stochastic nature of the process, the dominant wave number at small times was not always the calculated most likely fastest growing mode. It was also observed that the wave number that dominated most of the transition was sometimes overtaken by another mode toward the end of the transition, due no doubt to non-linear coupling that becomes important in the late stages. The fact that there is statistical noise in the final wave number also indicates that the fastest initial growth rate is not the *sole* determinant of the final state. Despite these phenomena, it can be seen in figure 3.6 that the linear perturbation theory for the fastest initial growth rate gives a quite good estimate of the final wave number of the cross-roll transition. Computationally, the linear perturbation theory obtains the initially fastest growing mode many, many orders of magnitude more efficiently than the full calculation of the final transition. Figure 3.6 shows that the calculated results overestimate the experimental measured final state (Busse and Whitehead 1971) by about

half a wave number. In the experiments, the cross-roll transition often begins independently over small domains, which grow and join up to occupy the entire convection box, often with defects (Busse and Whitehead 1971). In the calculations, combinations of ideal straight rolls are used that occupy the entire convection box. Also, in the computations, the initial *y*-rolls are constrained to the fundamental wave number a_x and up to five harmonics. Any transition pathway that proceeds by occupying a spectrum of nearby wave numbers is precluded in the computations. Such pathways may or may not contribute to the experimentally measured cross-roll transition. Finally, it is likely that mechanical noise from water flow in the heat bath in the experiments is much larger than the statistical noise used here. These differences may possibly account for the discrepancy between the calculated final wave numbers in figure 3.6 and the measured ones.

In figure 3.6, as in figure 3.5, it can be seen that the optimum final state is not equal to the state that maximizes the rate of (first) entropy production (in this case maximum heat flow), nor to the state that maximizes the sub-system first entropy. It is also not equal to the optimum state that follows a transition from conduction at this Rayleigh number.

It can be mentioned that the total entropy monotonically increased during these spontaneous transitions (Attard 2012, section 6.6). The sub-system entropy and the sub-system dependent part of the total entropy (i.e. the static entropy) were higher in the final state than in the initial state, as was the rate of entropy production (the heat flux, or the Nusselt number). The increase in these was not monotonic during a spontaneous transition.

The various approaches in figure 3.6 agree that the final steady state depends on the initial steady state. All three theories show that the final wave number varies with the initial wave number, $\bar{a}_y(a_x)$. Also, both the experiments and the theories identify a range of initial wave numbers for which no transition occurs.

The fact that the final wave number varies with the initial wave number precludes the existence of *any* one-time property that is optimized by the final state. The data in figure 3.6 are a graphic illustration of the general principle that for a non-equilibrium system there is no such thing as an optimum state. Rather, there is an optimum transition, or, equivalently, an optimum final state conditional upon a given initial state.

Summary

- The Second Law of Thermodynamics characterizes equilibrium systems. For these it gives a variational principle, namely that for a not-in-equilibrium system the equilibrium value of the constrained parameter corresponds to a maximum in the constrained entropy. Non-equilibrium systems are *not* characterized by any analogy of the Second Law based on extremizing the rate of change of entropy.
- The second entropy characterizes transitions between two states over a time interval, and hence also rates of change and fluxes. For a constrained non-equilibrium system, the second entropy is a maximum for the optimum rate or flux. The reduction condition says that the maximum value of the second entropy equals the first entropy for the initial state.
- Fluctuations of non-conserved extensive variables are Gaussian distributed according to a quadratic form for the second entropy. The most likely regression of such a fluctuation maximizes the second entropy. A small time expansion of the second entropy shows that a fluctuation most likely regresses at a constant rate in proportion to the gradient of the first entropy. The proportionality constant comes from the leading coefficient of the small time expansion of the second entropy and may be identified with the hydrodynamic transport matrix.
- The hydrodynamic transport matrix is symmetric, which is known as the Onsager reciprocal relations, and it can be related to the time correlation function, which is known as the Green–Kubo relations.

- For the case of steady-state thermodynamic non-equilibrium system, the appropriate extensive non-conserved variable to be constrained is typically the first moment of an extensive conserved variable. Its spontaneous fluctuation when isolated can also be induced by two spatially separated reservoirs with a difference between the respective conjugate field variables. The adiabatic (isolated) regression of the fluctuation is compensated by the rate of exchange with the reservoirs, which means that the structure of the induced moment is constant, as is the driven flux.
- Hydrodynamics can be formulated as a variational principle in which the second entropy is maximized by the optimum fluxes for the specified thermodynamic forces. Stochastic contributions to the conventional hydrodynamic equations are distributed according to the exponential of the second entropy.
- The conditionally optimum state or pattern of a non-equilibrium system is characterized by maximizing the second entropy for the transition, and it depends upon the given initial state.

References

Attard P 2005a Statistical mechanical theory for steady state systems. II. Reciprocal relations and the second entropy *J. Chem. Phys.* **122** 154101

Attard P 2005b Statistical mechanical theory for steady state systems. III. Heat flow in a Lennard-Jones fluid *J. Chem. Phys.* **122** 244105

Attard P 2006 Statistical mechanical theory for steady state systems. VI. Variational principles *J. Chem. Phys.* **125** 214502

Attard P 2012 *Non-Equilibrium Thermodynamics and Statistical Mechanics: Foundations and Applications* (Oxford: Oxford University Press)

van Beijeren H and Cohen E G D 1988 The effects of thermal noise in a Rayleigh Bénard cell near its first convective instability *J. Stat. Phys.* **53** 77

Biot M A 1955 Variational principles in irreversible thermodynamics with application to viscoelasticity *Phys. Rev.* **97** 1463

Biot M A 1975 A virtual dissipation principle and Lagrangian equations in nonlinear irreversible dynamics *Bull. Acad. Roy. de Belgique (Classe des Sciences)* **61** 6

Bochkov G N and Kuzovlev Y E 1980 On the variational principle of non-equilibrium non-linear statistical thermodynamics *JETP Lett.* **30** 46

Bodenschatz E, Pesch W and Ahlers G 2000 Recent developments in Rayleigh Bénard convection *Annu. Rev. Fluid Mech.* **32** 709

Busse F H 1967 On the stability of two-dimensional convection in a layer heated from below *J. Math. and Phys.* **46** 140

Busse F H and Whitehead J A 1971 Instabilities of convection rolls in a high Prandtl number fluid *J. Fluid Mech.* **47** 305

Dewar R 2003 Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states *J. Phys. Math. Gen.* **36** 631

Doob J L 1942 The Brownian movement and stochastic equations *Ann. Math.* **43** 351

Eddington A S 1928 *The Nature of the Physical World* (New York: Macmillan)

Fisher I 1906 *The Nature of Capital and Income* (New York: Macmillan)

Fox R F and Uhlenbeck G E 1970 Contributions to non-equilibrium thermodynamics. I. Theory of hydrodynamical fluctuations *Phys. Fluids* **13** 1893 Contributions to nonequilibrium thermodynamics. II. Fluctuation theory for the Boltzmann equation *Phys. Fluids* **13** 2881

Gage D H, Schiffer M, Kline S J and Reynolds W C 1966 *Non-Equilibrium Thermodynamics: Variational Techniques and Stability* ed Donnelly R J (ed) (Chicago, IL: University of Chicago Press)

Green M S 1954 Markoff random processes and the statistical mechanics of time-dependent phenomena. II. Irreversible processes in fluids *J. Chem. Phys.* **22** 398

de Groot S R and Mazur P 1984 *Non-Equilibrium Thermodynamics* (New York: Dover)

Gyarmati I 1968 On the most general form of the thermodynamic integral principle *Z. Phys. Chem.* **239** 133

Gyarmati I 1970 *Nonequilibrium Thermodynamics, Field Theory, and Variational Principles* (Berlin: Springer)

Hashitsume N 1952 A Statistical theory of linear dissipative systems, I *Prog. Theor. Phys.* **8** 461

Hashitsume N 1956 A Statistical Theory of Linear Dissipative Systems, II *Prog. Theor. Phys.* **15** 369

Hunt K L C, Hunt P M and Ross J 1987 Dissipation in steady states of chemical systems and deviations from minimum entropy production *Physica* A **147** 48

Keizer J and Fox R F 1974 Qualms regarding the range of validity of the Glansdorff-Prigogine criterion for stability of non-equilibrium states *Proc. Nat. Acad. Sci. USA* **71** 192

Keizer J 1987 *Statistical Thermodynamics of Non-Equilibrium Processes* (New York: Springer)

Kleidon A (ed) and Lorenz R D (ed) 2005 *Non-Equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond* (Berlin: Springer)

Kortweg D J 1883 On a general theorem of the stability of the motion of a viscous fluid *Phil. Mag.* **16** 112

Kubo R 1966 The fluctuation-dissipation theorem *Rep. Prog. Phys.* **29** 255

Kubo R, Toda M and Hashitsume N 1978 Statistical physics II *Non-Equilibrium Statistical Mechanics* (Berlin: Springer)

Landau L D and Lifshitz E M 1957 Hydrodynamic fluctuations *Sov. Phys. JETP* **5** 512

Landau L D and Lifshitz E M 1959 *Fluid Mechanics* (Oxford: Pergammon Press)

Lavenda B H 1985 *Nonequilibrium Statistical Thermodynamics* (Chichester: Wiley)

Martyushev L M and Seleznev V D 2006 Maximum entropy production principle in physics, chemistry and biology *Phys. Rep.* **426** 1

Onsager L 1931 Reciprocal relations in irreversible processes. I *Phys. Rev.* **37** 405 and Reciprocal relations in irreversible processes. II **38** 2265

Onsager L and Machlup S 1953 Fluctuations and irreversible processes *Phys. Rev.* **91** 1505

Ortiz de Zárate J M and Sengers J V 2006 *Hydrodynamic Fluctuations in Fluids and Fluid Mixtures* (Amsterdam: Elsevier)

Paltridge G W 1979 Climate and thermodynamic systems of maximum dissipation *Nature* **279** 630

Prigogine I 1967 *Introduction to Thermodynamics of Irreversible Processes* (New York: Interscience)

Reid W H and Harris D L 1958 Some further results on the Bénard problem *Phys. Fluids* **1** 102

Ross J and Vlad M O 2005 Exact solutions for the entropy production rate of several irreversible processes *J. Phys. Chem.* **109** 10607

Schneider E D and Kay J J 1994 Life as a manifestation of the second law of thermodynamics *Math. Comp. Mod.* **19** 25

Schöpf W and Rehberg I 1994 The influence of thermal noise on the onset of travelling-wave convection in binary fluid mixtures: an experimental investigation *J. Fluid. Mech.* **271** 235

Strutt J W 1871 (Rayleigh Lord) Some general theorems relating to vibrations *Proc. Math. Soc. London* **4** 357

Strutt J W 1913 (Rayleigh Lord) On the motion of a viscous fluid *Phil. Mag.* **26** 776

Swenson R and Turvey M T 1991 Thermodynamic reasons for perception-action cycles *Ecol. Psychol.* **3** 317

von Helmholtz H L F 1971 *Selected writings of Hermann von Helmholtz* (Middletown, Conn: Wesleyan University Press)

Willis G E, Deardorff J W and Somerville R C 1972 Roll-diameter dependence in Rayleigh convection and its effect upon the heat flux *J. Fluid Mech.* **54** 351

Zwanzig R 2001 *Non-Equilibrium Statistical Mechanics* (Oxford: Oxford University Press)

Entropy Beyond the Second Law (Second Edition)

Thermodynamics and statistical mechanics for equilibrium, non-equilibrium, classical, and quantum systems

Phil Attard

Chapter 4

Entropy for Brownian motion, fluctuations, and stochastic trajectories

'All chance [is] direction which thou canst not see'

<div align="right">(Pope 1734)</div>

'We ought to regard the present state of the Universe as the effect of its antecedent state and as the cause of the state that is to follow'

<div align="right">Laplace (Weinert 2005 p 197)</div>

The main aim of this chapter is to derive the equations of motion for macroscopic variables. Such variables range from the traditional Brownian particles to the more abstract thermodynamic fluctuations. Some of the results also apply to molecular coordinates, and to the evolution of the probability density. It will be shown in general that because the reservoir coordinates are projected out of the problem, the equations of motion are both stochastic and dissipative.

This dual consequence of the projection of the reservoir onto the sub-system is seldom taken into account in contemporary treatments of statistical systems. In particular, deterministic thermostatted equations of motion are widely used in computer simulations, as exemplified by the isokinetic thermostat,

$$\dot{\mathbf{p}} = -\nabla_{\mathbf{q}}\mathcal{H}(\mathbf{p},\mathbf{q}) + \alpha\mathbf{p}, \quad \alpha(\mathbf{\Gamma}) \equiv \frac{1}{\mathbf{p}\cdot\mathbf{p}}\,\mathbf{p}\cdot\nabla_{\mathbf{q}}\mathcal{H}(\mathbf{p},\mathbf{q}).$$

<div align="right">(4.1)</div>

Here $\mathbf{q}$ are the positions, $\mathbf{p}$ are the momenta, and $\mathcal{H}$ is the Hamiltonian. Because this neglects the stochastic part of the projection operation, it violates the fluctuation–dissipation theorem, one of the most fundamental theorems of statistical mechanics.

A related misunderstanding invokes Liouville's theorem for the evolution of the probability density together with such deterministic, non-Hamiltonian equations of motion,

$$\frac{\mathrm{d}\wp(X,t)}{\mathrm{d}t} = \frac{\partial\wp(X,t)}{\partial t} + \dot{X}^{\mathrm{det}}\cdot\nabla\wp(X,t).$$

<div align="right">(4.2)</div>

Here X is the macrostate, $\dot{X}^{\mathrm{det}}$ is its deterministic velocity, and $\wp(X,t)$ is its probability. It will be shown that in general this is an approximation rather than an exact theorem, and that its accuracy and reliability depends on the circumstances of each particular application.

4.1 Brownian motion

Brownian motion refers to the random movements of a microscopic solute or colloid particle in a solvent (Brown 1828, Haw 2002). This is a prototype of all stochastic processes. In science Einstein (1905) and Smoluchowski (1906) gave the earliest statistical treatments of Brownian motion, and these provided more or less direct evidence for the

existence of molecules and the validity of statistical mechanics. (Bachelier (1900, 1914) earlier derived these and other results in the context of financial markets and gambling.) The diffusion of the probability density that was the basis of their treatment was further developed by Fokker (1914) and Planck (1916). Langevin (1908) gave a stochastic dissipative equation that described Brownian motion itself, and this has become the general model for almost all the different types of random processes.

Here the stochastic dissipative equation for Brownian motion is derived from the second entropy, first in the Einstein–Smoluchowski form for position macrostates, and then in the more useful Langevin form for velocity macrostates. Following these the Fokker–Planck equation is discussed.

4.1.1 Position macrostates and Einstein's treatment

In this and the following sub-section we regard the Brownian particle's position $\mathbf{r}$ as labeling the macrostate of the system. The formalism for the time evolution of the macrostate is just the pure parity fluctuation case of chapter 3. The solvent in which the particle is embedded is treated as a reservoir and its microstates are not specified.

We shall suppose that the Brownian particle is trapped by a potential $U(\mathbf{r})$ centered on the origin, $U\prime(0) = 0$ and $U\prime\prime(0) > 0$, so that the average position vanishes,

$$\langle \mathbf{r} \rangle = 0.$$

(4.3)

The reason for requiring this is so that nonzero values of $\mathbf{r}$ can be treated as a fluctuation. With the pinning potential, $\langle r^2 \rangle < \infty$.

The limit $U(\mathbf{r}) \to 0$ gives results for the free Brownian particle. It will turn out that in this section the final results are independent of $U(\mathbf{r})$, and so this is a conceptually clear way of obtaining the free-particle results even in the thermodynamic limit.

For a solvent of temperature T, the reservoir entropy for the macrostate $\mathbf{r}$ is

$$S_{\mathrm{r}}(\mathbf{r}) = \frac{-U(\mathbf{r})}{T}.$$

(4.4)

For a Brownian particle, the reservoir entropy is the total entropy. The fluctuation matrix is the matrix of second derivatives of the total entropy,

$$S\prime\prime = \frac{-1}{T} U\prime\prime(0).$$

(4.5)

The probability is just the exponential of the entropy,

$$\wp(\mathbf{r}) = \frac{1}{Z(T)} \mathrm{e}^{S_{\mathrm{r}}(\mathbf{r})/k_{\mathrm{B}}T},$$

(4.6)

For a parabolic potential, this is a Gaussian in the position. Hence the average of the square of the position is

$$\langle \mathbf{r}\mathbf{r} \rangle = -k_{\mathrm{B}} S\prime\prime^{-1}$$

(4.7)

The second entropy in the pure parity case is now applied to the transition between positions of the Brownian particle. For this purpose equation (3.86) is

$$\mathbf{r}' = \mathbf{r} + \frac{|\tau|}{2}\Lambda_{\mathrm{r}}S\prime\prime\mathbf{r} + \tilde{\mathbf{R}}, \quad |\tau| \gg \tau_{\mathrm{molec}}.$$

Here τ_{molec} is a molecular time scale. Assuming a spherical particle, the transport coefficient is diagonal with the same values on the diagonal. It can be written as

$$\Lambda_{\mathrm{r}} = 2k_{\mathrm{B}}^{-1}D\,\mathrm{I}.$$

Here and below I is the 3 × 3 identity matrix. The quantity D is the diffusion constant, as will be justified by comparison with Einstein's result below.

This equation does not include explicitly any adiabatic contribution to the transition. For a small time step this would be $\mathbf{r}^0(\tau|\mathbf{r}) = \mathbf{r} + \tau\mathbf{v}$. In the present case this would vanish since the transition is not conditioned on the initial velocity, and its most likely or average value is zero, $\mathbf{v} = 0$. The direct adiabatic force due to the applied potential causes a change of particle velocity. But the particle velocity is projected out of the problem, taking with it the adiabatic force. These effects are subsumed in the reservoir entropy. That is, the effects of the external force are included in a cumulative manner in the dissipative term, $\nabla S_{\mathrm{r}} = S\prime\prime\mathbf{r}$.

The probability distribution for the stochastic part is the Gaussian, equation (3.88),

$$\wp(\tilde{\mathbf{R}}) = \frac{1}{Z\prime}\mathrm{e}^{-\Lambda_{\mathrm{r}}^{-1}:\tilde{\mathbf{R}}\tilde{\mathbf{R}}/2k_{\mathrm{B}}|\tau|}$$

$$= \frac{1}{(4\pi D\,|\tau|)^{3/2}}\mathrm{e}^{-\tilde{\mathbf{R}}\cdot\tilde{\mathbf{R}}/4D|\tau|}.$$

This is effectively the probability distribution of the step taken by a free Brownian particle over the time interval τ. It is the same as a step in a random walk. It is evidently a Gaussian, as sketched in figure 4.1, with the width increasing as the square root of the time interval.
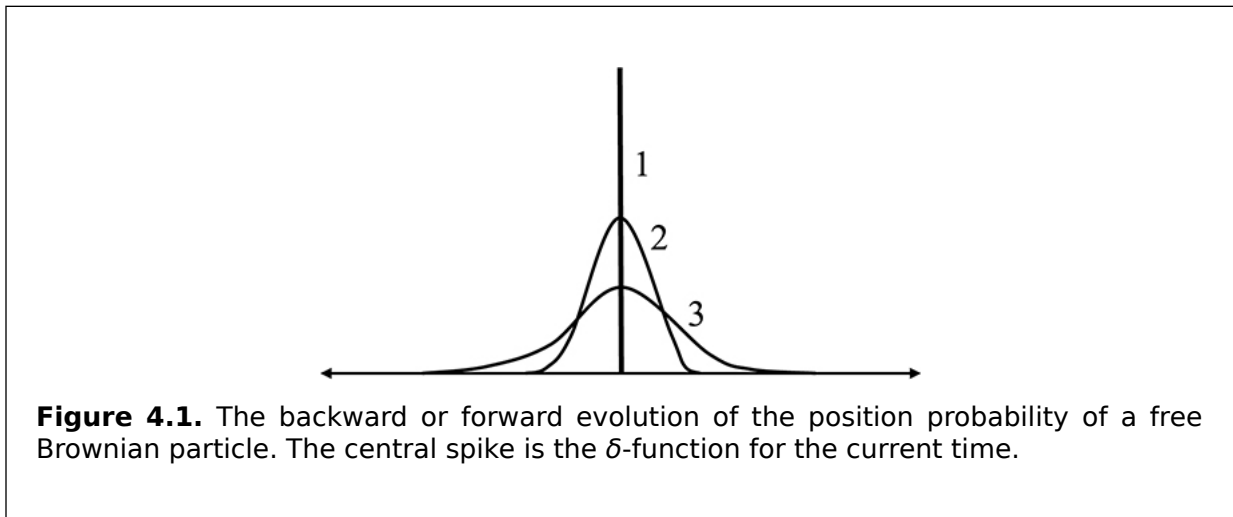
Figure 4.1. The backward or forward evolution of the position probability of a free Brownian particle. The central spike is the δ-function for the current time.

From this it follows that the stochastic change on average vanishes, $\langle\tilde{\mathbf{R}}\rangle = 0$, and it has variance

$$\left\langle \tilde{\mathbf{R}}\tilde{\mathbf{R}} \right\rangle = \mid \tau \mid k_{\mathrm{B}}\Lambda_{\mathrm{r}} = 2 \mid \tau \mid D\,\mathrm{I}.$$

This last result is a form of the fluctuation–dissipation theorem: the variance of the fluctuations is proportional to the strength of the dissipative term.

Note that there is no correlation between the stochastic change at successive time steps, $\left\langle \tilde{\mathbf{R}}(t+\tau)\tilde{\mathbf{R}}(t) \right\rangle = 0$. This means that this is a Markov system in which one can string successive transitions together to form a trajectory.

The correlation between the macrostate change over the transition and the initial macrostate is

$$
\begin{aligned}
\langle [\mathbf{r}(t+\tau) - \mathbf{r}(t)]\mathbf{r}(t) \rangle &= \frac{\mid \tau \mid}{2}\Lambda_{\mathrm{r}}S\prime\prime\langle \mathbf{r}(t)\mathbf{r}(t) \rangle \\
&= -k_{\mathrm{B}}\frac{\mid \tau \mid}{2}\Lambda_{\mathrm{r}}S\prime\prime S\prime\prime^{-1} \\
&= \frac{-k_{\mathrm{B}}\mid \tau \mid}{2}\Lambda_{\mathrm{r}} \\
&= -\mid \tau \mid D\mathrm{I}.
\end{aligned}
$$

This result is independent of the external potential $U(\mathbf{r})$. Hence taking the limit $U(\mathbf{r}) \to 0$ shows that this result holds also for the free Brownian particle.

One can write this in a more symmetric form as

$$
\begin{aligned}
&\langle [\mathbf{r}(t+\tau) - \mathbf{r}(t)][\mathbf{r}(t+\tau) - \mathbf{r}(t)] \rangle \\
&= \langle [\mathbf{r}(t+\tau) - \mathbf{r}(t)]\mathbf{r}(t+\tau) \rangle - \langle [\mathbf{r}(t+\tau) - \mathbf{r}(t)]\mathbf{r}(t) \rangle \\
&= 2 \mid \tau \mid D\mathrm{I}.
\end{aligned}
$$

Taking the trace of this dyadic matrix gives the result in the form originally given by Einstein (1905),

$$\langle [\mathbf{r}(t+\tau) - \mathbf{r}(t)] \cdot [\mathbf{r}(t+\tau) - \mathbf{r}(t)] \rangle = 6 \mid \tau \mid D.$$

The agreement justifies the identification of D as the diffusion constant for the particle in the particular solvent. One notes that the root mean square displacement of the Brownian particles grows with the square root of time. This is the well-known result for a random walk.

Differentiating the Einstein result with respect to τ yields

$$\langle \mathbf{v}(t+\tau) \cdot [\mathbf{r}(t+\tau) - \mathbf{r}(t)] \rangle = -\langle \mathbf{v}(t+\tau) \cdot \mathbf{r}(t) \rangle = 3\hat{\tau}D.$$

The first equality follows because variables of opposite time parity are instantaneously uncorrelated in an equilibrium system, $\langle \mathbf{v}(t+\tau) \cdot \mathbf{r}(t+\tau) \rangle = 0$. The result says that from the starting position, the particle moves toward the origin at constant velocity.

4.1.2 Langevin equation

We now turn to the stochastic dissipative equation given for Brownian motion by Langevin (1908). We shall derive the equations of motion from the pure parity second entropy result, equation (3.86), and show their equivalence to Langevin's result, giving an interpretation of his physical justification.

The Langevin equation treats the Brownian particle's position $\mathbf{r}$ and velocity $\mathbf{v}$ explicitly, and so we take $\{\mathbf{r}, \mathbf{v}\}$ as a macrostate of the system. The Hamiltonian is the sum of the kinetic and potential energies of the particle,

$$\mathcal{H}(\mathbf{r}, \mathbf{v}) = \mathcal{K}(\mathbf{v}) + U(\mathbf{r}), \quad \mathcal{K}(\mathbf{v}) = \frac{m}{2} v^2 = \frac{1}{2m} p^2,$$

(4.16)

where $\mathbf{p} = m\mathbf{v}$ is the particle's momentum and m is its mass.

The reservoir entropy for a position–velocity macrostate is just

$$S_{\mathrm{r}}(\mathbf{r}, \mathbf{v}) = \frac{-\mathcal{H}(\mathbf{r}, \mathbf{v})}{T},$$

(4.17)

and the fluctuation or second derivative matrix is

$$S'' = \frac{-1}{T} \begin{pmatrix} U''(0) & 0 \\ 0 & m \end{pmatrix}.$$

(4.18)

As has been mentioned, for the Brownian particle the reservoir entropy is the same as the total entropy.

Denote the trajectory of the Brownian particle in position–velocity space by $\{\mathbf{r}(t), \mathbf{v}(t)\}$. A transition over the time interval τ has the form

$$\mathbf{r}(t + \tau) = \mathbf{r}(t) + \tau \mathbf{v}(t)$$

$$\mathbf{v}(t + \tau) = \mathbf{v}(t) - \frac{\tau}{m} \nabla U(\mathbf{r}) + \frac{|\tau|}{m} \bar{\mathbf{F}}_{\mathrm{r}}(\mathbf{r}, \mathbf{v}) + \frac{1}{m} \tilde{\mathbf{F}}_{\mathrm{r}}(\tau).$$

(4.19)

This assumes that τ is of molecular or infinitesimal dimensions.

This transition contains both adiabatic and reservoir (solvent) contributions. The adiabatic contribution is

$$\mathbf{r}^0(\tau \mid \mathbf{r}, \mathbf{v}) = \mathbf{r} + \tau \mathbf{v},$$

$$\mathbf{v}^0(\tau \mid \mathbf{r}, \mathbf{v}) = \mathbf{v}(t) - \frac{\tau}{m} \nabla U(\mathbf{r}).$$

(4.20)

These are just Hamilton's equations of motion for the applied force $-\nabla U(\mathbf{r})$.

The reservoir contributions consist of a deterministic or most likely force $\bar{\mathbf{F}}_{\mathrm{r}}(\mathbf{r}, \mathbf{v})$, and a stochastic force $\tilde{\mathbf{F}}_{\mathrm{r}}(\tau)$. (The dependence of the stochastic force on the time interval will be obtained below.) These arise from the random (but statistically predictable) forces of the solvent molecules on the Brownian particle. These are real physical forces. These solvent forces effect the position evolution via the acceleration at $\mathcal{O}(\tau^2)$, which for infinitesimal τ is negligible over a single time step. The fact that the reservoir forces are irreversible (i.e. proportional to $|\tau|$ rather than to τ) is justified by the symmetry arguments used to derive the second entropy for the pure parity case, section 3.2.3.

Comparing this functional form to the second entropy result, equation (3.86), one sees several similarities and differences. First, the position evolution is purely adiabatic and is not treated as a stochastic transition governed by the second entropy. This is justified for infinitesimal time intervals, as argued above. The position is a 'slave' to the velocity over the time interval, which explains why the problem is treated as stochastic in the velocity

only. This is therefore a pure parity case, rather than a mixed parity case of position and velocity.

Second, whereas the present equation for the velocity contains both adiabatic and statistical contributions, the pure parity second entropy result, equation (3.86), contains only statistical contributions. This difference is explained by the fact that the fluctuations treated in chapter 3 were for macrostates with no external potential acting on the macrostate. Their evolution had to be treated statistically because the microstates necessary to determine adiabatic motion were unspecified. In the present case the applied potential acts directly on the transition between macrostates and this has to be directly accounted for.

For the present problem there is complete separation between the applied force acting on the Brownian particle, which gives the adiabatic contribution to the infinitesimal transition, and the solvent forces, which are characterized statistically by the second entropy. The adiabatic forces are reversible, $\propto \tau$, whereas the dissipative and stochastic forces are irreversible, $\propto |\tau|$, as the second entropy analysis showed.

Third, the present and former problem are treated at different time scales. The second entropy result, equation (3.86), was derived for small time intervals, which are much larger than the present infinitesimal time intervals. This difference in time scales raises the question of whether it is valid to apply the second entropy analysis over infinitesimal time scales as here. The original justification for requiring small time scales, say $|\tau| \gg |\tau_{\mathrm{molec}}|$, was that a fluctuation had an initial inertial period in which the velocities became organized into a steady flux. Time symmetry arguments showed that the term giving this steady flux was proportional to $|\tau|$ in the pure parity case. There are three arguments for extrapolating this result down to infinitesimal time scales in the present case. First, the present velocity macrostate is the end point of a continuous evolution, and the solvent already has had adequate time to become organized. This is particularly true in the case that the mass of the Brownian particle is much greater than the mass of the solvent molecules. Second, the relaxation time for velocity is much shorter than for position. And third, any term linearly proportional to $|\tau|$ can be broken down into a series of infinitesimal transitions each proportional to $|\tau_{\mathrm{molec}}| \ll |\tau|$, and over the longer term these in series return the original expression.

The fourth point in the comparison is that the second entropy approach explicitly determines the dissipative and stochastic forces in equation (3.86), whereas without additional assumptions these are undetermined in the Langevin equation. As an historical fact Langevin assumed that the dissipative force was the hydrodynamic drag force, and that the variance of the stochastic force was such that the equipartition theorem was satisfied. Such assumptions are quite plausible on physical grounds, and have been proven correct over a century of application. Nevertheless, it would be better to have a derivation of these from first principles, if for no other reason than to explain how it is possible that the hydrodynamic drag force can be applicable on molecular length scales. Also, a first principles derivation likely indicates the full generality of the stochastic dissipative equation, how to extend it to non-equilibrium systems, and how to apply them beyond the physical sciences to fields such as sociology, economics, financial markets, health, etc. Obviously such extensions cannot invoke hydrodynamic drag forces or the equipartition theorem.

This last point can be made explicit by applying equation (3.86) to the present problem. Adding the adiabatic transition, the stochastic dissipative equation for Brownian motion is

$$\mathbf{r}(t+\tau) = \mathbf{r}(t) + \tau \mathbf{v}(t)$$

$$\mathbf{v}(t+\tau) = \mathbf{v}(t) - \frac{\tau}{m}\nabla U(\mathbf{r}) + \frac{|\tau|}{2}\Lambda_v S\!\mathnormal{I\!I}_{vv}\mathbf{v} + \tilde{\mathbf{R}}_{\mathrm{v}}$$

$$= \mathbf{v}(t) - \frac{\tau}{m}\nabla U(\mathbf{r}) - \frac{|\tau|\gamma}{m}\mathbf{v} + \tilde{\mathbf{R}}_{\mathrm{v}}.$$

(4.21)

Here the entropy for the velocity macrostate is $S_{\mathrm{r}}(\mathbf{v}) = -mv^2/2T$, which has second derivative, $S\!\mathnormal{I\!I}_{vv} = -m\mathbf{I}/T$, the transport matrix is proportional to the identity matrix, $\Lambda_v = 2m^{-2}T\gamma\mathbf{I}$, and the stochastic force is Gaussian distributed, equation (3.88), $\wp(\tilde{\mathbf{R}}_{\mathrm{v}}) = Z'^{-1}\mathrm{e}^{-\Lambda_v^{-1}:\tilde{\mathbf{R}}_{\mathrm{v}}\tilde{\mathbf{R}}_{\mathrm{v}}/2k_{\mathrm{B}}|\tau|} = Z'^{-1}\mathrm{e}^{-m^2\tilde{R}_v^2/4\gamma|\tau|k_{\mathrm{B}}T}$. Shortly γ will be identified with the friction or drag coefficient.

One sees that in the second entropy theory the dissipative force in general takes the form of a drag force, $\bar{\mathbf{F}}_{\mathrm{r}} \propto \mathbf{v}$. It is not that the dissipative force arises from hydrodynamic drag, but rather that hydrodynamic drag forces themselves are dissipative forces. The very general statistical considerations of the second entropy show that dissipative forces are proportional to the gradient of the entropy. Since the entropy contains the kinetic energy, which is quadratic in the velocity, its velocity gradient is proportional to the velocity. The present result for a Brownian particle in a solvent would also apply to a Brownian particle in a solid, where no hydrodynamic forces are present. (Friction forces are also dissipative forces that arise from the gradient of the entropy.) Even though hydrodynamic forces and friction forces only hold on macroscopic length scales, dissipative forces are statistical in origin and apply on molecular and even sub-molecular length scales.

In the second entropy analysis there is a single parameter—the transport coefficient—that determines both the variance of the stochastic force and the magnitude of the dissipative force. In contrast, conventional treatments of the theory of Langevin (1908) regard the two forces as separate and determine the variance of the stochastic force from the equipartition theorem (cf Pathria 1972 section 13.4).

Figure 4.2 shows the most likely trajectory of a Brownian particle. This contains the adiabatic contributions, which are time reversible, and the dissipative contributions, which are time irreversible. Hence the discontinuous first derivative in the trajectory at $t = 0$. The stochastic contributions perturb the trajectory, increasingly so as time progresses or regresses from the present.

Figure 4.2. The most likely trajectory of a Brownian particle forward and backward in time from the present time (circle). The shaded areas indicate the stochastic contribution within a one-standard deviation envelope.
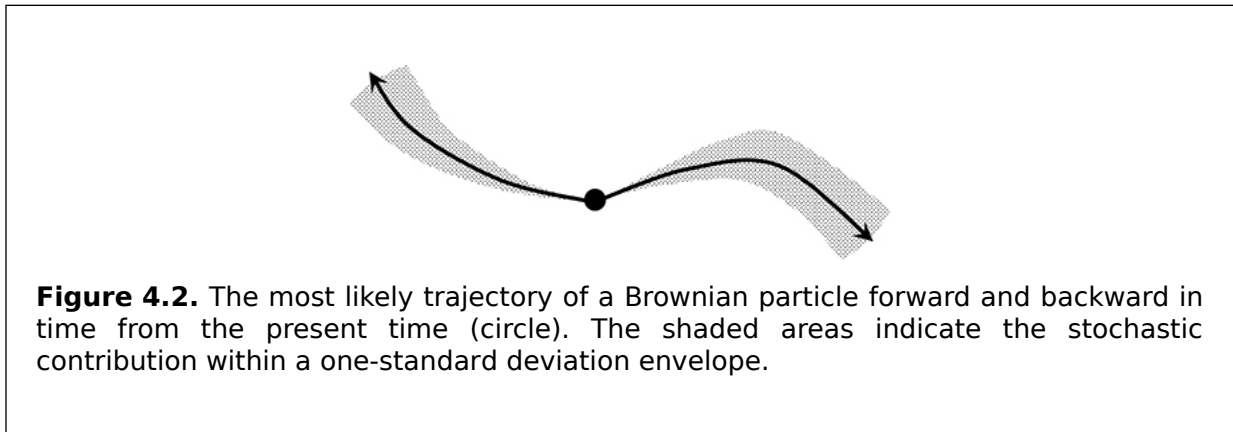
The discontinuity in the derivative of the trajectory means that one has to specify whether the time derivative of a function on the trajectory is the forward or the backward derivative,

$$\frac{\mathrm{d}^{\pm} f(\mathbf{v}(t), t)}{\mathrm{d}t} = \frac{\partial f(\mathbf{v}(t), t)}{\partial t} + \dot{\mathbf{v}}^{\pm}(t) \cdot \nabla_{v} f(\mathbf{v}(t), t)$$

$$= \frac{\partial f(\mathbf{v}(t), t)}{\partial t} + \frac{\mathbf{v}(t \pm |\tau|) - \mathbf{v}(t)}{\pm |\tau|} \cdot \nabla_{v} f(\mathbf{v}(t), t)$$

$$= \frac{\partial f(\mathbf{v}(t), t)}{\partial t} + \left[\frac{-1}{m} \nabla U(\mathbf{r}) \mp \frac{\gamma}{m} \mathbf{v} \pm \frac{1}{|\tau|} \tilde{\mathbf{R}}_{\mathrm{v}} \right] \cdot \nabla_{v} f(\mathbf{v}(t), t).$$

The upper sign refers to the forward derivative and the lower sign to the backward derivative.

4.1.3 Markov behavior

The second entropy analysis of Brownian motion in terms of position (i.e. Einstein's result) is consistent with the analysis in terms of velocity (i.e. Langevin's result). This may be seen as follows.

For a free Brownian particle, $U(\mathbf{r}) = 0$, the most likely rate of change of velocity given by equation (4.21) is

$$\dot{\overline{\mathbf{v}}}(t) = \frac{-\hat{\tau} \gamma}{m} \mathbf{v}(t).$$

Since τ is molecular scale, this takes the coarse velocity to equal the instantaneous velocity, $[\mathbf{v}(t + \tau) - \mathbf{v}(t)]/\tau = \dot{\mathbf{v}}(t)$.

According to this, the rate of change of velocity depends only on the current velocity, and so this is a Markov process. The solution is

$$\overline{\mathbf{v}}(t) = \mathrm{e}^{-|t| \gamma/m} \mathbf{v}(0).$$

This says that the velocity decays exponentially from the current value (and to the current value).

The time correlation of the velocity is also exponential,

$$\langle \mathbf{v}(t) \cdot \mathbf{v}(0) \rangle = \langle \overline{\mathbf{v}}(t) \cdot \mathbf{v}(0) \rangle$$

$$= \mathrm{e}^{-|t| \gamma/m} \langle \mathbf{v}(0) \cdot \mathbf{v}(0) \rangle$$

$$= \frac{3 k_{\mathrm{B}} T}{m} \mathrm{e}^{-\gamma |t|/m}.$$

The final equality follows because $\langle \mathbf{v} \mathbf{v} \rangle = -k_{\mathrm{B}} S''^{-1}_{vv} = (k_{\mathrm{B}} T/m)\, \mathbf{I}$, which is equivalent to the equipartition theorem.

Now in the analysis of the Brownian particle from the perspective of position, the Einstein result linked the diffusion constant to the correlation of velocity and position, equation (4.15). Writing τ_r as the time step used for the position equation (the interval for velocity is much smaller, as sketched in figure 4.3), that result may be written as

$$D = \frac{\hat{\tau}_r}{3} \langle \mathbf{v}(t + \tau_r) \cdot [\mathbf{r}(t + \tau_r) - \mathbf{r}(t)] \rangle$$

$$= \frac{\hat{\tau}_r}{3} \int_0^{\tau_r} dt\prime \, \langle \mathbf{v}(t + \tau_r) \cdot \mathbf{v}(t + t\prime) \rangle$$

$$= \frac{\hat{\tau}_r}{3} \int_0^{\tau_r} dt\prime \, \frac{3k_BT}{m} e^{-\gamma|\tau_r - t\prime|/m}$$

$$= \frac{\hat{\tau}_r k_B T}{m} \frac{m\hat{\tau}_r}{\gamma} \left[1 - e^{-\gamma|\tau_r|/m} \right]$$

$$= \frac{k_B T}{\gamma}, \quad |\tau_r| \gg \frac{m}{\gamma}.$$

This says that the diffusion constant is inversely proportional to the drag constant provided that the time interval for positional diffusion is relatively large. This shows the consistency of the Einstein result for free-particle diffusion with the Langevin result for the evolution of the free-particle velocity.
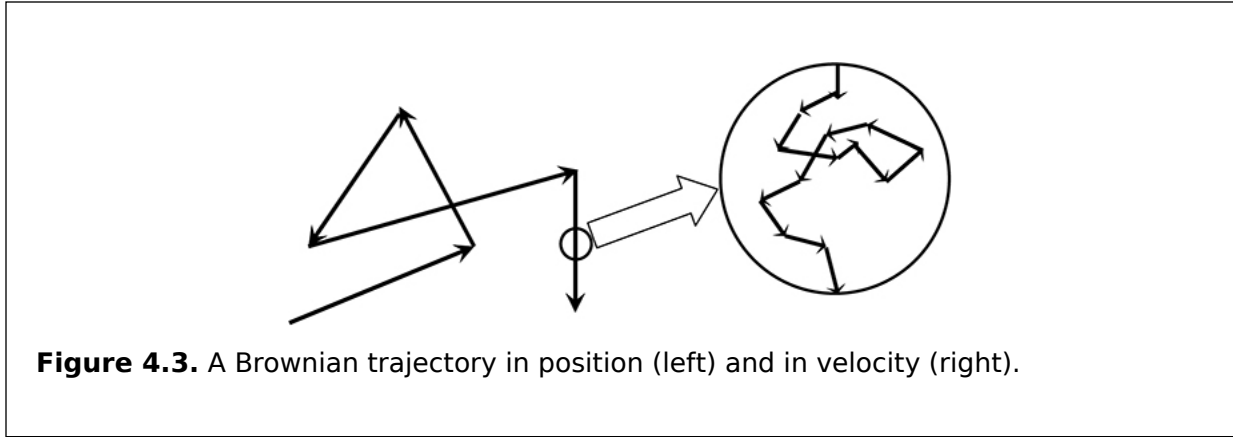
Figure 4.3. A Brownian trajectory in position (left) and in velocity (right).

One can also find the regime of validity for a pinned Brownian particle. For the case of a parabolic pinning potential,

$$U(\mathbf{r}) = \frac{\kappa}{2} r^2,$$

the reservoir entropy is

$$S(\mathbf{r}, \mathbf{v}) = \frac{-m}{2T} v^2 - \frac{\kappa}{2T} r^2.$$

In this case the stochastic dissipative equation of motion for the position, equation (4.8), becomes

$$\mathbf{r}\prime = \mathbf{r} - |\tau_r| \frac{\kappa D}{k_B T} \mathbf{r} + \tilde{\mathbf{R}}, \quad |\tau_r| \gg \tau_{\text{molec}}.$$

In order for the discretization to be valid, the change in position must be small compared to the position itself, which means

$$\frac{m}{\gamma} \ll \mid \tau_r \mid \ll \frac{k_{\mathrm{B}}T}{\kappa D} = \frac{\gamma}{\kappa}.$$

(4.30)

The lower limit is that obtained above for the free particle from the Markov analysis of velocity. This says that if the drag coefficient is large, or the curvature of the potential is small, then there is a time interval in which the stochastic dissipative equation of motion for the position can validly be applied in the case of an applied potential.

4.2 Fokker–Planck equation

We shall now derive the Fokker–Planck equation in general for stochastic dissipative equations of motion, and then use it to show that the Langevin equation yields a stationary equilibrium probability distribution when the fluctuation–dissipation theorem is satisfied. We shall also show that the stochastic dissipative equation derived from the second entropy analysis also yields a stationary probability distribution. For a more specialised treatment of the Fokker–Planck equation, see Risken (1984).

In the present book, the probability is taken to be the exponential of the entropy, and so it does not make much sense to speak of the evolution of an arbitrary probability density in the equilibrium case. Similarly, the transition probability is taken to be fixed by the exponential of the second entropy, which in turn fixes the stochastic, dissipative equations of motion. So again it is strange to speak of arbitrary equations of motion.

Nevertheless, one of the benefits that arise from the following derivation of the Fokker–Planck equation is a consistency rule that arbitrary stochastic dissipative equations of motion have to obey in order for the equilibrium probability density to be stationary. This leads to the fluctuation–dissipation theorem independent of the second entropy. It is shown that Langevin's equation not only has to have such a quantitative relationship, but that the qualitative form of the equation is fixed and unchangeable. Also the Fokker–Planck equation provides a basis for discussing Liouville's theorem.

4.2.1 General expression

We consider a generic stochastic dissipative transition,

$$X\prime = X + \tau \dot{X}^{\mathrm{det}} + \tilde{R}.$$

(4.31)

Here the macrostate X is to be considered quite generally. It can be a single coordinate, or a vector of multiple coordinates, possibly of mixed parity. The deterministic part of the transition consists of the adiabatic evolution and the most likely contribution due to the reservoir. For future use, the stochastic part is the difference between the actual and the most likely destination, $X\prime - \overline{X\prime} = \tilde{R}$. (We habitually use the word 'stochastic' to mean the zero mean part of the random reservoir contribution. Strictly speaking, the entire reservoir contribution is stochastic; it can be divided into a most likely deterministic part, $\bar{R} = \tau[\dot{X}^{\mathrm{det}} - \dot{X}^0]$, and a zero mean random part, $\tilde{R}$. It is the latter that we mean by the word 'stochastic'.)

We shall suppose that the probability distribution for the stochastic part of the transition is Gaussian

$$\wp(\tilde{R}) = \frac{1}{Z\prime}\mathrm{e}^{-\Lambda^{-1}:\tilde{R}\tilde{R}/2k_{\mathrm{B}}|\tau|}, \quad Z\prime \equiv \{\mathrm{Det}\ [2\pi k_{\mathrm{B}} \mid \tau \mid \Lambda]\}^{1/2}.$$

Clearly $\langle \tilde{R} \rangle = 0$. This Gaussian is just the conditional transition probability, which means
that the unconditional transition probability can be written as

$$\wp(X\prime, X \mid \tau) = \wp(X\prime| X, \tau)\wp(X) = \wp(\tilde{R})\wp(X).$$

With these, and using the conservation law for probability in the equilibrium case, equation (3.17), one has

$$
\begin{aligned}
\wp(X\prime, t + \tau) &= \int dX \, \wp(X\prime, X \mid \tau) \\
&= \frac{1}{Z\prime} \int dX \, e^{-\Lambda^{-1}:\left[X\prime-\overline{X\prime}\right]^2 / 2k_B|\tau|} \wp(X, t) \\
&= \frac{1}{Z\prime} \int d\overline{X\prime} \, |\frac{d\overline{X\prime}}{dX}|^{-1} e^{-\Lambda^{-1}:\left[X\prime-\overline{X\prime}\right]^2 / 2k_B|\tau|} \\
&\quad \times \{\wp(X\prime, t) + (X - X\prime) \cdot \nabla \wp(X\prime, t) \\
&\quad + \frac{1}{2}(X - X\prime)(X - X\prime) : \nabla\nabla\wp(X\prime, t) + \cdots\}.
\end{aligned}
$$

The evolution of a volume element is sketched in figure 4.4. With this and a little thought (cf equation (5.7)) one can see that the compressibility of the equations of motion is

$$
\begin{aligned}
|\frac{d\overline{X\prime}}{dX}| &= |I + \tau \frac{d\dot{X}^{det}}{dX}| \\
&= 1 + \tau \, \mathrm{Tr} \, \frac{d\dot{X}^{det}}{dX} + \mathcal{O}(\tau^2) \\
&= 1 + \tau \nabla \cdot \dot{X}^{det}.
\end{aligned}
$$

Note that the adiabatic equations of motion are incompressible, $\nabla \cdot \dot{X}^0 = 0$, and so only the dissipative part contributes to this. Writing $X\prime - X = \tau \dot{X}^{det} + \tilde{R}$, and neglecting quadratic powers of $\tau \dot{X}^{det}$ and odd powers of $\tilde{R}$, the evolution becomes

$$\wp(X\prime,t+\tau) = \frac{1}{Z\prime}\left[1-\tau\nabla\cdot\dot{X}^{\mathrm{det}}\right]\int\mathrm{d}\tilde{R}\,\mathrm{e}^{-\Lambda^{-1}:\tilde{R}\tilde{R}/2k_{\mathrm{B}}|\tau|}\Big\{\wp(X\prime,t)$$

$$-(\tau\dot{X}^{\mathrm{det}}+\tilde{R})\cdot\nabla\wp(X\prime,t)+\frac{1}{2}(\tau\dot{X}^{\mathrm{det}}+\tilde{R})(\tau\dot{X}^{\mathrm{det}}+\tilde{R}):\nabla\nabla\wp(X\prime,t)$$

$$=\frac{1}{Z\prime}\left[1-\tau\nabla\cdot\dot{X}^{\mathrm{det}}\right]\int\mathrm{d}\tilde{R}\,\mathrm{e}^{-\Lambda^{-1}:\tilde{R}\tilde{R}/2k_{\mathrm{B}}|\tau|}\Big\{\wp(X\prime,t)$$

$$-\tau\dot{X}^{\mathrm{det}}\cdot\nabla\wp(X\prime,t)+\frac{1}{2}\tilde{R}\tilde{R}:\nabla\nabla\wp(X\prime,t)\Big\}$$

$$=\wp(X\prime,t)-\tau\nabla\cdot\dot{X}^{\mathrm{det}}\,\wp(X\prime,t)-\tau\dot{X}^{\mathrm{det}}\cdot\nabla\wp(X\prime,t)+\frac{k_{\mathrm{B}}\,|\,\tau\,|}{2}\Lambda:\nabla\nabla\wp($$

(4.3
6)

The neglected terms are $\mathcal{O}(\tau^2)$. To leading order $\tau\dot{X}^{\mathrm{det}}$ can be evaluated at $X\prime$, which is why it has been taken outside of the integral twice here. This gives the partial time derivative as

$$\frac{\partial\wp(X,t)}{\partial t}=-[\nabla\cdot\dot{X}^{\mathrm{det}}]\wp(X,t)-\dot{X}^{\mathrm{det}}\cdot\nabla\wp(X,t)+\frac{\hat{\tau}k_{\mathrm{B}}}{2}\Lambda:\nabla\nabla\wp(X,t).$$

(4.3
7)

This is the Fokker–Planck equation.
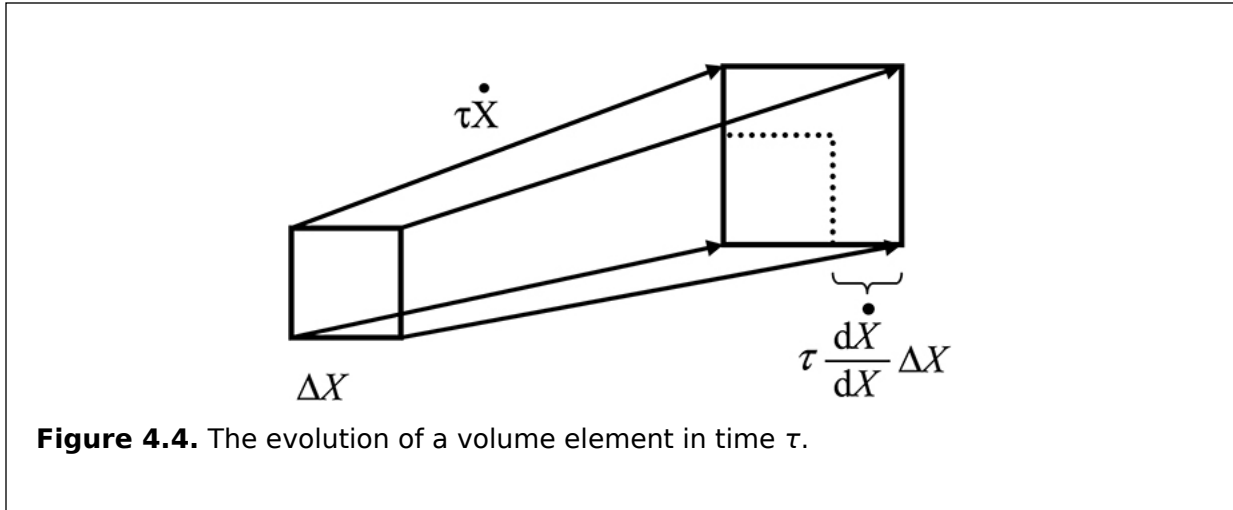
Figure 4.4. The evolution of a volume element in time τ.

Note that this partial time derivative on the left-hand side does not distinguish notationally between the forward and backward time derivatives. In contrast the right-hand side contains irreversible terms in the dissipative contribution to the deterministic velocity and in the stochastic contribution, with $\hat{\tau}\equiv\tau/\,|\,\tau\,|=+1$ for the forward time derivative and $\hat{\tau}=-1$ for the backward time derivative. It will be shown next (and also in section 5.4.3) that for equilibrium systems, the right-hand side is zero with the irreversible terms canceling each other. (In chapter 6 the same cancelation will be shown to occur for non-equilibrium steady state systems.)

4.2.2 Second entropy

The equilibrium probability is just the exponential of the total entropy,

$$\wp(X, t) = \frac{1}{Z} e^{S(X)/k_{\mathrm{B}}}.$$

In the second entropy case, the deterministic part of the transition is

$$\dot{X}^{\mathrm{det}} = \dot{X}^0 + \frac{\hat{\tau}}{2} \Lambda \nabla S(X).$$

Because in the equilibrium case $S(X) = -\mathcal{H}(X)/T$ it follows that $\dot{X}^0 \cdot \nabla S(X) = 0$. This assumes that the sub-system state X has no internal entropy, so that the total entropy is the same as the reservoir entropy. This is a valid assumption for a Brownian particle, or for a point in classical phase space, but it does not hold, for example, for a fluctuation of a non-conserved variable in the sub-system, such as the first energy moment treated in section 3.4.

A second result required is that the adiabatic equations of motion are incompressible, $\nabla \cdot \dot{X}^0 = 0$. These two results mean that there is no adiabatic contribution to the Fokker–Planck equation. What remains is

$$\frac{\partial \wp(X, t)}{\partial t} = \left\{ \frac{-\hat{\tau}}{2} \Lambda : \nabla \nabla S(X) - \frac{\hat{\tau}}{2k_{\mathrm{B}}} \Lambda : \nabla S(X) \nabla S(X) \right.$$
$$\left. + \frac{\hat{\tau}}{2k_{\mathrm{B}}} \Lambda : \nabla S(X) \nabla S(X) + \frac{\hat{\tau}}{2} \Lambda : \nabla \nabla S(X) \right\} \wp(X, t)$$
$$= 0.$$

This vanishes identically, as it must in the present equilibrium case.

4.2.2.1 Mixed parity

In the second entropy, mixed parity case, the deterministic part of the transition is given by equation (3.135) with the adiabatic evolution added,

$$\dot{X}^{\mathrm{det}} = \dot{X}^0 + \frac{\hat{\tau}}{2} [\Lambda - 2\hat{\tau}\Theta] S\prime\prime X \equiv \dot{X}^0 + \frac{\hat{\tau}}{2} L(\hat{\tau}) \nabla S(X).$$

The random forces remain Gaussian distributed with the variance $\Lambda k_{\mathrm{B}} \mid \tau \mid$ unchanged from the pure parity case.

One can use the above pure parity result for the partial time derivative of the probability with the first two transport matrices replaced by $L(\hat{\tau})$,

$$\frac{\partial \wp(X,t)}{\partial t} = \left\{ \frac{-\hat{\tau}}{2} L(\hat{\tau}) : \nabla\nabla S(X) - \frac{\hat{\tau}}{2k_{\mathrm{B}}} L(\hat{\tau}) : \nabla S(X)\nabla S(X) \right.$$

$$\left. + \frac{\hat{\tau}}{2k_{\mathrm{B}}} \Lambda : \nabla S(X)\nabla S(X) + \frac{\hat{\tau}}{2} \Lambda : \nabla\nabla S(X) \right\} \wp(X,t)$$

$$= \left\{ \frac{-\hat{\tau}}{2} \Lambda : \nabla\nabla S(X) - \frac{\hat{\tau}}{2k_{\mathrm{B}}} \Lambda : \nabla S(X)\nabla S(X) \right.$$

$$\left. + \frac{\hat{\tau}}{2k_{\mathrm{B}}} \Lambda : \nabla S(X)\nabla S(X) + \frac{\hat{\tau}}{2} \Lambda : \nabla\nabla S(X) \right\} \wp(X,t)$$

$$= 0.$$

(4.42)

The second equality follows because any symmetric scalar product with the antisymmetric matrix Θ vanishes. Hence for both the pure and mixed parity case, the second entropy formulation of the stochastic dissipative equations of motion yield a stationary equilibrium probability distribution.

4.2.3 Dissipation in the Langevin equation

Now it is shown that the solvent force, $\mathbf{F}_{\mathrm{r}}$, in the Langevin equation must necessarily be of the form of a drag force in order for the equilibrium probability distribution to be stable.

In the case of the Langevin equation, $X \equiv \{\mathbf{r}, \mathbf{v}\}$, the deterministic rates of change are

$$\dot{\mathbf{r}}^{\mathrm{det}} = \dot{\mathbf{r}}^0$$

$$\dot{\mathbf{v}}^{\mathrm{det}} = \dot{\mathbf{v}}^0 + \frac{\hat{\tau}}{m} \mathbf{F}_{\mathrm{r}}(\mathbf{r}, \mathbf{v}).$$

(4.43)

The velocity gradient of the total entropy is $\nabla_v S = -m\mathbf{v}/T$. Again using the equilibrium probability, $\wp(\mathbf{r}, \mathbf{v}; t) = e^{S(\mathbf{r}, \mathbf{v})/k_{\mathrm{B}}}/Z$, the Fokker–Planck equation in this case is

$$\frac{\partial \wp(X,t)}{\partial t} = \left\{ \frac{-\hat{\tau}}{m} \nabla_v \cdot \mathbf{F}_{\mathrm{r}} - \frac{\hat{\tau}}{mk_{\mathrm{B}}} \mathbf{F}_{\mathrm{r}} \cdot \nabla_v S(X) \right.$$

$$\left. + \frac{\hat{\tau}}{2k_{\mathrm{B}}} \Lambda_v : \nabla_v S(X)\nabla_v S(X) + \frac{\hat{\tau}}{2} \Lambda_v : \nabla_v \nabla_v S(X) \right\} \wp(X,t)$$

$$= \left\{ \frac{-\hat{\tau}}{m} \nabla_v \cdot \mathbf{F}_{\mathrm{r}} + \frac{\hat{\tau}}{k_{\mathrm{B}}T} \mathbf{F}_{\mathrm{r}} \cdot \mathbf{v} + \frac{\hat{\tau}m^2}{2k_{\mathrm{B}}T^2} \Lambda_v : \mathbf{v}\mathbf{v} - \frac{\hat{\tau}m}{2T} \mathrm{Tr}\,\Lambda_v \right\} \wp(X,t)$$

$$= \left\{ \frac{-\hat{\tau}}{m} \nabla_v \cdot \mathbf{F}_{\mathrm{r}} + \frac{\hat{\tau}}{k_{\mathrm{B}}T} \mathbf{F}_{\mathrm{r}} \cdot \mathbf{v} + \frac{\hat{\tau}m^2 \lambda_v}{2k_{\mathrm{B}}T^2} \mathbf{v} \cdot \mathbf{v} - \frac{3\hat{\tau}m\lambda_v}{2T} \right\} \wp(X,t),$$

(4.44)

since $\Lambda_v = \lambda I$. To make this vanish for all $\mathbf{v}$, it is clear that one must take

$$\mathbf{F}_{\mathrm{r}} = -\gamma \mathbf{v},$$

(4.45)

which gives

$$\frac{\partial \wp(X,t)}{\partial t} = \left\{ \frac{3\hat{\tau}}{m} \gamma - \frac{\hat{\tau}}{k_{\mathrm{B}}T} \gamma \mathbf{v} \cdot \mathbf{v} + \frac{\hat{\tau}m^2 \lambda_v}{2k_{\mathrm{B}}T^2} \mathbf{v} \cdot \mathbf{v} - \frac{3\hat{\tau}m\lambda_v}{2T} \right\} \wp(X,t).$$

This vanishes when the strength of the dissipation, the drag force, is related to the strength of the fluctuations via

$$\gamma = \frac{m^2 \lambda_v}{2T}.$$

This is another version of the fluctuation–dissipation theorem. This theorem was originally derived by Langevin (1908) by demanding that the equipartition theorem be satisfied. Since the equilibrium probability yields the equipartition theorem, the present derivation based on the stationarity of the former includes the latter as a special case.

4.2.4 No dissipation without fluctuation

The Langevin equation for velocity, equation (4.19), may be rewritten in terms of momentum $\mathbf{p} = m\mathbf{v}$ as

$$\dot{\mathbf{p}} = \frac{-\partial \mathcal{H}(\mathbf{p}, \mathbf{q})}{\partial \mathbf{q}} - \frac{\gamma}{m}\mathbf{p} + \frac{1}{\tau}\tilde{\mathbf{F}}_{\mathrm{r}}(\tau).$$

Here the position is written $\mathbf{q} \equiv \mathbf{r}$, and it has been assumed that τ is infinitesimal. This may be interpreted as applying to the coordinates of a single Brownian particle, or to the phase space of the molecules of a system.

The fluctuation–dissipation theorem says that the variance of the stochastic force is

$$\left\langle \tilde{\mathbf{F}}_{\mathrm{r}} \tilde{\mathbf{F}}_{\mathrm{r}} \right\rangle = m^2 k_{\mathrm{B}} \mid \tau \mid \lambda_v \mathbf{I} = \frac{2k_{\mathrm{B}}T}{\gamma} \mid \tau \mid \mathbf{I}.$$

Comparing the above stochastic dissipative Langevin equation with the dissipative equations of motion with the isokinetic thermostat given at the start of the chapter, equation (4.1), one sees that the latter has the same dissipative term, $\alpha \equiv \gamma/m$, but it completely neglects the stochastic term, $\langle \tilde{\mathbf{F}}_{\mathrm{r}} \tilde{\mathbf{F}}_{\mathrm{r}} \rangle = 0$. This is a violation of the fluctuation–dissipation theorem.

The consequence of this violation is that the equilibrium probability distribution *cannot* be stationary under the isokinetic equations of motion despite (or because of) the fact that the kinetic energy is a constant of the motion and the equipartition theorem is obeyed. It follows that any property of the system (apart from the kinetic energy) that is obtained as a time average over the isokinetic trajectory cannot be equal to its correct average value obtained with the proper probability distribution.

The approach of using deterministic equations of motion—dissipation without fluctuation—is very commonly used for non-equilibrium molecular dynamics simulations. Although we shall not analyze non-equilibrium systems until chapter 6, one can anticipate that any dissipative equations of motion for the non-equilibrium system based upon an isokinetic or any similar deterministic thermostat, and any other deterministic term designed to mimic a particular thermodynamic flow, in the absence of a stochastic contribution that obeys a generalized fluctuation–dissipation theorem, will likewise yield erroneous results.

4.3 Stochastic calculus

The stochastic calculus refers to that branch of mathematics that deals with integrals and derivatives of functions of stochastic variables. The issue addressed here is whether or not the stochastic calculus is relevant to statistical mechanics in the physical sciences. For the

conventional view of the stochastic calculus and the Fokker–Planck equation, see Gardener (1983) or Risken (1984); for a more detailed discussion of what follows, see Attard (2012, section 11.3.3).

There are two commonly used versions of the stochastic calculus, one due to Itô and the other to Stratonovitch, as is now explained. In the simplest case, consider a trajectory with a random displacement added at each discrete node, $t_n = n\Delta_t$,

$$x_n = \bar{x}(t_n) + \tilde{R}_n,$$

(4.50)

or

$$x_{n+1} = x_n + \Delta_t \dot{\bar{x}}(t_n) + [\tilde{R}_{n+1} - \tilde{R}_n].$$

(4.51)

We shall return to the variance of the increment in the random displacement below. For the present purpose of connecting with the mathematics literature on the stochastic calculus, one can consider it to be linear in the time interval, $\langle [\tilde{R}_{n+1} - \tilde{R}_n]^2 \rangle = 2D \mid \Delta_t \mid$, as in Einstein's treatment of a free Brownian particle.

The point is that the above stochastic equation specifies the trajectory at the nodes, but it is silent about what happens between the nodes. One is free to make any convenient choice, but any such choice has consequences.

The Itô calculus says that the random displacement is applied entirely at the beginning of the node, and so between nodes one has

$$x(t) = \bar{x}(t) + \tilde{R}_n, \quad t_n < t < t_{n+1}.$$

(4.52)

The Stratonovitch calculus says that between nodes the average random displacement at the termini is applied,

$$x(t) = \bar{x}(t) + \frac{1}{2}[\tilde{R}_{n+1} + \tilde{R}_n], \quad t_n < t < t_{n+1}.$$

(4.53)

These are sketched in figure 4.5. These are not the only choices that can be made, but they are the most common.

Figure 4.5. From top to bottom, (a) the most likely trajectory with four nodes, (b) four random displacements, (c) the Itô trajectory, and (d) the Stratonovitch trajectory.
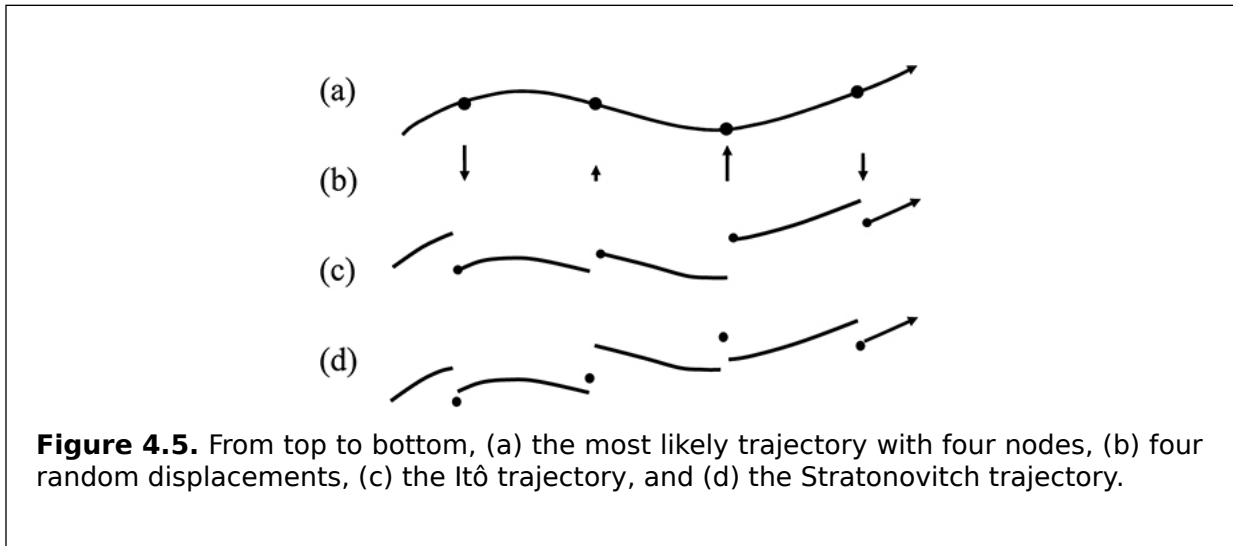
The choice of calculus makes a difference for the integral of a function over a trajectory. One has

$$G(t) = \int_{x_0}^{x(t)} \mathrm{d}x(t\prime)\, g(x(t\prime))$$

$$= \sum_{i=0}^{n-1} [x_{i+1} - x_i]\, g(x(t\prime)),$$

$$= \sum_{i=0}^{n-1} [\Delta_t \dot{\bar{x}}(t_i) + \tilde{R}_{i+1} - \tilde{R}_i]\, g(x(t\prime)).$$

(4.54)

Here $t\prime = t_i + \alpha\Delta_t$, and $\alpha \in [0,1)$. Ordinarily the precise value of α is immaterial; different choices of $\alpha \in [0,1)$ change $g(x(t\prime))$ by $\mathcal{O}(\Delta_t)$, and hence $G(t)$ by $\mathcal{O}(n\Delta_t^2) = \mathcal{O}(\Delta_t)$, which is negligible in the limit $\Delta_t \to 0$.

In the stochastic calculus the choice of α makes a difference to the final result. To see this, note that the value of the function within the interval differs between the two versions,

$$g(x(t\prime)) = \begin{cases} g(x_i) + [\alpha\Delta_t\dot{\bar{x}}(t_i) + \tilde{R}_i]g\prime(x_i) + \cdots, & \text{Itô,} \\ g(x_i) + [\alpha\Delta_t\dot{\bar{x}}(t_i) + (\tilde{R}_i + \tilde{R}_{i+1})/2]g\prime(x_i) + \cdots, & \text{Strat.} \end{cases}$$

(4.55)

Subtracting one version of the integral from the other and averaging yields a result proportional to the variance of the increment in the random displacement

$$\left\langle G^{\text{Strat.}}(t) - G^{\text{Ito}}(t) \right\rangle = \sum_{i=0}^{n-1} \left\langle [\Delta_t\dot{\bar{x}}(t_i) + R_{i+1} - R_i]\,[(\tilde{R}_{i+1} - \tilde{R}_i)/2]g\prime(x_i) \right\rangle$$

$$= \frac{1}{2}\sum_{i=0}^{n-1} \left\langle [\tilde{R}_{i+1} - \tilde{R}_i]^2 \right\rangle g\prime(x_i)$$

$$= D\,|\,\Delta_t\,|\sum_{i=0}^{n-1} g\prime(x_i).$$

(4.56)

Since there are $n = t/\Delta_t$ terms here, the difference between the two version of the integral is $\mathcal{O}(1)$, which cannot be neglected even for an infinitesimal time step. This shows that one has to make a particular choice for the stochastic calculus, either Itô or Stratonovitch, or some other, and the value of any integral over a stochastic trajectory is dependent on the choice.

In the physical sciences this situation is unacceptable because one cannot have any such ambiguity in the value of physical variables.

The resolution of the problem comes from noticing that in the final equality above it was assumed that the variance is twice the diffusion constant times the length of the time step. This corresponds to the Einstein result, that the variance in the displacement of a free Brownian particle scales with the length of the time interval. In the mathematics literature, and particularly in financial applications of stochastic difference equations, this is taken as an axiomatic definition of stochastic behavior. It is also assumed that the time interval can

be regarded as infinitesimal on the scale of the considered motion (otherwise the discretization of the trajectory would yield inaccurate results).

In fact, however, in statistical mechanics as applied to real physical systems, the variance must equal the position self-correlation function, and this has different asymptotic behavior in different time regimes. As discussed in connection with figure 3.2, one has

$$\left\langle [\tilde{R}_{n+1} - \tilde{R}_n]^2 \right\rangle = 2k_{\mathrm{B}} q_0^{\mathrm{s}}(\Delta_t)$$

$$\sim \begin{cases} k_{\mathrm{B}} \ddot{q}_0^{\mathrm{s}}(0)\Delta_t^2, & \Delta_t \to 0, \\ 2D \mid \Delta_t \mid, & \Delta_t \gg \tau_{\mathrm{molec}}. \end{cases}$$

(4.57)

As in equation (4.13), the symmetrized position autocorrelation function is

$$q_0^{\mathrm{s}}(\tau) \equiv \frac{1}{2k_{\mathrm{B}}} \left\langle [x(t+\tau) - x(t)] \, [x(t+\tau) - x(t)] \right\rangle_0,$$

(4.58)

and its 'acceleration' is

$$\ddot{q}_0^{\mathrm{s}}(\tau) = \frac{-1}{2k_{\mathrm{B}}} [\langle \ddot{x}(t+\tau)x(t)\rangle_0 + \langle x(t)\ddot{x}(t+\tau)\rangle_0]$$

$$= \frac{1}{k_{\mathrm{B}}} \langle \dot{x}(t+\tau)\dot{x}(t)\rangle_0.$$

(4.59)

The sub-script 0 signifies a free Brownian particle, in which case $\ddot{q}_0^{\mathrm{s}}(0) = T/m$, but the analysis is in fact more general than this (cf section 4.6.2).

This result shows that in physical applications, the behavior of the variance of the increment in the random displacement is quadratic in the time interval for intervals of the order of molecular time scales, and is linear in the time interval for intervals longer than the relaxation time of the system.

If in a particular application one uses an infinitesimal time step, (because, for example, the function of interest $g(x(t))$ is rapidly varying), then the variance is quadratic in the time step, and the difference between the Itô and the Stratinovich calculus is $\mathcal{O}(n\Delta_t^2) \sim t\Delta_t$, which is negligible, and the choice of the stochastic calculus makes no difference. If on the other hand one uses a relatively large time step beyond the system relaxation time (because $\mid \Delta g(x(t\prime)) \mid \ll \mid g(x(t\prime)) \mid$ over the time step), then the variance would indeed be linear in the time step, as mathematicians invariably assume. However, since the whole theory requires that the stochastic influence be a small perturbation on the adiabatic evolution (otherwise the reservoir influence would have to be treated explicitly rather than statistically), the contribution of the stochastic forces to the change in a function over the time interval must be a small perturbation on the deterministic change, $\mid Dg\prime(x(t)) \mid \ll \mid \dot{x}(t)g(x(t)) \mid$. If this holds, then the difference between the two versions of the stochastic calculus is negligible. If it does not hold then one must use instead a time step smaller than the system relaxation time (because this reduces the stochastic influence, $\mathcal{O}(\Delta_t^2)$, more quickly than the adiabatic influence, $\mathcal{O}(\Delta_t)$).

One concludes from these arguments that the stochastic calculus of Itô and Stratinovich plays no rôle in the statistical mechanics of physical systems.

In some books and papers, the Fokker–Planck equation is derived using the stochastic calculus, and different versions result depending upon the choice. The differences depend upon the gradient of the diffusion constant with respect to the current trajectory variable, namely $\nabla \cdot D$ (Ermak and McCammon 1978, Gardiner 1983, Risken 1984, Tough *et al* 1986,

Keizer 1987, Allen and Tildesley 1987, and Mazo 2008). These competing versions of the Fokker–Planck equation only occur for the standard 'no memory' Langevin equation; no such differences occur for the generalized Langevin equation that has memory.

It should be understood that the divergence of the transport coefficient in that analysis is with respect to the actual trajectory variable, which means that it is only nonzero if $D = D(x(t))$ in the present case. However, since the transport coefficient is an average, as given for example by the time correlation function, or by Green–Kubo theory, in physical problems it is impossible for the transport coefficient to depend on the current point on the trajectory. (It may depend upon space and time, for example $D(\bar{x}(t), t)$, but not upon the system state x. Obviously, $\nabla \cdot D(\bar{x}(t), t) = 0$.) By the fluctuation–dissipation theorem, this independence holds equivalently for the variance of the stochastic forces. In a system without memory, the transport coefficient must also be independent of prior points on the trajectory. In the opinion of the present author, for systems without memory, $\nabla \cdot D = 0$, and there is no ambiguity in the Fokker–Planck equation or in the evolution of any physical variable.

There remains the case of non-Markov systems, which is to say systems with memory. For these the generalized Langevin equation is required,

$$\dot{x}(t) = \frac{1}{T} \int_0^t dt' \, \ddot{q}_0^{\mathrm{s}}(t - t') F(x(t'), t') + \tilde{R}(t),$$

(4.60)

where $F(x, t)$ is an externally applied force. This and variants are derived in section 4.6. Detailed analysis (Attard 2012, chapters 10 and 11) shows that the different versions of the stochastic calculus have no influence on the Fokker–Planck equation or on any function of the trajectory for systems with memory.

4.4 Generalized equipartition theorem

For an equilibrium system with kinetic energy $\mathcal{K}(\mathbf{v}) = mv^2/2$, and potential energy $U(\mathbf{r})$, the entropy is $S_{\mathrm{r}}(\mathbf{v}, \mathbf{r}) = -[\mathcal{K}(\mathbf{v}) + U(\mathbf{r})]/T$, and the probability is $\wp(\mathbf{v}, \mathbf{r}) = Z(T)^{-1} e^{S_{\mathrm{r}}(\mathbf{v}, \mathbf{r})/k_{\mathrm{B}}}$. This is for a Brownian particle with no internal entropy so that the total entropy is the same as the reservoir entropy. Because of the Gaussian form of the velocity distribution, the variance of the velocity is

$$\langle \mathbf{v}\mathbf{v} \rangle = \frac{k_{\mathrm{B}}T}{m} \mathrm{I}.$$

(4.61)

This is the equipartition theorem. It holds for any system with the entropy having the stated form.

For the case that there are N particles, so that $\mathbf{v}$ is a $3N$-dimensional vector, then the trace of this is extensive, and fluctuations are relatively negligible. In this case the theorem applies locally,

$$\langle \mathbf{v} \cdot \mathbf{v} \rangle \approx \mathbf{v} \cdot \mathbf{v} = \frac{3Nk_{\mathrm{B}}T}{m}.$$

(4.62)

Obviously one *could* choose sets of velocities where this fails badly. But if one chooses likely velocities, or if the velocities arise spontaneously in the system, then this is an increasingly good approximation as $N \to \infty$.

The equipartition theorem is a rather trivial consequence of the Gaussian nature of the probability. It is possible to obtain a powerful generalization of this theorem even when the

probability is non-Gaussian. The derivation is quite simple and is based on the fact that the averand on the left-hand side can be considered to be in essence the dyadic product of the velocity gradient of the entropy, and the right-hand side can be considered to be the dyadic second velocity gradient of the entropy.

To this end represent the phase space of the sub-system of interest by X, and the reservoir entropy by $S_{\mathrm{r}}(X, t)$. For greatest generality we consider a general non-equilibrium system dependent on time t. Explicit expressions for the non-equilibrium entropy are given in chapter 6 but will not be required here. The probability is as usual $\wp(X, t) = Z(t)^{-1} \mathrm{e}^{S_{\mathrm{r}}(X,t)/k_{\mathrm{B}}}$. (It will be shown in section 5.2 that phase space is the set of microstates, and that they have no internal entropy, $S_{\mathrm{total}}(X, t) = S_{\mathrm{r}}(X, t)$. When X represents a macrostate with internal entropy, the total entropy $S_{\mathrm{total}}(X, t) = S_{\mathrm{s}}(X, t) + S_{\mathrm{r}}(X, t)$ has to be used in place of the reservoir entropy.)

In view of the observation made in the penultimate paragraph above, the average of the dyadic product of the phase space gradient of the entropy is

$$
\begin{aligned}
k_{\mathrm{B}}^{-2} \langle [\nabla S_{\mathrm{r}}(X, t)][\nabla S_{\mathrm{r}}(X, t)] \rangle &= k_{\mathrm{B}}^{-2} \int \mathrm{d}X \, \wp(X, t)[\nabla S_{\mathrm{r}}(X, t)][\nabla S_{\mathrm{r}}(X, t)] \\
&= k_{\mathrm{B}}^{-1} \int \mathrm{d}X \, [\nabla \wp(X, t)][\nabla S_{\mathrm{r}}(X, t)] \\
&= -k_{\mathrm{B}}^{-1} \int \mathrm{d}X \, \wp(X, t) \nabla \nabla S_{\mathrm{r}}(X, t) \\
&= -k_{\mathrm{B}}^{-1} \langle \nabla \nabla S_{\mathrm{r}}(X, t) \rangle.
\end{aligned}
$$

(4.63)

The third equality follows after an integration by parts, the integrated portion vanishing because the probability vanishes on the boundaries of the integration region. For the case of position components, this gives the virial equation provided that the total potential is used (i.e. including any one-body contributions from the walls containing the sub-system). This is the first generalized equipartition theorem. For a single component, say x, it reads

$$
\left\langle \left(\frac{\partial S_{\mathrm{r}}(X, t)}{k_{\mathrm{B}} \partial x} \right)^2 \right\rangle + \left\langle \frac{\partial^2 S_{\mathrm{r}}(X, t)}{k_{\mathrm{B}} \partial x^2} \right\rangle = 0.
$$

(4.64)

Taking $x = v$ and the reservoir entropy as above, one sees that this reduces to the usual equipartition theorem.

The first generalized equipartition theorem is equivalent to

$$
\int \mathrm{d}X \, \nabla \nabla \wp(X, t) = 0.
$$

(4.65)

This can be seen directly to be true by performing a single integration and noting that $\oint \mathrm{d}\mathbf{n}_X \cdot \nabla \wp(X, t) = 0$, since the probability and its gradient can be taken to vanish on the boundary. This is a global theorem, but one suspects that it may be a reasonable approximation applied point by point. That is

$$
\begin{aligned}
\nabla \nabla \wp(X, t) &\equiv \left[k_{\mathrm{B}}^{-1} \nabla \nabla S_{\mathrm{r}}(X, t) + k_{\mathrm{B}}^{-2} \{ \nabla S_{\mathrm{r}}(X, t) \} \{ \nabla S_{\mathrm{r}}(X, t) \} \right] \wp(X, t) \\
&\approx 0.
\end{aligned}
$$

(4.66)

Of course there are values of X when the term in brackets has a large magnitude, but one expects such X to be rather improbable, $\wp(X,t) \approx 0$. Conversely, probable values of X correspond to the approximate cancelation of the two terms in the brackets. It is important to note that for this approximation to be true locally in phase space, both factors (the bracketed terms and the probability density) are required.

This local approximation can be made with even greater certainty by summing over components. For example, summing the diagonal elements of the dyadic matrix yields

$$\left[k_{\mathrm{B}}^{-1} \nabla^2 S_{\mathrm{r}}(X,t) + k_{\mathrm{B}}^{-2} \{\nabla S_{\mathrm{r}}(X,t)\} \cdot \{\nabla S_{\mathrm{r}}(X,t)\} \right] \wp(X,t) \approx 0.$$

(4.67)

The reason that one can have confidence in this result is that the bracketed term is extensive, which is to say it scales with the size of the system. It will be recalled that fluctuations in extensive variables are relatively negligible. Hence in the thermodynamic limit one can write this as

$$k_{\mathrm{B}}^{-1} \nabla^2 S_{\mathrm{r}}(X,t) + k_{\mathrm{B}}^{-2} \{\nabla S_{\mathrm{r}}(X,t)\} \cdot \{\nabla S_{\mathrm{r}}(X,t)\} = 0.$$

(4.68)

It is implicitly understood that the values of X in this arise spontaneously in the system.

In the case of N particles in three dimensions, for an equilibrium system with the kinetic energy given above, $S_{\mathrm{r}}(X,t) = -mv^2/2T - U(\mathbf{r})/T$, using the velocity gradient in the last expression gives

$$\frac{-3Nm}{k_{\mathrm{B}}T} + \frac{m^2}{k_{\mathrm{B}}^2 T^2} v^2 = 0, \quad \text{or} \quad v^2 = 3Nmk_{\mathrm{B}}T.$$

(4.69)

One often sees the equipartition function written in this form. Again the reason that one can expect it to apply locally and not just on average is that it is extensive. Provided that these are velocities that have arisen in the system, the fluctuations that would invalidate this result are unlikely to occur. Arguably it would be better to show the probability density explicitly here to ensure that unlikely fluctuations that violate the equipartition theorem have zero weight.

There is another result that can be called a generalized equipartition theorem. Consider the average of the dyadic,

$$\begin{aligned} \left\langle X \frac{\partial S_{\mathrm{r}}(X,t)}{\partial X} \right\rangle &= \frac{1}{Z} \int dX \; e^{S_{\mathrm{r}}(X,t)/k_{\mathrm{B}}} X \frac{\partial S_{\mathrm{r}}(X,t)}{\partial X} \\ &= \frac{k_{\mathrm{B}}}{Z} \int dX \; X \frac{\partial e^{S_{\mathrm{r}}(X,t)/k_{\mathrm{B}}}}{\partial X} \\ &= \frac{-k_{\mathrm{B}}}{Z} \int dX \; e^{S_{\mathrm{r}}(X,t)/k_{\mathrm{B}}} \frac{\partial X}{\partial X} \\ &= -k_{\mathrm{B}} \, \mathrm{I}. \end{aligned}$$

(4.70)

The third equality follows from an integration by parts. This may be called the second generalized equipartition theorem.

Choosing $S_{\mathrm{r}}(X,t) = -\mathcal{H}(\mathbf{q},\mathbf{p})/T$, for a single momentum component this gives the conventional equipartition theorem,

$$\left\langle p_{j\alpha}^2 \right\rangle_{N,V,T} = m k_B T.$$

More generally it gives

$$\left\langle p_{j\alpha} p_{i\gamma} \right\rangle_{N,V,T} = \delta_{j,i} \delta_{\alpha,\gamma} m k_B T,$$

$$\left\langle q_{j\alpha} \frac{\partial U(\mathbf{q})}{\partial q_{i\gamma}} \right\rangle_{N,V,T} = \delta_{j,i} \delta_{\alpha,\gamma} k_B T,$$

and

$$\left\langle p_{j\alpha} \frac{\partial U(\mathbf{q})}{\partial q_{i\gamma}} \right\rangle_{N,V,T} = \left\langle q_{j\alpha} p_{i\gamma} \right\rangle_{N,V,T} = 0.$$

This last result exemplifies the principle that in an equilibrium system there can be no instantaneous correlation between variables of different time parity.

Finally, it should be pointed out that in general one should replace the reservoir entropy in the two generalized theorems everywhere above by the total entropy. In most cases this makes no difference to the results, but for some choices of macrostate *X* the sub-system entropy is non-constant and its inclusion does make a difference.

4.5 Liouville's approximation

The stochastic dissipative equation in the pure parity case, was given above as equation (4.31),

$$X(t+\tau) = X(t) + \tau \dot{X}^{\mathrm{det}}(t) + \tilde{R}(t).$$

Although derived with a Brownian particle in mind, the quantity *X* can be interpreted quite generally. The deterministic velocity can be written as the sum of the adiabatic and the most likely reservoir-induced velocity, equation (4.39),

$$\dot{X}^{\mathrm{det}} = \dot{X}^0 + \dot{\bar{X}} = \dot{X}^0 + \frac{\hat{\tau}}{2} \Lambda \nabla S(X),$$

Here $S(X)$ is the total entropy, which for a Brownian particle and for phase space, is the same as the reservoir entropy. (For the mixed parity case, the transport matrix is instead $L(\hat{\tau})$, equation (3.135).) The stochastic contribution is Gaussian distributed, equation (4.32),

$$\wp(\tilde{R}) = \frac{1}{Z\prime} e^{-\Lambda^{-1}:\tilde{R}\tilde{R}/2k_B|\tau|}, \quad Z\prime \equiv \left\{ \mathrm{Det} \left[2\pi k_B \mid \tau \mid \Lambda \right] \right\}^{1/2}.$$

The time development of a probability distribution due to this stochastic dissipative equation is given by the Fokker–Planck equation, derived above as equation (4.36). This can be rearranged as

$$\frac{\partial \wp(X,t)}{\partial t} = -\nabla \cdot \dot{X}^{\text{det}} \wp(X,t) - \dot{X}^{\text{det}} \cdot \nabla \wp(X,t) + \frac{\hat{\tau} k_{\text{B}}}{2} \Lambda : \nabla\nabla \wp(X,t)$$

$$= -\nabla \cdot \left\{ \dot{X}^{\text{det}} \wp(X,t) \right\} + \frac{\hat{\tau} k_{\text{B}}}{2} \nabla \cdot \{\Lambda \nabla \wp(X,t)\}.$$

(4.7
8)

The final term can be written like this because the transport matrix Λ is independent of
X (cf the discussion of the stochastic calculus, section 4.3).

Written in this way, the right-hand side of the Fokker–Planck equation represents a
conservation law for the probability, $\partial \wp(X,t)/\partial t = -\nabla \cdot J_{\wp}(X,t)$, with the probability
flux being

$$J_{\wp}(X,t) = \dot{X}^{\text{det}} \wp(X,t) + \frac{\hat{\tau} k_{\text{B}}}{2} \Lambda \nabla \wp(X,t).$$

(4.7
9)

The first term on the right-hand side represents the flux of probability carried by the
deterministic motion, and the second term is that carried by the stochastic part of the
equations of motion.

Liouville's theorem for an open sub-system corresponds to the neglect of the stochastic
contribution in the equations of motion and in the Fokker–Planck equation. In this case the
total time derivative is

$$\frac{\mathrm{d}\wp(X,t)}{\mathrm{d}t} = \frac{\partial \wp(X,t)}{\partial t} + \dot{X}^{\text{det}}(X,t) \cdot \nabla \wp(X,t)$$

$$= -\left[\nabla \cdot \dot{X}^{\text{det}}(X,t)\right] \wp(X,t).$$

(4.8
0)

The first line is exact for deterministic equations of motion, and the second line follows
from the Fokker–Planck equation with the stochastic contribution neglected. This is called
Liouville's theorem, and it was critically referred to at the start of this chapter, equation
(4.2). This gives the probability flux as $J_{\wp}(X,t) = \dot{X}^{\text{det}}(X,t)\wp(X,t)$.

It would be better to call this Liouville's approximation. As has been mentioned several
times, when one projects the reservoir coordinates out of the problem, one must
necessarily have both dissipative (most likely) and stochastic (zero mean) forces in the
equations of motion. When a reservoir is present, it is not possible to have a dissipative
part without a fluctuation part. The only legitimate case with zero dissipative part and zero
fluctuation part is the adiabatic equations of motion, which are valid for an isolated sub-
system that does not interact with its environment or a reservoir. In this case, and only this
case, Liouville's theorem is an exact theorem and it reads

$$\frac{\mathrm{d}\wp(X,t)}{\mathrm{d}t} = -\left[\nabla \cdot \dot{X}^{0}(X,t)\right] \wp(X,t)$$

$$= -\dot{X}^{0}(X,t) \cdot \nabla \wp(X,t).$$

(4.8
1)

The second equality follows because adiabatic equations of motion are incompressible,
$\nabla \cdot \dot{X}^{0} = 0$, (assuming X represents a point in phase space).

At the time that Liouville derived his theorem the only equations of motion that were
used were Hamilton's. Over the years workers have developed ad hoc non-Hamiltonian
equations of motion. One such example is the isokinetic equations of motion, equation

(4.1), which were criticized in section 4.2.4. Unfortunately, some have naively assumed that Liouville's theorem holds also for these.

This raises three questions. Is it reasonable to neglect stochastic contributions to the evolution of the probability density? Is it a reasonable approximation to neglect stochastic contributions to the equations of motion themselves? And, is it reasonable to approximate the evolution of an arbitrary function as deterministic?

First, in equation (4.66) it was shown that the dyadic gradient of the probability density could be neglected not only globally, but also locally for extensive formulations such as the Laplacian. The relevance of this is that the stochastic contributions to the rate of change of the probability density, equation (4.78), involve the Laplacian of the probability density. If the local version of the generalized equipartition theorem, equation (4.66), is a reasonable approximation, then so also is Liouville's approximation for the evolution of the probability density. The double scaler product with the transport matrix makes the term extensive, and so local fluctuations from the generalized equipartition theorem are negligible in the thermodynamic limit. Arguably, Liouville's approximation gives the evolution of the probability density for the exact stochastic dissipative equations of motion, with the generalized equipartition theorem applied locally making the stochastic contribution negligible.

Second, deterministic, non-Hamiltonian equations of motion are invalid for a sub-system interacting with a reservoir or environment. Since the condition for the equilibrium probability density to be stationary is that the fluctuation–dissipation theorem be satisfied, and since the latter says that the variance of the fluctuations must be linearly proportional to the strength of the dissipation, it is not possible to have a dissipative deterministic non-Hamiltonian term in the equations of motion without having the corresponding stochastic term that is related by the fluctuation–dissipation theorem. We shall analyze non-equilibrium systems in chapter 6, but here one can already see that if purely deterministic non-Hamiltonian equations are forbidden for equilibrium systems, then they must necessarily be invalid for non-equilibrium systems. In general deterministic, non-Hamiltonian equations of motion are unphysical as they violate the equilibrium or non-equilibrium fluctuation– dissipation theorem.

Deterministic, non-Hamiltonian equations of motion suffer from two additional errors. Not including the stochastic contribution to the equations of motion neglects the randomness that is an inevitable consequence of the projection from the total system to the sub-system. Because of this, purely deterministic non-Hamiltonian equations of motion cannot represent a real physical system. In addition, lacking a physical basis, the ad hoc non-Hamiltonian dissipative term has not only the wrong magnitude, but often also the wrong functional form. It is an illusion to imagine that the fitting parameters that this freedom from the laws of nature allow in any way compensates for the uncertainty that it introduces into the results.

Third, the evolution of an arbitrary function is dependent upon the chosen equations of motion. Since there is no dissipation without fluctuation, it is not reasonable to approximate such evolution as deterministic.

4.6 Generalized Langevin equation

4.6.1 Perturbation theory

Consider a Brownian particle at $\mathbf{r}$, acted upon by a time-varying external force $\mathbf{F}(\mathbf{r}, t) = -\nabla U(\mathbf{r}, t)$. The Brownian particle is the sub-system and the solvent is the thermal reservoir with temperature T. The aim here is to account for solvent memory effects when the external potential varies rapidly. To this end a perturbation or linear response theory for the particle motion is developed.

All of the change in external energy over a specified trajectory due to the movement of the Brownian particle comes from the solvent, and this changes the reservoir entropy from what it would be for the same trajectory in the absence of the external potential,

$$S_{\mathrm{r}}(t \mid [\mathbf{r}], [U]) = S_{\mathrm{r},0}(t \mid [\mathbf{r}]) - \frac{1}{T} \int_0^t \mathrm{d}t\prime\, \dot{\mathbf{r}}(t\prime) \cdot \nabla U(\mathbf{r}(t\prime), t\prime).$$

(4.8 2)

The quantity $\dot{\mathbf{r}} \cdot \nabla U$ is the rate of change of external energy due to the particle motion, and, by energy conservation, this is equal and opposite to the rate of change of solvent energy additional to what would have occurred for the bare particle on the same trajectory.

The average velocity at time t in the presence of the external potential is

$$\langle \mathbf{r}(t) \rangle_U = \frac{\displaystyle\int \mathrm{d}[\mathbf{r}]\; \mathrm{e}^{S_{\mathrm{r}}(t|[\mathbf{r}],[U])/k_{\mathrm{B}}} \dot{\mathbf{r}}(t)}{\displaystyle\int \mathrm{d}[\mathbf{r}]\; \mathrm{e}^{S_{\mathrm{r}}(t\prime|[\mathbf{r}],[U])/k_{\mathrm{B}}}}$$

$$= \frac{\displaystyle\int \mathrm{d}[\mathbf{r}]\; \mathrm{e}^{S_{\mathrm{r},0}(t|[\mathbf{r}])/k_{\mathrm{B}}}\left[1 - \beta \int_0^t \mathrm{d}t\prime\, \dot{\mathbf{r}}(t\prime) \cdot \nabla U(\mathbf{r}(t\prime), t\prime)\right] \dot{\mathbf{r}}(t)}{\displaystyle\int \mathrm{d}[\mathbf{r}]\; \mathrm{e}^{S_{\mathrm{r},0}(t|[\mathbf{r}])/k_{\mathrm{B}}}\left[1 - \beta \int_0^t \mathrm{d}t\prime\, \dot{\mathbf{r}}(t\prime) \cdot \nabla U(\mathbf{r}(t\prime), t\prime)\right]}$$

$$= -\beta \int_0^t \mathrm{d}t\prime\, \langle \dot{\mathbf{r}}(t\prime) \cdot \nabla U(\mathbf{r}(t\prime), t\prime) \dot{\mathbf{r}}(t) \rangle_0$$

$$= -\beta \int_0^t \mathrm{d}t\prime\, \langle \dot{\mathbf{r}}(t) \dot{\mathbf{r}}(t\prime) \rangle_0 \cdot \nabla U(\mathbf{r}(t\prime), t\prime)$$

$$= \frac{1}{T} \int_0^t \mathrm{d}t\prime\, \ddot{q}_0(t - t\prime) \mathbf{F}(\mathbf{r}(t\prime), t\prime).$$

(4.8 3)

In the second equality the exponentials have been linearized with respect to the external potential. In the third equality the facts that the velocity in the bare system averages to zero, $\langle \dot{\mathbf{r}}(t\prime) \rangle_0 = 0$, and that it is uncorrelated with the current position, $\langle \dot{\mathbf{r}}(t\prime) \mathbf{r}(t\prime) \rangle_0 = 0$, have been used. In the fourth equality the force has been taken outside of the average because in the bare system $\mathbf{r}(t\prime)$ is uncorrelated with $\dot{\mathbf{r}}(t) \dot{\mathbf{r}}(t\prime)$, $t > t\prime$. In the final equality, the bare velocity autocorrelation function has been defined as in equation (4.59),

$$\ddot{q}_0(\tau) = k_{\mathrm{B}}^{-1} \langle \dot{\mathbf{r}}(t + \tau) \dot{\mathbf{r}}(t) \rangle_0.$$

(4.8 4)

This is a symmetric, even, matrix, $\ddot{q}_0(\tau) = \ddot{q}_0(-\tau) = \ddot{q}_0(\tau)^{\mathrm{T}}$.

The final result says that the velocity autocorrelation function of the bare Brownian particle is the response function that gives the current velocity due to the preceding values of the external force. This function contains the memory effects of the solvent. Because these are short-ranged, the result is independent of the lower limit of the integral.

In the event that the external potential is slowly varying, $\mathbf{F}(\mathbf{r}(t\prime), t\prime) \approx \mathbf{F}(\mathbf{r}(t), t)$, this becomes

$$\dot{\bar{\mathbf{r}}}(t) = \langle \dot{\mathbf{r}}(t) \rangle_U$$
$$\approx \frac{1}{T} \int_0^t \mathrm{d}t\prime \, \ddot{q}_0(t - t\prime) \mathbf{F}(\mathbf{r}(t), t)$$
$$= \beta \langle \dot{\mathbf{r}}(t)[\mathbf{r}(t) - \mathbf{r}(0)] \rangle_0 \mathbf{F}(\mathbf{r}(t), t)$$
$$= \beta D \mathbf{F}(\mathbf{r}(t), t).$$

(4.85)

The final equality uses equation (4.15) for the diffusion constant. This is known as the Smoluchowski equation and it is applicable in the large drag limit (Smoluchowski 1906). It is equivalent to the over damped case in which the acceleration is set to zero in the Langevin equation (4.19), so that the drag force is always equal and opposite to the applied force, $\dot{\bar{\mathbf{r}}}(t) = \gamma^{-1} \mathbf{F}(\bar{\mathbf{r}}(t), t) = \beta D \mathbf{F}(\bar{\mathbf{r}}(t), t)$.

One can add the random force to the full result to obtain the Langevin equation with memory,

$$\dot{\mathbf{r}}(t) = \mathbf{R}_v(t) + \frac{1}{T} \int_{-\infty}^t \mathrm{d}t\prime \, \ddot{q}_0(t - t\prime) \mathbf{F}(\mathbf{r}(t\prime), t\prime).$$

(4.86)

One can discretize this to generate the stochastic, dissipative trajectory of the Brownian particle driven by a rapidly varying external force. The variance of the random force (actually force divided by mass) can be obtained by setting the external force to zero, multiplying the respective sides of this by their value at $t = t\prime$, and taking the average,

$$\langle \mathbf{R}_v(t) \mathbf{R}_v(t\prime) \rangle_0 = \langle \dot{\mathbf{r}}(t) \dot{\mathbf{r}}(t\prime) \rangle_0 = k_B \ddot{q}_0(t - t\prime).$$

(4.87)

When memory effects are important, this form of the fluctuation–dissipation theorem tells how the random forces at different times are correlated.

The present perturbation theory has two advantages. First, an explicit formula is given for the memory function, namely it is the velocity autocorrelation function. Second, the difficult part of the calculation, the memory function, need be done for the bare system once only, and the same memory function can be used for arbitrary applied time-dependent forces.

4.6.2 Thermodynamic fluctuations with memory

We now analyze the effects of memory on the evolution of thermodynamic fluctuations. This has many similarities to the just treated case of a driven Brownian particle, with the main differences coming from the more general mixed parity system treated here. This gives rise to some terms that would otherwise be zero in the pure parity case of a Brownian particle.

What are here called thermodynamic variables can also be called phase functions or dynamical variables, and what is here called a perturbation approach can also be called linear response theory.

4.6.2.1 Bare system

Let x denote a vector of fluctuating thermodynamic variables arranged so that they vanish on average $\langle x \rangle = 0$. Each variable has pure time parity, but the vector is of mixed parity, $x(\mathbf{\Gamma}^\dagger) = \varepsilon x(\mathbf{\Gamma})$, where, as in section 3.3, the parity matrix is diagonal with elements $\varepsilon_{jk} = [x_j(\mathbf{\Gamma}^\dagger)/x_j(\mathbf{\Gamma})]\delta_{jk}$.

The velocity autocorrelation matrix is

$$\ddot{q}_0(\tau) = k_{\mathrm{B}}^{-1}\langle x(t+\tau)\ddot{x}(t)\rangle_0$$
$$= -k_{\mathrm{B}}^{-1}\langle \dot{x}(t+\tau)\dot{x}(t)\rangle_0.$$

(4.88)

(This assumes that x is real. For the case of Fourier transforms, this is defined with a complex conjugate.) From the parity rules one has $\ddot{q}_0(-\tau) = \varepsilon\ddot{q}_0(\tau)\varepsilon = \ddot{q}_0(\tau)^{\mathrm{T}}$. This is one difference from the pure parity case, equations (4.59) and (4.84).

The position–velocity correlation matrix is

$$\dot{q}_0(\tau) = k_{\mathrm{B}}^{-1}\langle x(t+\tau)\dot{x}(t)\rangle_0$$
$$= -k_{\mathrm{B}}^{-1}\langle \dot{x}(t+\tau)x(t)\rangle_0,$$

(4.89)

with $\dot{q}_0(-\tau) = -\varepsilon\dot{q}_0(\tau)\varepsilon = -\dot{q}_0(\tau)^{\mathrm{T}}$. Note that $\dot{q}_0(\tau) \to 0$, $\tau \to \infty$. At $\tau = 0$ this is an antisymmetric matrix, $\dot{q}_0(0) = -\dot{q}_0(0)^{\mathrm{T}} \neq 0$. This is in contrast with Brownian motion above, where the pure parity position–velocity correlation function vanishes at $\tau = 0$.

The position autocorrelation matrix is

$$q_0(\tau) = -k_{\mathrm{B}}^{-1}\langle x(t+\tau)x(t)\rangle_0,$$

(4.90)

with $q_0(\tau) \to 0$, $\tau \to \infty$. This differs from the symmetrized position autocorrelation function defined for a free Brownian particle, equation (4.57), because the present fluctuations are localized about the origin, unlike the free Brownian particle. The parity rule is $q_0(-\tau) = \varepsilon q_0(\tau)\varepsilon = q_0(\tau)^{\mathrm{T}}$. Note the negative sign in this definition of the position autocorrelation matrix, which is implied by the choice of a positive sign in the definition of the velocity autocorrelation matrix. With this convention, it is generally the case that the diagonal elements of the autocorrelation matrix are global maxima for the velocity and global minima for the position at $\tau = 0$. These definitions of the time correlation matrices may be confirmed by differentiation.

The preceding observation about global extrema is based on the generic properties of the time correlation function. In probability theory, the covariance inequality, which is a form of the Cauchy–Schwarz inequality, states that in the present case of real variables with zero mean,

$$\langle AB\rangle_0^2 \leqslant \langle A^2\rangle_0 \langle B^2\rangle_0.$$

(4.91)

In the present context, one can evaluate A at $t + \tau$ and B at t, and each can be any of the components of x, $\dot{x}$, or $\ddot{x}$. For example, taking $A = x_j(t+\tau)$ and $B = x_k(t)$, this says

$$\{q_0(\tau)\}_{jk}^2 \leqslant \{q_0(0)\}_{jj}\{q_0(0)\}_{kk}.$$

(4.92)

Summing over both indices, this is $\mathrm{Tr}\,[q_0(\tau)q_0(-\tau)] \leqslant [\mathrm{Tr}\,q_0(0)]^2$. Taking $j = k$, this is $\{q_0(\tau)\}_{jj}^2 \leqslant \{q_0(0)\}_{jj}^2$, which shows that the magnitude of a diagonal element of the time correlation function decays from its magnitude at $\tau = 0$.

Consider a trajectory $x[t]$. This can be discretized into n nodes of spacing τ, and described by a vector consisting of the final position and the velocities leading up to that position, $\mathbf{X}^{(n+1)} = \{x(t), \dot{x}(t), \dot{x}(t-\tau), \ldots, \dot{x}(t-(n-1)\tau)\} = \{x(t), \dot{\mathbf{x}}^{(n)}\}$. The time step may be positive or negative. Each entry here is itself a vector of thermodynamic

variables. Since we are dealing with fluctuations, we take the trajectory entropy to be a quadratic form,

$$S_0(\mathbf{X}^{(n+1)}) = \frac{1}{2} S_0^{(n+1,n+1)} : \mathbf{X}^{(n+1)} \mathbf{X}^{(n+1)}$$

$$= \frac{1}{2} S_{0;xx} : x^2 + S_{0;\dot{x}x}^{(n)} : \dot{\mathbf{x}}^{(n)} x + \frac{1}{2} S_{0;\dot{x}\dot{x}}^{(nn)} : \dot{\mathbf{x}}^{(n)} \dot{\mathbf{x}}^{(n)},$$

(4.93)

where $x = x(t)$ and the matrices have an implicit dependence on the time step τ. The fluctuation matrix $S_0^{(n+1,n+1)}$ has been broken into four sub-matrices in an obvious notation.

The trajectory correlation matrix is

$$Q_0^{(n+1,n+1)} \equiv k_B^{-1} \left\langle \mathbf{X}^{(n+1)} \mathbf{X}^{(n+1)} \right\rangle_0.$$

(4.94)

This can also be decomposed into sub-matrices,

$$Q_{0;xx} \equiv k_B^{-1} \langle x(t)x(t) \rangle_0 = -q_0(0),$$

(4.95)

$$Q_{0;\dot{x}\dot{x}}^{(nn)} \equiv k_B^{-1} \left\langle \dot{\mathbf{x}}^{(n)} \dot{\mathbf{x}}^{(n)} \right\rangle_0 = \ddot{Q}_0^{(nn)},$$

(4.96)

and

$$Q_{0;x\dot{x}}^{(n)} \equiv k_B^{-1} \left\langle x(t) \dot{\mathbf{x}}^{(n)} \right\rangle_0,$$

(4.97)

with $Q_{0;\dot{x}x}^{(n)} = [Q_{0;x\dot{x}}^{(n)}]^{\mathrm{T}}$.

The elements of the velocity autocorrelation matrix are

$$\ddot{Q}_{0;jk}^{(nn)} = k_B^{-1} \langle \dot{x}(t_j) \dot{x}(t_k) \rangle_0 = \ddot{q}_0(t_j - t_k).$$

(4.98)

In the present mixed parity case $\ddot{Q}_0^{(nn)}$ is symmetric, $\ddot{Q}_{0;k\alpha,j\gamma}^{(nn)} = \ddot{Q}_{0;j\gamma,k\alpha}^{(nn)}$, where Roman letters range over the time nodes and Greek letters range over the thermodynamic components. Equivalently, $\ddot{q}_0(t) = \ddot{q}_0(-t)^{\mathrm{T}}$ or $\ddot{q}_{0;\alpha\gamma}(t) = \ddot{q}_{0;\gamma\alpha}(-t)$.

The position–velocity time correlation n-vector has elements

$$Q_{0;x\dot{x};k}^{(n)} \equiv k_B^{-1} \langle x(t) \dot{x}(t_k) \rangle_0 = \dot{q}_0(t - t_k).$$

(4.99)

Each of these elements is actually a matrix of the thermodynamic variables, $\left\{ Q_{0;x\dot{x};k}^{(n)} \right\}_{\alpha\beta} = \left\{ Q_{0;\dot{x}x;k}^{(n)} \right\}_{\beta\alpha} = \dot{q}_{0;\alpha\beta}(t - t_k) = -\dot{q}_{0;\beta\alpha}(t_k - t)$. Recall that $t_1 \equiv t$.

As usual, the fluctuation matrix is the negative inverse of the correlation matrix $Q_0^{(n+1,n+1)} S_0^{(n+1,n+1)} = -\mathbf{I}^{(n+1,n+1)}$. It is straightforward to show that the fluctuation sub-matrices are given by

$$S_{0;xx} = -\left[1 - [Q_{0;xx}]^{-1}Q_{0;\dot{x}\dot{x}}^{(n)}[Q_{0;\dot{x}\dot{x}}^{(nn)}]^{-1}[Q_{0;\dot{x}x}^{(n)}]^{\mathrm{T}}\right]^{-1}[Q_{0;xx}]^{-1},$$

(4.100)

$$S_{0;\dot{x}\dot{x}}^{(nn)} = -\left[\mathbf{I}^{(nn)} - [Q_{0;\dot{x}\dot{x}}^{(nn)}]^{-1}Q_{0;\dot{x}x}^{(n)}[Q_{0;xx}]^{-1}Q_{0;x\dot{x}}^{(n)}\right]^{-1}[Q_{0;\dot{x}\dot{x}}^{(nn)}]^{-1},$$

(4.101)

and

$$S_{0;\dot{x}x}^{(n)} = \left[\mathbf{I}^{(nn)} - [Q_{0;\dot{x}\dot{x}}^{(nn)}]^{-1}Q_{0;\dot{x}x}^{(n)}[Q_{0;xx}]^{-1}Q_{0;x\dot{x}}^{(n)}\right]^{-1}[Q_{0;\dot{x}\dot{x}}^{(nn)}]^{-1}Q_{0;\dot{x}x}^{(n)}[Q_{0;xx}]^{-1}.$$

(4.102)

The last two give

$$-[S_{0;\dot{x}\dot{x}}^{(nn)}]^{-1}S_{0;\dot{x}x}^{(n)} = Q_{0;\dot{x}x}^{(n)}[Q_{0;xx}]^{-1}.$$

(4.103)

For the terminal node this is $Q_{0;\dot{x}x;1}^{(n)}[Q_{0;xx}]^{-1} = -\dot{q}_0(0)^{\mathrm{T}}q_0(0)^{-1}$, which will be used below.

The inverse of the velocity fluctuation matrix will turn out to be the memory matrix,

$$M^{(nn)} \equiv [S_{0;\dot{x}\dot{x}}^{(nn)}]^{-1} = -Q_{0;\dot{x}\dot{x}}^{(nn)} + Q_{0;\dot{x}x}^{(n)}[Q_{0;xx}]^{-1}Q_{0;x\dot{x}}^{(n)}.$$

(4.104)

The first term in essence is the unconditional contribution to the entropy from the direct correlation between the velocities at a pair of nodes. The second term is the indirect contribution conditional on the fixed terminal value of the fluctuation $x(t)$. For the case of a free Brownian particle, $Q_{0;xx} \to \infty$, leaving $M^{(nn)} = -Q_{0;\dot{x}\dot{x}}^{(nn)}$.

The components of the memory matrix are

$$M_{jk}^{(nn)} = -\ddot{q}_0(t_j - t_k) - \dot{q}_0(t - t_j)^{\mathrm{T}}q_0(0)^{-1}\dot{q}_0(t - t_k).$$

(4.105)

Recall that $t_1 \equiv t$. The memory matrix is in total symmetric, $M_{j\alpha,k\gamma}^{(nn)} = M_{k\gamma,j\alpha}^{(nn)}$.

The terminal case, $j = 1$, will be required below, and so with $t_1 \equiv t$ one can define the memory vector with components

$$M_k^{(n)} \equiv M_{1k}^{(nn)} = -\ddot{q}_0(t - t_k) - \dot{q}_0(0)^{\mathrm{T}}q_0(0)^{-1}\dot{q}_0(t - t_k).$$

(4.106)

In the continuum limit this is the memory function $M(t - t\prime) \equiv -\ddot{q}_0(t - t\prime) - \dot{q}_0(0)^{\mathrm{T}}q_0(0)^{-1}\dot{q}_0(t - t\prime)$. The memory function is a matrix in the thermodynamic variables with symmetry $M_{\alpha\gamma}(t - t\prime) = M_{\gamma\alpha}(t\prime - t)$. Because the correlation functions are short-ranged, so is the memory function, $M(t) \to 0$, $|t| \to \infty$. For the pure parity case, $\dot{q}_0(0) = 0$. Hence the pure parity memory function is just the velocity autocorrelation function alone.

4.6.2.2 Regression of fluctuation
The derivative of the trajectory entropy with respect to the velocities is

$$\frac{\partial S_0(\mathbf{X}^{(n+1)})}{\partial \dot{\mathbf{x}}^{(n)}} = S_{0;xx}^{(n)} x(t) + S_{0;xx}^{(nn)} \dot{\mathbf{x}}^{(n)}.$$

Setting this to zero gives the most likely trajectory conditional on the terminal position $x_1 = x(t_1)$ as

$$\begin{aligned}
\overline{\dot{\mathbf{x}}}^{(n)} &= -\left[S_{0;xx}^{(nn)}\right]^{-1} S_{0;xx}^{(n)\mathrm{T}} x_1 \\
&= Q_{0;xx}^{(n)} [Q_{0;xx}]^{-1} x_1 \\
&= \left\langle \dot{\mathbf{x}}^{(n)} x_1 \right\rangle_0 \langle x_1 x_1 \rangle_0^{-1} x_1.
\end{aligned}$$

The second equality follows from equation (4.103).

This must be the same as that given by the definitions of the time correlation functions, which may be confirmed by multiplying the nth element of this on the right by x_1 and taking the average. One obtains for the left-hand side

$$\mathrm{LHS} = \left\langle \overline{\dot{x}}_n x_1 \right\rangle_0 = \langle \dot{x}_n x_1 \rangle_0 = k_\mathrm{B} \dot{q}_0(t_1 - t_n)^\mathrm{T},$$

and for the right-hand side

$$\mathrm{RHS} = -\dot{q}_0(t_1 - t_n)^\mathrm{T} q_0(0)^{-1} \langle x_1 x_1 \rangle_0 = k_\mathrm{B} \dot{q}_0(t_1 - t_n)^\mathrm{T}.$$

These are equal, which confirms the validity of the approach.

The element $\overline{\dot{x}}_n$, is the most likely velocity after an interval $t_n - t_1$ given that the system was (will be) at x_1 at t_1. This is explicitly

$$\begin{aligned}
\overline{\dot{x}}(t_n \mid x_1, t_1) &= \left\langle \dot{x}(t_n) x_1 \right\rangle_0 \langle x_1 x_1 \rangle_0^{-1} x_1 \\
&= \dot{q}_0(t_n - t_1) \, S x_1.
\end{aligned}$$

This uses the fluctuation form for the equilibrium entropy,

$$S(x) = \frac{1}{2} S'' \colon xx, \quad S'' = -k_\mathrm{B} \langle xx \rangle_0^{-1} = q_0(0)^{-1}.$$

The result for the most likely velocity is really just a form of Onsager's regression hypothesis, with $F = \partial S/\partial x = S'' x$ being the thermodynamic force, and $\dot{q}_0(t_n - t_1)$ being the mixed parity transport matrix for the given time interval.

This result for the regression of a fluctuation ignores the history of the system beyond t_1 (prior to t_1 if $t_n > t_1$; after t_1 if $t_1 > t_n$). This is equivalent to assuming that the system is dynamically disordered at t_1. This remark is of relevance in interpreting the perturbation theory that is now developed.

4.6.2.3 Time-varying perturbation force

Now add to the bare system an external time-varying force that acts on the fluctuation, $\mathbf{F}(x,t) = -\nabla_x U(x,t)$. For example x might include the electric or magnetic polarization of the system, and the external force would contain the corresponding applied field. We

shall continue to regard *x* as possibly a mixed parity vector, even if the external forces couple directly to only some of the variables.

As in section 4.6.1, the rate of change of entropy due to the motion of the thermodynamic variables is

$$\dot{S}(x,t) = \frac{-1}{T}\,\dot{x} \cdot \nabla_x U(x,t) = \frac{1}{T}\dot{x} \cdot \mathbf{F}(x,t).$$

(4.113)

The quantity $\dot{x} \cdot \nabla_x U = -\dot{x} \cdot \mathbf{F}(x,t)$ is the rate of change of external energy due to the particle motion. This energy comes from the solvent reservoir itself, which decreases the entropy of the reservoir accordingly; *T* is the temperature of the reservoir.

As in equation (4.82), for a given trajectory the reservoir entropy is just the bare trajectory entropy plus the change in entropy over the trajectory due to the motion in the external field,

$$S(\mathbf{X}^{(n+1)}, [U]) = S_0(\mathbf{X}^{(n+1)}) - \frac{1}{T}\int_0^t dt\prime\,\dot{x}(t\prime) \cdot \nabla_x U(x(t\prime), t\prime)$$

$$= \frac{1}{2}S_0^{(n+1,n+1)} : \mathbf{X}^{(n+1)}\mathbf{X}^{(n+1)} + \frac{\tau}{T}\dot{\mathbf{x}}^{(n)} \cdot \mathbf{F}^{(n)}.$$

(4.114)

Differentiating this and using equation (4.107) gives

$$\frac{\partial S(\mathbf{X}^{(n+1)}, [U])}{\partial \dot{\mathbf{x}}^{(n)}} = S_{0;\dot{x}x}^{(n)} x(t) + S_{0;\dot{x}\dot{x}}^{(nn)}\dot{\mathbf{x}}^{(n)} + \frac{\tau}{T}\mathbf{F}^{(n)} + \mathcal{O}(F^2).$$

(4.115)

This neglects a contribution from the dependence of the force at time *t*⁣′ on the position $x(t\prime)$, which depends on the velocities $\dot{x}(t\prime\prime)$, $t\prime\prime \in [t, t\prime]$. This neglected term contains the product of $\dot{x}(t\prime)$ and the derivative of the force, $\nabla_{x(t\prime)}\mathbf{F}(x(t\prime), t\prime)\partial x(t\prime)/\partial \dot{x}(t\prime\prime)$. Since the most likely value of the velocity is a linear function of the force and of the position $x(t)$, and since the most likely value of the latter is also linearly proportional to the force, the neglected term is quadratic in the force.

Setting the derivative to zero gives the most likely velocities and hence the trajectory as

$$\overline{\dot{\mathbf{x}}}^{(n)} = -[S_{0;\dot{x}\dot{x}}^{(nn)}]^{-1}S_{0;\dot{x}x}^{(n)}x_1 - \frac{\tau}{T}[S_{0;\dot{x}\dot{x}}^{(n)}]^{-1}\mathbf{F}^{(n)}$$

$$= Q_{0;\dot{x}x}^{(n)}[Q_{0;xx}]^{-1}x_1 - \frac{\tau}{T}M^{(n)}\mathbf{F}^{(n)}.$$

(4.116)

This uses equation (4.103) and the memory matrix given in equation (4.104), $M^{(n)} \equiv [S_{0;\dot{x}\dot{x}}^{(n)}]^{-1}$. This confirms the above assertion that the most likely value of the velocity is a linear function of the force and of the position $x(t)$. This is essentially the same as the result for Brownian motion, section 4.6.1, with differences arising from terms that are nonzero in the present mixed parity case.

Evaluating this at the terminal node, $\overline{x}(t_1)$, and writing $t_1 \equiv t$, and transforming back to the continuum this is

$$\overline{\dot{x}}(t) = -\dot{q}_0(0)^{\mathrm{T}} q_0(0)^{-1} x(t) - \frac{1}{T} \int_0^t \mathrm{d}t\prime \; M(t - t\prime) \mathbf{F}(x(t\prime), t\prime).$$

The first term on the right-hand side is just the result for the (instantaneous) regression of a fluctuation in the absence of an external field, equation (4.111), $\overline{\dot{x}}(t_n \mid x_1, t_1) = \dot{q}_0(t_n - t_1)\, S x_1$. The memory function is given by equation (4.106),

$$M(t - t\prime) \equiv -\ddot{q}_0(t - t\prime) - \dot{q}_0(0)^{\mathrm{T}} q_0(0)^{-1}\dot{q}_0(t - t\prime)$$
$$= -k_{\mathrm{B}}^{-1}\langle \dot{x}(t)\dot{x}(t\prime)\rangle_0 + k_{\mathrm{B}}^{-1}\langle \dot{x}(t)x(t)\rangle_0\langle x(t)x(t)\rangle_0^{-1}\langle x(t)\dot{x}(t\prime)\rangle_0.$$

The memory function is short-ranged and so the integral can be truncated after some fixed interval beyond the relaxation time.

One can perform similar analysis to obtain the most likely velocity at t, given the position at some prior time $t\prime$ and the velocities leading up to t, $\dot{\mathbf{x}}$. In this case the memory function changes to reflect this different condition (Attard 2012, equation (10.161)).

4.6.2.4 Internal thermodynamic force

As mentioned, the way in which the external force enters the above equations is that it gives the rate of entropy production on the trajectory, $\dot{S}(x, t) = \dot{x} \cdot \mathbf{F}(x, t)/T$. But as pointed out by Onsager (1931) in his regression hypothesis, the system behaves the same whether the current state is bought about by a previously applied external force or by a spontaneous fluctuation of the system itself. In the latter case the internal thermodynamic force is the gradient of the entropy, and the rate of change of system entropy is just $\dot{S}(x) = \dot{x} \cdot \nabla_x S(x)$. Hence for a 'bare' system one can write

$$\frac{\mathbf{F}(x, t)}{T} \Rightarrow \nabla_x S(x(t)) = S\prime\prime x(t),$$

the second equality holding in the linear regime. With this replacement the most likely velocity in the bare system without an applied force is

$$\overline{\dot{x}}(t) = -\dot{q}_0(0)^{\mathrm{T}} q_0(0)^{-1} x(t) - \int_0^t \mathrm{d}t\prime \; M(t - t\prime)\nabla_x S(x(t\prime))$$
$$= \dot{q}_0(0)q_0(0)^{-1} x(t)$$
$$+ \int_0^t \mathrm{d}t\prime \; \left\{ \ddot{q}_0(t - t\prime) - \dot{q}_0(0)q_0(0)^{-1}\dot{q}_0(t - t\prime) \right\} S\prime\prime x(t\prime).$$

This result gives $\overline{\dot{x}}(t \mid \mathbf{X})$, the most likely velocity conditional on the specified preceding trajectory $\mathbf{X} = \{x, \dot{\mathbf{x}}\}$, leading up to the current position $x = x(t)$. The result for the regression of a fluctuation, equation (4.111), namely $\overline{\dot{x}}(t\prime \mid x, t) = \dot{q}_0(t\prime - t)\, S\prime\prime x$, gives the most likely current velocity given the position at a different time and not specifying the trajectory leading up to that position. The present result is useful when memory effects are non-negligible and one has knowledge of how the system reached its current state.

One can form dyadic matrices by multiplying this on the right by $k_{\mathrm{B}}^{-1} x(0)$. Taking the average gives

$$\dot{q}_0(t) \equiv -k_{\mathrm{B}}^{-1}\langle \dot{x}(t)x(0)\rangle_0$$
$$= -k_{\mathrm{B}}^{-1}\langle \overline{\dot{x}}(t)x(0)\rangle_0$$
$$= -k_{\mathrm{B}}^{-1}\dot{q}_0(0)q_0(0)^{-1}\langle x(t)x(0)\rangle_0$$
$$\quad - k_{\mathrm{B}}^{-1}\int_0^t dt\prime \left\{\ddot{q}_0(t-t\prime) - \dot{q}_0(0)q_0(0)^{-1}\dot{q}_0(t-t\prime)\right\}S\prime\prime\langle x(t\prime)x(0)\rangle_0$$
$$= \dot{q}_0(0)q_0(0)^{-1}q_0(t)$$
$$\quad + \int_0^t dt\prime \left\{\ddot{q}_0(t-t\prime) - \dot{q}_0(0)q_0(0)^{-1}\dot{q}_0(t-t\prime)\right\}q_0(0)^{-1}q_0(t\prime).$$

This is, presumably, an identity.

4.6.2.5 Generalized Langevin equation

Adding a stochastic zero mean term to the above result for the most likely velocity due to a perturbing time-dependent external force gives the generalized Langevin equation

$$\dot{x}(t) = R(t) + \dot{q}_0(0)q_0(0)^{-1}x(t) - \frac{1}{T}\int_0^t dt\prime \, M(t-t\prime)\mathbf{F}(x(t\prime),t\prime).$$

The covariance of the random forces is given by the memory matrix

$$\langle R(t)R(0)\rangle_0 = -k_{\mathrm{B}}M(t),$$

where the memory matrix is given by equation (4.118).

This is readily checked for the discretized trajectory, in which case one has

$$\langle \mathbf{R}^{(n)}\mathbf{R}^{(n)}\rangle = \left\langle \left[\dot{\mathbf{x}}^{(n)} - \overline{\dot{\mathbf{x}}}^{(n)}\right]\left[\dot{\mathbf{x}}^{(n)} - \overline{\dot{\mathbf{x}}}^{(n)}\right]\right\rangle$$
$$= \left\langle \dot{\mathbf{x}}^{(n)}\,\dot{\mathbf{x}}^{(n)}\right\rangle_0$$
$$= -k_{\mathrm{B}}[S_{0;\dot{x}\dot{x}}^{(nn)}]^{-1}$$
$$= -k_{\mathrm{B}}M^{(nn)}.$$

The covariance of the fluctuation in velocity, the first equality, can be taken to be the same as in the bare system, the second equality. The fourth equality is the definition of the memory function as the velocity autocorrelation matrix, equation (4.104). The probability distribution of the last n random forces is just the corresponding multivariate Gaussian.

The memory matrix satisfies $M(-t) = M(t)^{\mathrm{T}}$. The relationship between the covariance of the random 'force' and the memory function is the same as for Brownian motion, equation (4.86). That the memory function gives both the covariance of the random forces and the dissipative force is the fluctuation–dissipation theorem for a system with memory. The generalized Langevin equation in this form has the advantage that the memory function is constructed from the bare time correlation functions and it can be used for arbitrary applied forces.

For the case of the internal thermodynamic force for a fluctuation in the bare system, $\mathbf{F}(x,t)/T \Rightarrow \nabla_x S(x(t)) = S\prime\prime x(t)$, the generalized Langevin equation is

$$\dot{x}(t) = R(t) + \dot{q}_0(0)q_0(0)^{-1}x(t) - \int_0^t \mathrm{d}t\prime\, M(t - t\prime)S\prime\prime x(t\prime).$$

(4.1
25)

The memory function, equation (4.118), and the covariance of the random force remain unchanged. It can be shown that these results are formally the same as the generalized Langevin equation given by the projector operator formalism of Zwanzig (1961) and Mori (1965) (see Attard 2012 section 10.7). The present approach gives the memory function explicitly in terms of the time correlation functions.

Summary

- Brownian motion is the archetypical stochastic process in which hidden forces contribute as well as the extant adiabatic forces. These hidden forces arise from projection of reservoir interactions, they must be treated statistically, and they obey certain thermodynamic and time symmetry rules.
- The Langevin equation and the fluctuation–dissipation theorem are consequences of the second entropy for transitions. The dissipative force drives the system up the reservoir entropy gradient, and the fluctuation force randomizes the transition, with a bias toward more probable sub-system macrostates. These compete with each other, reaching a balance in the equilibrium macrostate.
- The Fokker–Planck equation gives the evolution of the probability density under stochastic, dissipative equations of motion. The equilibrium probability density is stationary under the second entropy stochastic, dissipative equations of motion.
- The equation due to Liouville for the time derivative of the probability density is an exact theorem for adiabatic equations of motion. It is an approximation for deterministic, dissipative equations of motion that is valid when the generalized equipartition theorem can be applied locally.
- The first generalized equipartition theorem relates the dyadic of the gradient of the reservoir entropy to the dyadic gradient of the reservoir entropy. These two are equal on average, and they may be approximately equal locally on likely points in phase space. The second generalized equipartition theorem says that the average of dyad formed from the state coordinate and the coordinate gradient of the reservoir entropy is proportional to the identity matrix.
- The Langevin equation can be generalized to replace the friction coefficient by a memory function for a Brownian particle or thermodynamic variable that is driven by a time-dependent external potential. The formalism holds also for thermodynamic fluctuations.

References

Allen M P and Tildesley D J 1987 *Computer Simulations of Liquids* (Oxford: Oxford University Press)

Attard P 2012 *Non-Equilibrium Thermodynamics and Statistical Mechanics: Foundations and Applications* (Oxford: Oxford University Press)

Bachelier L 1900 Théorie de la Spéculation *Annales Scientifiques de L'École Normale Supérieure* 19 21

Bachelier L 1914 *Le Jeu, la Chance et le Hasard* (Paris: Ernest Flammarion)

Boltzmann L 1871 Über das wärmegleichgewicht zwischen mehratomigen gasmolekule *Wien. Ber.* **63** 397, 679, 712

Brown R 1828 A brief account of microscopical observations made in the months of June, July, and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies *Phil. Mag.* **4** 161 (series 2)

Einstein A 1905 Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen *Ann. Phys.* **17** 549

Ermak D L and McCammon J A 1978 Brownian dynamics with hydrodynamic interactions *J. Chem. Phys.* **69** 1352

Fokker A 1914 Die mittlere energie rotierender elektrischer dipole im strahlungsfeld *Ann. Phys.* **43** 810

Gardiner C W 1983 *Handbook of Stochastic Methods* (New York: Springer)

Haw M D 2002 Colloidal suspensions, Brownian motion, molecular reality: a short history *J. Phys.: Condens. Matter* **14** 7769

Keizer J 1987 *Statistical Thermodynamics of Non-Equilibrium Processes* (New York: Springer)

Langevin P 1908 Sur la théorie du mouvement Brownien *C. R. Acad. Sci. Paris* **146** 530

Mazo R M 2008 *Brownian Motion: Fluctuations, Dynamics, and Applications* (Oxford: Oxford University Press)

Mori H 1965 Transport, collective motion, and Brownian motion *Progr. Theor. Phys.* **33** 423 and 34 399

Münster A 1969 *Statistical Thermodynamics* **vol 1** (Berlin: Springer)

Onsager L 1931 Reciprocal relations in irreversible processes. I *Phys. Rev.* **37** 405 and 38 2265

Pathria R K 1972 *Statistical Mechanics* (Oxford: Pergamon)

Planck M 1916 Über einen satz der statistischen dynamik und seine erweiterung in der quanten theorie *Sitz. Ber. Preuss. Akad. Wiss.* **324**

Pope A 1734 *Essay on Man*

Risken H 1984 *The Fokker–Planck equation* (Berlin: Springer)

Silver B L 1998 *The Ascent of Science* (Oxford: Oxford University Press)

Tough R J A, Pusey P N, Lekerkerker H N W and van den Broeck C 1986 Stochastic descriptions of the dynamics of interacting Brownian particles *Mol. Phys.* **59** 595

von Smoluchowski M 1906 Zur kinetischen theorie der Brownschen molekularbewegung und der suspensionen *Ann. Phys.* **21** 756

Weinert F 2005 *The Scientist as Philosopher: Philosophical Consequences of Great Scientific Discoveries* (Berlin: Springer)

Zwanzig R 1961 *Lectures in Theoretical Physics* ed Britton W E (ed) , Downs B W (ed) and Downs J (ed) (New York: Wiley Interscience)

Entropy Beyond the Second Law (Second Edition)

Thermodynamics and statistical mechanics for equilibrium, non-equilibrium, classical, and quantum systems

Phil Attard

Chapter 5

Equilibrium statistical mechanics and entropy in classical phase space

'Indeed it is clear that any individual uniform distribution, which might arise after a certain time from some particular initial state, is just as improbable as an individual non-uniform distribution.... It is only because there are many more uniform distributions than non-uniform ones that the distribution of states will become uniform in the course of time.'

(Boltzmann 1877)

'A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it'

(Planck 1949)

The main aim of this chapter is to formulate classical equilibrium statistical mechanics from first principles. The approach is to derive the Maxwell–Boltzmann distribution from Hamilton's classical equations of motion in phase space. The stochastic, dissipative equations of motion for phase space for an open sub-system are also derived. Two approximations for the information entropy are tested against exact results for the two-dimensional Ising model.

The Boltzmann distribution is often invoked for an equilibrium system,

$$\wp_\alpha = \frac{1}{Z(T)} e^{-U_\alpha/k_B T}.$$

(5.1)

Here $\wp_\alpha$ is the probability of the state α, U_α is its potential energy, and T is the temperature. Despite its use being common in papers and textbooks, as a general result there is a quite serious problem with this that is discussed in section 5.3.3. The point of deriving the Maxwell–Boltzmann distribution from first principles is to explain why equilibrium statistical mechanics is formulated the way it is, and why the Boltzmann distribution in this form cannot be blindly applied to the states of an equilibrium system

5.1 Phase space and Hamilton's equations of motion

The preceding chapter dealt with the properties of Brownian motion, which is a generic model of the stochastic behavior of a macroscopic coordinate due to the hidden deterministic behavior of microscopic coordinates. Most of the results carry over directly to the stochastic behavior of microscopic coordinates of a sub-system due to the hidden deterministic behavior of a reservoir with which it interacts. Here and in the remainder of this chapter the focus is on the formulation of classical equilibrium statistical mechanics in terms of classical phase space, eventually leading to the Maxwell–Boltzmann probability distribution and to the stochastic, dissipative equation of motion in phase space.

5.1.1 Classical phase space

In this section points in phase space are identified as the microstates of a classical system. At the simplest level this is the space of positions and linear momenta of the particles in the sub-system. In the present case the sub-system is isolated from the rest of the Universe and is the same as the system. It is emphasized that this is the adiabatic case in which no interactions with a reservoir are permitted.

For N particles in three dimensions phase space is $6N$-dimensional, with a phase space point being denoted $\boldsymbol{\Gamma} = \{\mathbf{q}, \mathbf{p}\}$, with the positions being

$$\mathbf{q} = \{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N\}, \quad \mathbf{q}_j = \{q_{jx}, q_{jy}, q_{jz}\},$$

(5.2)

and the momenta being

$$\mathbf{p} = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N\}, \quad \mathbf{p}_j = \{p_{jx}, p_{jy}, p_{jz}\}.$$

(5.3)

At this level of description the particles are modeled as spherical. Although it is possible to augment phase space to include orientational degrees of freedom, the present choice is sufficiently realistic to derive the formalism of statistical mechanics without needlessly complicating the mathematical analysis.

The assertion that $6N$-dimensional classical phase space represents the microstates of the isolated sub-system is a statement about the level of description that is necessary and sufficient to describe the physical properties of the present sub-system with an acceptable level of accuracy. It is a statement that for this particular sub-system not only are orientational effects negligible, but so also are quantum effects, the electronic degrees of freedom, intra-nuclear configurations, etc. These effects may be subsumed grossly into an effective interaction potential. But in any case, having decided upon classical phase space, no finer level of description will be pursued.

Of course, this raises the question: if a point in classical phase space is an effective microstate, then might it not have some internal entropy, and might not this entropy vary from point to point? In other words, what is the entropy of a point in classical phase space?

Before addressing these questions we first give Hamilton's equations of motion for classic phase space. A derivation of the primacy of classical phase space and of Hamilton's classical equations of motion from the underlying quantum mechanical equations for the system is given in chapter 7.

5.1.2 Hamilton's equations of motion

Hamilton's equations of motion are

$$\dot{\mathbf{p}}^0 = \frac{-\partial \mathcal{H}(\boldsymbol{\Gamma})}{\partial \mathbf{q}}, \quad \text{and} \quad \dot{\mathbf{q}}^0 = \frac{\partial \mathcal{H}(\boldsymbol{\Gamma})}{\partial \mathbf{p}}.$$

(5.4)

The Hamiltonian, $\mathcal{H}(\boldsymbol{\Gamma})$, is the total energy of the isolated sub-system, and it is the sum of the kinetic and potential energies, $\mathcal{H}(\boldsymbol{\Gamma}) = \mathcal{K}(\mathbf{p}) + U(\mathbf{q})$. Usually the kinetic energy is

$$\mathcal{K}(\mathbf{p}) = \frac{1}{2m} \sum_{j=1}^{N} \sum_{\alpha=x,y,z} p_{j\alpha}^2 = \frac{p^2}{2m}.$$

(5.5)

Here m is the mass of the particles, assuming they are all identical. The momentum derivative of this yields the velocity, $\dot{\mathbf{q}}^0 = \partial \mathcal{K}(\mathbf{p})/\partial \mathbf{p} = \mathbf{p}/m$.

The superscript 0 here and throughout (excepting the quantum adiabatic evolution in subsection 9.4.4) signifies the Hamiltonian or adiabatic evolution. In the present case the sub-system is isolated, and Hamilton's equations of motion are exact and entirely characterize the motion of a point in phase space. In later parts of this chapter and in most of this book the sub-system interacts with a reservoir, and Hamilton's equations of motion for the sub-system is just part of the total evolution. In such cases the superscript 0 is necessary to distinguish the adiabatic part internal to the sub-system from the dissipative and the stochastic parts due to the reservoir, and from the total evolution that is the sum of all three parts.

The energy is a constant of the motion of the isolated sub-system,

$$
\begin{aligned}
\frac{\mathrm{d}\mathcal{H}(\mathbf{\Gamma})}{\mathrm{d}t} &= \dot{\mathbf{\Gamma}}^0 \cdot \nabla \mathcal{H}(\mathbf{\Gamma}) \\
&= \dot{\mathbf{q}}^0 \cdot \nabla_q \mathcal{H}(\mathbf{\Gamma}) + \dot{\mathbf{p}}^0 \cdot \nabla_p \mathcal{H}(\mathbf{\Gamma}) \\
&= \sum_{j=1}^{N} \sum_{\alpha=x,y,z} \left\{ \frac{\partial \mathcal{H}(\mathbf{\Gamma})}{\partial p_{j\alpha}} \frac{\partial \mathcal{H}(\mathbf{\Gamma})}{\partial q_{j\alpha}} - \frac{\partial \mathcal{H}(\mathbf{\Gamma})}{\partial q_{j\alpha}} \frac{\partial \mathcal{H}(\mathbf{\Gamma})}{\partial p_{j\alpha}} \right\} \\
&= 0.
\end{aligned}
$$

(5.6)

In addition to the energy being a constant of the adiabatic motion, it is also approximately linear additive, which is to say it is extensive with the sub-system size. (It is exactly linear additive in the thermodynamic limit when interactions with and across boundaries are relatively negligible.) Because of these two facts energy plays a pre-eminent role in the analysis that follows.

There are six other constants of the adiabatic motion, namely the components of the linear and angular momentum. These tend to play a comparatively minor role in equilibrium thermodynamics and statistical mechanics. (They can be important in hydrodynamics and non-equilibrium thermodynamics.) The reason that momentum is usually ignored is possibly that the conservation law for them depends upon the system having linear and rotational symmetry. Since almost all systems have a fixed boundary, this symmetry is broken, and in practice linear and angular momentum are not conserved if the boundary interactions are accounted for.

In addition to conserving energy, Hamilton's equations of motion are incompressible. That is

$$
\begin{aligned}
\nabla \cdot \dot{\mathbf{\Gamma}}^0 &= \nabla_q \cdot \dot{\mathbf{q}}^0 + \nabla_p \cdot \dot{\mathbf{p}}^0 \\
&= \sum_{j=1}^{N} \sum_{\alpha=x,y,z} \left\{ \frac{\partial^2 \mathcal{H}(\mathbf{\Gamma})}{\partial q_{j\alpha} \partial p_{j\alpha}} - \frac{\partial^2 \mathcal{H}(\mathbf{\Gamma})}{\partial p_{j\alpha} q_{j\alpha}} \right\} \\
&= 0.
\end{aligned}
$$

(5.7)

This vanishing of the compressibility means that phase space volume is conserved during its adiabatic evolution.

To see the relationship between compressibility and volume change, let $\mathbf{\Gamma}_1$ be the initial phase space point, and let $\mathbf{\Gamma}_2 = \mathbf{\Gamma}_1 + \Delta_t \dot{\mathbf{\Gamma}}_1^0$ be the adiabatically evolved point in an infinitesimal time step Δ_t. The relative change in volume element is

$$\left|\frac{d\boldsymbol{\Gamma}_2}{d\boldsymbol{\Gamma}_1}\right| = \left|\mathbf{I} + \Delta_t \frac{d\dot{\boldsymbol{\Gamma}}_1^0}{d\boldsymbol{\Gamma}_1}\right|$$

$$= 1 + \Delta_t \operatorname{Tr} \frac{d\dot{\boldsymbol{\Gamma}}_1^0}{d\boldsymbol{\Gamma}_1} + \mathcal{O}(\Delta_t^2)$$

$$= 1 + \Delta_t \nabla \cdot \dot{\boldsymbol{\Gamma}}^0.$$

(5.8)

The second equality follows because the determinant is the product of the eigenvalues, each one of which is of the form $1 + \Delta_t \lambda_k$, and $\prod_k (1 + \Delta_t \lambda_k) = 1 + \Delta_t \sum_k \lambda_k + \mathcal{O}(\Delta_t^2)$, where the sum of the eigenvalues is just the trace of the corresponding matrix. One sees that if the trajectory velocity has no divergence, then the ratio of volume elements on a trajectory is unity, which is to say they are incompressible.

The incompressibility of phase space during adiabatic evolution will be used to obtain an important result for the conservation of probability density on an adiabatic trajectory below.

5.1.2.1 Uniqueness of the adiabatic trajectory

The adiabatic trajectory passing through $\boldsymbol{\Gamma}_0$ at time t_0 is $\boldsymbol{\Gamma}^0(t \mid \boldsymbol{\Gamma}_0, t_0) = \boldsymbol{\Gamma}^0(t - t_0 \mid \boldsymbol{\Gamma}_0)$. The equality holds because an equilibrium system is homogeneous in time. For Hamilton's equations of motion, for an infinitesimal time step Δ_t, the evolution of the adiabatic trajectory is

$$\boldsymbol{\Gamma}_2 \equiv \boldsymbol{\Gamma}^0(t + \Delta_t \mid \boldsymbol{\Gamma}_1) = \boldsymbol{\Gamma}_1 + \Delta_t \dot{\boldsymbol{\Gamma}}^0(\boldsymbol{\Gamma}_1).$$

(5.9)

By the fundamental theorem of calculus, the full trajectory is obtained by integrating Hamilton's equations,

$$\boldsymbol{\Gamma}^0(t \mid \boldsymbol{\Gamma}_0) = \boldsymbol{\Gamma}_0 + \int_0^t dt' \, \dot{\boldsymbol{\Gamma}}^0(\boldsymbol{\Gamma}^0(t' \mid \boldsymbol{\Gamma}_0)).$$

(5.10)

Since t here may be taken to be positive or negative, Hamilton's equations specify a unique trajectory passing through each point in the isolated sub-system's phase space. Each phase space point has a unique destination in a transition forward in time, and a unique destination in a transition backward in time. This means that adiabatic trajectories can never cross or terminate.

This in turn means that a region of phase space adiabatically evolves in a unique and well-defined fashion: phase space points initially interior to the region can never cross the evolving boundaries as they themselves evolve. Conversely points initially exterior to the specified region can never enter the region. The evolution of a region of phase space is sketched in figure 5.1.
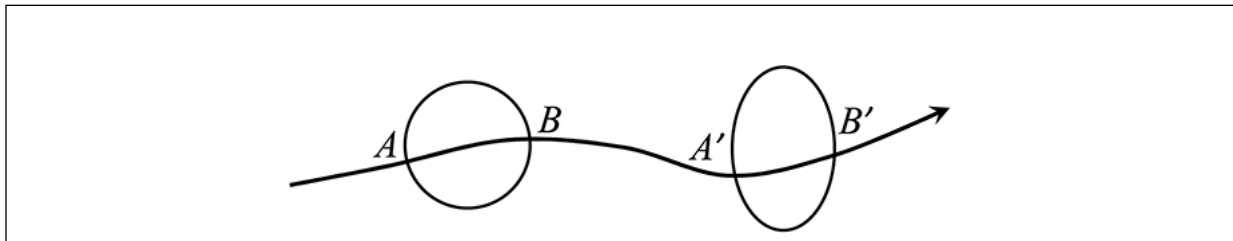
Figure 5.1. An adiabatic trajectory in phase space (curved arrow), and two regions of phase space (circle and ellipse) whose boundaries are related by adiabatic evolution. The labels refer to the times when the trajectory enters and exits the regions considered as fixed in phase space.

Because adiabatic trajectories cannot cross the boundary of an evolving region of phase space, the number of trajectories inside an evolving region is a constant of the adiabatic motion. By the incompressibility of Hamilton's equations proved above, the volume of an evolving region is also a constant of the adiabatic motion. This is also sketched in figure 5.1, where the regions have different shapes but the same volume. These two constants of the motion can be combined to conclude that the number density of trajectories in phase space is a constant of the adiabatic motion. This result will be used in the following subsection, section 5.2.1, to discuss a result due to Boltzmann, namely the uniformity of phase space weight on an adiabatic trajectory.

5.1.2.2 Reversibility of Hamilton's equations

Hamilton's equations of motion are reversible. To see what this means, define the conjugate of a phase space point of the isolated sub-system as the point with all the momenta reversed,

$$\mathbf{\Gamma} = \{\mathbf{q}, \mathbf{p}\} \Leftrightarrow \mathbf{\Gamma}^\dagger = \{\mathbf{q}, -\mathbf{p}\}.$$

(5.11)

The energy is insensitive to the direction of the particle's velocities,

$$\mathcal{H}(\mathbf{\Gamma}) = \mathcal{H}(\mathbf{\Gamma}^\dagger).$$

(5.12)

The adiabatic rate of change of a conjugate point is the negative conjugate of that of the original point,

$$\dot{\mathbf{\Gamma}}^0(\mathbf{\Gamma}^\dagger) = \left\{ \dot{\mathbf{q}}^0(\mathbf{\Gamma}^\dagger), \dot{\mathbf{p}}^0(\mathbf{\Gamma}^\dagger) \right\}$$

$$= \left\{ \frac{\partial \mathcal{H}(\mathbf{\Gamma}^\dagger)}{\partial \mathbf{p}^\dagger}, \frac{-\partial \mathcal{H}(\mathbf{\Gamma}^\dagger)}{\partial \mathbf{q}^\dagger} \right\}$$

$$= \left\{ \frac{-\partial \mathcal{H}(\mathbf{\Gamma})}{\partial \mathbf{p}}, \frac{-\partial \mathcal{H}(\mathbf{\Gamma})}{\partial \mathbf{q}} \right\}$$

$$= -\dot{\mathbf{\Gamma}}^0(\mathbf{\Gamma})^\dagger.$$

(5.13)

Consider an adiabatic transition over a single time step, $\mathbf{\Gamma}_2 = \mathbf{\Gamma}_1 + \Delta_t \dot{\mathbf{\Gamma}}^0(\mathbf{\Gamma}_1)$. To leading order it does not matter whether the velocity is evaluated at $\mathbf{\Gamma}_1$ or $\mathbf{\Gamma}_2$, $\dot{\mathbf{\Gamma}}^0(\mathbf{\Gamma}_1) = \dot{\mathbf{\Gamma}}^0(\mathbf{\Gamma}_2) + \mathcal{O}(\Delta_t)$. Hence in view of the fact that a unique adiabatic trajectory passes through each point in phase space, and the above result for the velocity of the conjugate point, one has three symmetries

$$\mathbf{\Gamma}_2 = \mathbf{\Gamma}_1 + \Delta_t \dot{\mathbf{\Gamma}}^0(\mathbf{\Gamma}) \Leftrightarrow \mathbf{\Gamma}_1 = \mathbf{\Gamma}_2 - \Delta_t \dot{\mathbf{\Gamma}}^0(\mathbf{\Gamma}), \text{(temporal)}$$

$$\Leftrightarrow \mathbf{\Gamma}_1^\dagger = \mathbf{\Gamma}_2^\dagger + \Delta_t \dot{\mathbf{\Gamma}}^0(\mathbf{\Gamma}^\dagger), \text{(conjugate)}$$

$$\Leftrightarrow \mathbf{\Gamma}_2^\dagger = \mathbf{\Gamma}_1^\dagger - \Delta_t \dot{\mathbf{\Gamma}}^0(\mathbf{\Gamma}^\dagger), \text{(microscopic)}.$$

(5.14)

Here $\mathbf{\Gamma}$ can be either $\mathbf{\Gamma}_1$ or $\mathbf{\Gamma}_2$. One sees that microscopic reversibility is a combination of temporal and conjugate reversibility.

Integrating these over time an adiabatic trajectory has symmetries

$$\mathbf{\Gamma}_2 = \mathbf{\Gamma}^0(\tau \mid \mathbf{\Gamma}_1) \Leftrightarrow \mathbf{\Gamma}_1 = \mathbf{\Gamma}^0(-\tau \mid \mathbf{\Gamma}_2), \text{(temporal)}$$

$$\Leftrightarrow \mathbf{\Gamma}_1^\dagger = \mathbf{\Gamma}^0(\tau \mid \mathbf{\Gamma}_2^\dagger), \text{(conjugate)}$$

$$\Leftrightarrow \mathbf{\Gamma}_2^\dagger = \mathbf{\Gamma}^0(-\tau \mid \mathbf{\Gamma}_1^\dagger), \text{(microscopic)}.$$

(5.15)

The first of these can be called the temporal or time reversibility of Hamilton's equations of motion. Microscopic reversibility, the final one, says that if $\mathbf{\Gamma}_2$ is the adiabatic destination of $\mathbf{\Gamma}_1$ going forward in time, then $\mathbf{\Gamma}_2^\dagger$ is the adiabatic destination of $\mathbf{\Gamma}_1^\dagger$ going backward in time. This is sketched in figure 5.2. Note that time goes in the opposite direction on these trajectories. This is what is meant by the microscopic reversibility of Hamilton's equations of motion.

Figure 5.2. Microscopic reversibility for two adiabatic trajectories, from $\mathbf{\Gamma}_1$ (upper) and $\mathbf{\Gamma}_1^\dagger$ (lower). The arrows point in the direction of the transition; the two trajectories move in opposite directions in time.
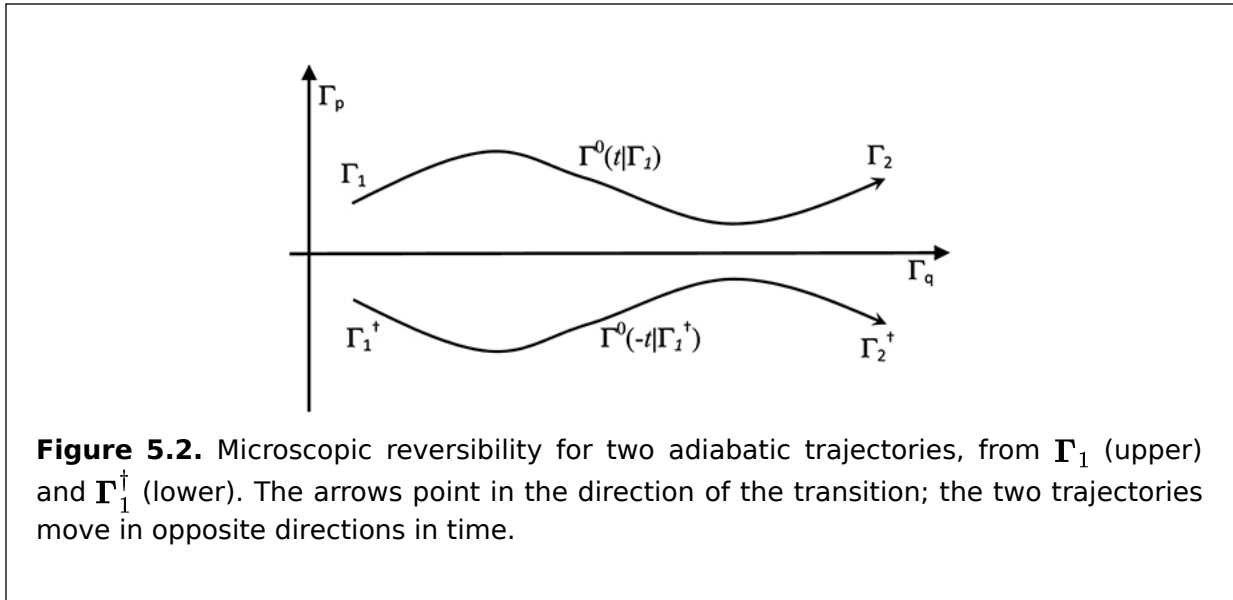
5.1.2.3 Time-dependent potential

At this point a small digression can be made to address the nature of microscopic reversibility in the case that the sub-system has applied to it an external potential that varies explicitly with time, $U(\mathbf{\Gamma}, t)$.

Consider the time interval $[t_1, t_2]$, and let the termini of the adiabatic trajectory satisfy $\mathbf{\Gamma}_2 = \mathbf{\Gamma}^0(t_2 \mid \mathbf{\Gamma}_1, t_1)$. This of course can equally well be written $\mathbf{\Gamma}_1 = \mathbf{\Gamma}^0(t_1 \mid \mathbf{\Gamma}_2, t_2)$. On this time interval one can also write $\mathbf{\Gamma}\prime = \mathbf{\Gamma}^0(t\prime \mid \mathbf{\Gamma}_1, t_1) = \mathbf{\Gamma}^0(t\prime \mid \mathbf{\Gamma}_2, t_2)$, $t\prime \in [t_1, t_2]$.

Define the conjugate potential on this time interval as

$$\tilde{U}(\mathbf{\Gamma}, t) \equiv U(\mathbf{\Gamma}, t_1 + t_2 - t), \quad t \in \left[t_1, t_2\right].$$

This is in effect the temporal mirror image of the original potential on the time interval. In so far as the applied potential arises from atoms and molecules external to the system, reversing its time dependence in this way as well as conjugating the momenta of the particles of the sub-system is equivalent to reversing the momenta of all the atoms and molecules in the Universe.

This conjugate potential is required for the reverse trajectory,

$$\mathbf{\Gamma}_2 = \mathbf{\Gamma}^0(t_2 \mid \mathbf{\Gamma}_1, t_1; [U]) \Leftrightarrow \mathbf{\Gamma}_1^\dagger = \mathbf{\Gamma}^0(t_2 \mid \mathbf{\Gamma}_2^\dagger, t_1; [\tilde{U}]),$$

and

$$\mathbf{\Gamma}\prime = \mathbf{\Gamma}^0(t\prime \mid \mathbf{\Gamma}_1, t_1; [U]) \Leftrightarrow \mathbf{\Gamma}\prime^\dagger = \mathbf{\Gamma}^0(t\prime\prime \mid \mathbf{\Gamma}_2^\dagger, t_1; [\tilde{U}]), \quad t\prime\prime \equiv t_1 + t_2 - t\prime.$$

This is one form of microscopic reversibility for a time-dependent potential.

5.2 Entropy of a point in phase space

The aim in this section is to derive the entropy of a point in the phase space of the isolated system, $S(\mathbf{\Gamma})$. By the axioms of chapter 1, this is in essence equivalent to the weight density $w(\mathbf{\Gamma})$ and the probability density $\wp(\mathbf{\Gamma})$. This 'internal' entropy of a phase space point for the isolated system is a necessary prerequisite for the derivation of the Maxwell–Boltzmann probability for an open sub-system that can exchange energy with a heat reservoir (see section 5.3).

In the following sub-sections, three approaches are discussed, each with the same aim of proving that the entropy density is constant in phase space, which means that it can be set to zero, $S(\mathbf{\Gamma}) = 0$. Equivalently, the weight density and the probability density are constant, $w(\mathbf{\Gamma}) = \text{const.}$ and $\wp(\mathbf{\Gamma}) = \text{const.}$ The first approach, section 5.2.1, is the ergodic hypothesis, which essentially takes the result to be an axiom. The second approach, section 5.2.2, essentially proves that the probability density is uniform on an adiabatic trajectory, and assumes that this implies the final result. The third approach, section 5.2.3, derives the final result from time homogeneity.

5.2.1 Ergodic hypothesis

In most books on statistical mechanics the uniform weight density of phase space is but one of a several axioms that together form what is called the ergodic hypothesis (Boltzmann 1871). This hypothesis is traditionally central to the formulation of statistical mechanics; Münster (1969), for example, discusses the role of the former in the axiomatic development of the latter.

The ergodic hypothesis comprises three points: for an isolated sub-system
1. all points in phase space with a given energy lie on a single trajectory;
2. the probability density in phase space is uniform on an energy hypersurface;
3. the probability density in phase space is uniform.

The second of these is sometimes called the equal *a priori* hypothesis (Pathria 1972, section 2.3).

The first of these assertions says that a trajectory is dense on the energy hypersurface that it is confined to. It further says that the hypersurface is not broken up into distinct regions each covered by its own trajectory that can't communicate with the rest.

Since adiabatic trajectories do not terminate, they must be in the form of loops. Hence, since every point on the energy hypersurface belongs to an adiabatic trajectory, it is obvious that the entire energy hypersurface must be covered by one or more loops, each distinct and non-communicating. According to this first axiom, there is only one such loop for the whole energy hypersurface.

In the formal development of statistical mechanics this first axiom appears to be essential, because it is necessary to go from a property established for a single trajectory to a property of the entire energy hypersurface (and thence to a property of the entire phase space). In the following sub-subsection, following Boltzmann it is established that the probability density on a given trajectory has a uniform value. But it does not follow that this value must be the same on all distinct trajectories with the same energy. Hence without the first axiom of the ergodic hypothesis it is not possible to prove that the second axiom is a consequence of Boltzmann's result.

It strains credulity that an energy hypersurface with Avogadro's number of dimensions should be covered by one and only one trajectory loop. As a contrary example, if the isolated sub-system was itself composed of isolated sub-systems, then there would be at least one distinct adiabatic trajectory loop for each such sub-system, the starting position of each being arbitrary. Further, the total energy can be arbitrarily partitioned amongst the isolated sub-sub-systems. Each combination of starting positions or energies corresponds to a different trajectory loop of the total sub-system. This example obviously violates the first axiom of the ergodic hypothesis.

Some texts emphasize the second assertion in the ergodic hypothesis, which is claimed to be self-evident: because there is nothing to choose between phase space points with the same energy, they must all have the same weight. (Against this one could point out that the speed of the trajectory varies on the energy hypersurface, and this could reasonably be expected to influence the weight.) Equivalently, on the basis of the first axiom, it is claimed that the system must pass thorough all possible points 'without fear or favor' (Pathria 1972, p 38).

Whether or not one finds the second assertion self-evident, it is the third assertion that is the one that is really needed to obtain the Maxwell–Boltzmann probability density. The point is that even if one accepts the second assertion, or the result derived next—that the probability density is uniform on a trajectory—one still has to find a way to prove the third assertion.

5.2.2 Constant probability on a trajectory

Consider the adiabatic trajectory $\mathbf{\Gamma}^0(t \mid \mathbf{\Gamma}_0)$, which passes through the point $\mathbf{\Gamma}_0$ at time $t = 0$. Consider also the possibly time-dependent phase space probability density, $\wp(\mathbf{\Gamma}, t)$ that has initial value $\wp(\mathbf{\Gamma}, 0) = \wp_0(\mathbf{\Gamma})$. Using the adiabatic equations of motion for the present isolated sub-system, the Fokker–Planck equation, equation (4.36), for the partial time derivative is

$$
\begin{aligned}
\frac{\partial \wp(\mathbf{\Gamma}, t)}{\partial t} &= -\nabla \cdot \dot{\mathbf{\Gamma}}^0 \wp(\mathbf{\Gamma}, t) - \dot{\mathbf{\Gamma}}^0 \cdot \nabla \wp(\mathbf{\Gamma}, t) \\
&= -\dot{\mathbf{\Gamma}}^0 \cdot \nabla \wp(\mathbf{\Gamma}, t).
\end{aligned}
$$

(5.19)

It is worth emphasizing that this result is a direct consequence of the conservation law for the transition probability, equation (3.17). The second equality follows because the adiabatic equations of motion are incompressible.

With this, the total rate of change of the probability density is

$$\frac{\mathrm{d}\wp(\mathbf{\Gamma}, t)}{\mathrm{d}t} = \frac{\partial \wp(\mathbf{\Gamma}, t)}{\partial t} + \dot{\mathbf{\Gamma}}^0 \cdot \nabla \wp(\mathbf{\Gamma}, t)$$
$$= -\dot{\mathbf{\Gamma}}^0 \cdot \nabla \wp(\mathbf{\Gamma}, t) + \dot{\mathbf{\Gamma}}^0 \cdot \nabla \wp(\mathbf{\Gamma}, t)$$
$$= 0.$$

(5.20)

The first equality is the definition of the total rate of change with time. Combining this with the Fokker–Planck equation gives the second equality. This is just Liouville's theorem, equation (4.80), which is exact for an adiabatic trajectory. This vanishing of the total time derivative says that the probability density in phase space is a constant of the adiabatic motion, $\wp(\mathbf{\Gamma}^0(t \mid \mathbf{\Gamma}_0), t) = \wp_0(\mathbf{\Gamma}_0)$.

One can give a straightforward physical interpretation to this result. As has been mentioned above in connection with figure 5.1, the adiabatic incompressibility of phase space means that volume is conserved during their evolution under Hamilton's equations of motion. Since adiabatic trajectories do not cross, trajectories inside the volume stay there for their entire evolution. Combined with the conservation law for weight during a transition, equation (3.17), this implies that the total weight inside the volume is a constant of the adiabatic motion. Since the volume itself is constant, this means that the weight density, and hence the probability density is constant along the length of a trajectory.

Boltzmann gave an equivalent argument that is also illuminating. In figure 5.1 are shown two regions, the second of which evolved from the first adiabatically. An adiabatic trajectory is also shown, along with the times it entered, t_A and $t_{A'}$, and exited, t_B and $t_{B'}$, the respective regions (considered now as fixed in phase space). Since the evolution of the region is defined by the evolution of the boundary, the difference in these respective times must be equal, $t_{B'} - t_B = t_{A'} - t_A$. This means that the trajectory spent an equal amount of time in each region, $t_B - t_A = t_{B'} - t_{A'}$. Assuming that the weight of a region of phase space is proportional to the time spent in it, it follows that regions of phase space lying on a single trajectory have the same weight density.

Both arguments show that the probability density is a constant of the adiabatic motion $\wp(\mathbf{\Gamma}^0(t \mid \mathbf{\Gamma}_0), t) = \wp_0(\mathbf{\Gamma}_0)$.

Now for the present equilibrium system, the probability density must be independent of time,

$$\frac{\partial \wp(\mathbf{\Gamma}, t)}{\partial t} = 0.$$

(5.21)

Hence $\wp(\mathbf{\Gamma}, t) = \wp(\mathbf{\Gamma})$. These last two results imply that $\wp(\mathbf{\Gamma}^0(t \mid \mathbf{\Gamma}_0)) = \wp_0(\mathbf{\Gamma}_0)$. This says that the probability density is uniform along the length of an adiabatic trajectory. That is, if $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ lie on the same trajectory, $\mathbf{\Gamma}_2 = \mathbf{\Gamma}^0(t_2 \mid \mathbf{\Gamma}_1, t_1)$, then they must have the same probability density, $\wp_0(\mathbf{\Gamma}_2) = \wp_0(\mathbf{\Gamma}_1)$.

With the first part of the ergodic hypothesis, this result becomes that all points that lie on the same energy hypersurface have equal probability density. As mentioned above, this is the equal *a priori* hypothesis (Pathria 1972, section 2.3). In equation form this says that given that the isolated sub-system has energy E, then the phase space probability density is

$$\wp(\mathbf{\Gamma} \mid E) = \frac{\delta(\mathcal{H}(\mathbf{\Gamma}) - E)}{\Omega(E)}.$$

(5.22)

Here the normalizing factor is essentially the number of phase space points in the energy hypersurface,

$$\Omega(E) = \int d\mathbf{\Gamma} \; \delta(\mathcal{H}(\mathbf{\Gamma}) - E).$$

(5.23)

However, even if one accepts this result (notwithstanding the arguments given above against the first axiom), one still has to convert this result from a conditional probability density to an unconditional probability density. That is, it is the phase space weight $w(\mathbf{\Gamma})$ that is required.

Formally the unconditional probability density is

$$\wp(\mathbf{\Gamma}) = \int dE \; \wp(\mathbf{\Gamma} \mid E)\wp(E)$$
$$= \int dE \; \frac{\delta(\mathcal{H}(\mathbf{\Gamma}) - E)}{\Omega(E)} \frac{e^{S(E)/k_{\mathrm{B}}}}{W}.$$

(5.24)

Again formally, the entropy of an energy macrostate is given by the weight of the macrostate, which is just the sum total of the phase space weight in the macrostate

$$e^{S(E)/k_{\mathrm{B}}} \equiv W(E) = \int d\mathbf{\Gamma} \; \omega(\mathbf{\Gamma}) \, \delta(\mathcal{H}(\mathbf{\Gamma}) - E).$$

(5.25)

With this the phase space probability density

$$\wp(\mathbf{\Gamma}) = \int dE \; \frac{\delta(\mathcal{H}(\mathbf{\Gamma}) - E)}{\Omega(E)} \frac{W(E)}{W}$$
$$= \frac{W(\mathcal{H}(\mathbf{\Gamma}))}{\Omega(\mathcal{H}(\mathbf{\Gamma})) \, W}.$$

(5.26)

In general, since neither $W(\mathcal{H}(\mathbf{\Gamma}))$ nor $\Omega(\mathcal{H}(\mathbf{\Gamma}))$ are known (contrast equation (5.23) with equation (5.25)), this is not a constant in phase space.

However, if one assumes uniform phase space weight, $w(\mathbf{\Gamma}) = 1$ (any set of weights can be re-scaled by a positive factor, and so if this is constant, then that constant may be set to unity), then one has

$$W(E) = \Omega(E) \text{ iff } \omega(\mathbf{\Gamma}) = 1,$$

(5.27)

and the phase space probability density becomes

$$\wp(\mathbf{\Gamma}) = \frac{1}{W},$$

(5.28)

which is the desired constant in phase space. Regrettably this result begs the question: it shows that phase space has uniform weight if it is assumed that phase space has uniform weight.

5.2.3 Time and energy

Now the weight density of phase space is derived from a more satisfactory perspective that avoids invoking the three axioms of the ergodic hypothesis.

The adiabatic trajectory $\mathbf{\Gamma}^0(t \mid \mathbf{\Gamma}_0)$ that is determined by the Hamiltonian is confined to the energy hypersurface $\mathcal{H}(\mathbf{\Gamma}) = E$. This suggests proceeding in two stages to obtain the phase space weight density: first obtain the weight density on an energy hypersurface, and then obtain the density of energy hypersurfaces in phase space.

Define a coordinate system for the energy hypersurface, $\gamma(\mathbf{\Gamma})$. The relationship is invertible, $\mathbf{\Gamma}(\gamma, E)$. The trajectory may be written as $\gamma(t) = \gamma(\mathbf{\Gamma}(t))$. It is axiomatic that an average of a phase space function is a simple time average over a trajectory, which is to say that time is homogeneous. In consequence of this, the weight density on the energy hypersurface must be inversely proportional to the speed of the trajectory at that point,

$$\omega(\gamma \mid E) \propto \frac{1}{|\dot{\gamma}^0|}$$

$$= \frac{1}{|\dot{\mathbf{\Gamma}}^0|}.$$

(5.29)

The second equality follows because on a trajectory over a time step, $|\gamma_2 - \gamma_1| = |\mathbf{\Gamma}_2 - \mathbf{\Gamma}_1|$. Since a time average is a simple average, the weight is large in regions of slow speed because the system spends more time there. This physical picture is incompatible with the equal *a priori* hypothesis. In the left of figure 5.3 can be seen an adiabatic trajectory of varying speed. Closely spaced marks signify slow speed, in which regions the trajectory spends more time, and which therefore have larger weight.

Figure 5.3. Left: an adiabatic trajectory in phase space marked in equal time increments. Right: contours of the energy hypersurface. The boxes signify regions of phase space, with the dashed region having greater weight than the dotted region in each case, all other things being equal.
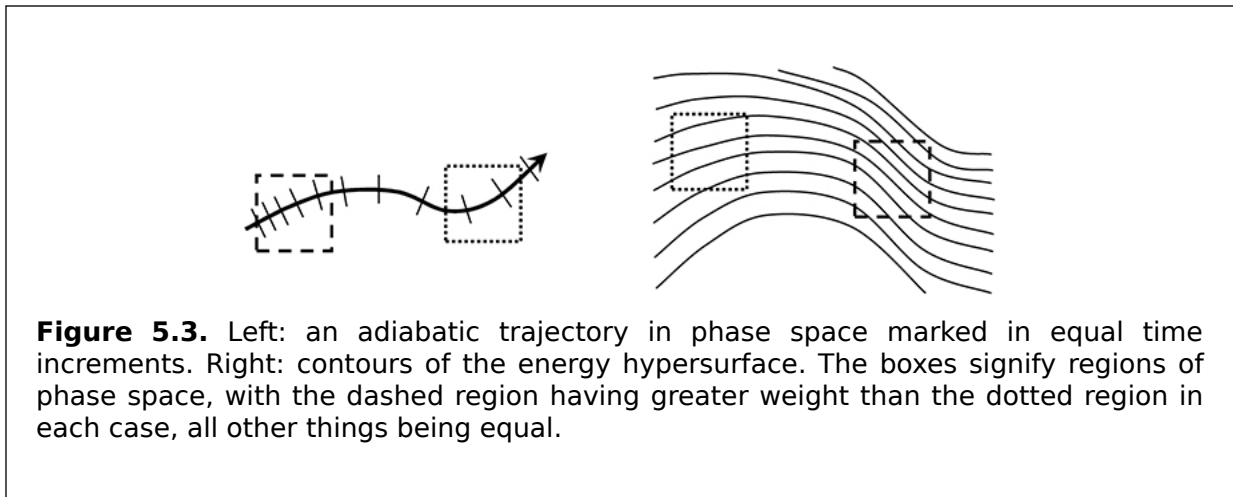
The proportionality constant neglected here is essentially the length of the time interval over which the weights are measured. It is important to note that this cannot depend upon the energy E, and that it must be the same constant on different energy hypersurfaces. The assertion that time is uniform, and that the weight of a state is proportional only to the time spent in the state, precludes any such energy dependence.

The weight density of phase space itself is related to this this by the usual rule for the transformation of densities,

$$\omega(\mathbf{\Gamma} \mid E) = \omega(\boldsymbol{\gamma} \mid E)\frac{\mathrm{d}\boldsymbol{\gamma}}{\mathrm{d}\mathbf{\Gamma}}, \quad \mid \mathcal{H}(\mathbf{\Gamma}) - E \mid < \mathrm{d}E$$

$$= \frac{1}{\mid \dot{\mathbf{\Gamma}}^0 \mid} \frac{\mid \nabla\mathcal{H}(\mathbf{\Gamma}) \mid}{\mathrm{d}E}, \quad \mid \mathcal{H}(\mathbf{\Gamma}) - E \mid < \mathrm{d}E$$

$$= \frac{\mid \nabla\mathcal{H}(\mathbf{\Gamma}) \mid}{\mid \dot{\mathbf{\Gamma}}^0 \mid} \delta(\mathcal{H}(\mathbf{\Gamma}) - E).$$

(5.30)

The second factor from the transformation of coordinates, $\mid \nabla\mathcal{H}(\mathbf{\Gamma}) \mid /\mathrm{d}E$, is essentially the number of energy hypersurfaces per unit phase space. One can perhaps see the need and form for this transformation more clearly by noting that the volume of the energy hypersurface can be equally written

$$A(E) = \oint_E \mathrm{d}\boldsymbol{\gamma}$$

$$= \int \mathrm{d}\mathbf{\Gamma} \mid \nabla\mathcal{H}(\mathbf{\Gamma}) \mid \delta(\mathcal{H}(\mathbf{\Gamma}) - E).$$

(5.31)

With this transformation rule, the expression for the phase space weight gives the same total weight whether expressed as an integral over the energy hypersurface, or as an integral over phase space,

$$W(E) = \oint_E \mathrm{d}\boldsymbol{\gamma}\, \omega(\boldsymbol{\gamma} \mid E)$$

$$= \int \mathrm{d}\mathbf{\Gamma}\, \omega(\boldsymbol{\gamma} \mid E) \mid \nabla\mathcal{H}(\mathbf{\Gamma}) \mid \delta(\mathcal{H}(\mathbf{\Gamma}) - E)$$

$$= \int \mathrm{d}\mathbf{\Gamma}\, \omega(\mathbf{\Gamma} \mid E).$$

(5.32)

The result for $\omega(\mathbf{\Gamma} \mid E)$ says that the weight is larger for steep gradients because in such regions there are more energy hypersurfaces per unit phase space distance. This real physical effect is also inconsistent with the equal *a priori* hypothesis. In the right of figure 5.3 can be seen the contours of the energy hypersurface of an isolated system. The box crossed by many contour lines has greater weight than the box crossed by few. The former lies in a region with steeper energy gradient than the latter.

The function $\omega(\mathbf{\Gamma} \mid E)$ is the phase space weight density conditional on a given energy. The unconditional weight of an isolated system phase space point is

$$\omega(\mathbf{\Gamma}) = \int \mathrm{d}E\, \omega(\mathbf{\Gamma} \mid E)$$

$$= \frac{\mid \nabla\mathcal{H}(\mathbf{\Gamma}) \mid}{\mid \dot{\mathbf{\Gamma}}^0 \mid}.$$

(5.33)

As was argued above, there cannot be an additional weight factor $\tilde{w}(E) \Rightarrow \tilde{w}(\mathcal{H}(\mathbf{\Gamma}))$ included in this final result because this would violate the principle of time homogeneity.

This has a straightforward physical interpretation. The weight of a phase space point is proportional to the time the system spends in the volume element about that point, which is to say that it is inversely proportional to the speed of the trajectory at that point. Since the system is confined to an energy hypersurface, the weight of a phase space point is also proportional to the number of energy hypersurfaces per unit volume, which is to say that it is proportional to the gradient of the energy at that point.

Both of these properties—the trajectory speed and the energy gradient—are local properties in phase space. The big advantage of this derivation of the weight density of phase space, $\omega(\mathbf{\Gamma})$, is that it relies only on the properties of the particular point, $\mathbf{\Gamma}$. This contrasts with the ergodic hypothesis, for example, where one has to postulate a global property, namely that a single trajectory passes infinitesimally close to every point on an energy hypersurface, and also that the weights on all energy hypersurfaces are equal.

In order to evaluate the energy gradient and the speed that appear here, position and momentum components have to be added together, and these have different dimensions. To get around this problem a metric element, the length scales l_q and l_p, can be introduced into the definition of the scalar product and the gradient operator (Galea and Attard [2002]). Scalar products have the form

$$\mathbf{\Gamma}_1 \cdot \mathbf{\Gamma}_2 = \frac{1}{l_q^2}\mathbf{q}_1 \cdot \mathbf{q}_2 + \frac{1}{l_p^2}\mathbf{p}_1 \cdot \mathbf{p}_2,$$

(5.34)

and gradients have the form

$$\nabla f(\mathbf{\Gamma}) = l_q^2 \frac{\partial f(\mathbf{\Gamma})}{\partial \mathbf{q}} + l_p^2 \frac{\partial f(\mathbf{\Gamma})}{\partial \mathbf{p}}.$$

(5.35)

The ratio of the particular values chosen for these length scales will change the value of the length of the gradient and of the speed of the trajectory, but it will make no difference to the final conclusion that the two cancel.

With these scale factors, the square of the magnitude of the gradient of the Hamiltonian is

$$\begin{aligned}
\mid \nabla \mathcal{H}(\mathbf{\Gamma}) \mid^2 &= \frac{l_q^4}{l_q^2} \frac{\partial \mathcal{H}}{\partial \mathbf{q}} \cdot \frac{\partial \mathcal{H}}{\partial \mathbf{q}} + \frac{l_p^4}{l_p^2} \frac{\partial \mathcal{H}}{\partial \mathbf{p}} \cdot \frac{\partial \mathcal{H}}{\partial \mathbf{p}} \\
&= l_q^2 \dot{\mathbf{p}}^0 \cdot \dot{\mathbf{p}}^0 + l_p^2 \dot{\mathbf{q}}^0 \cdot \dot{\mathbf{q}}^0 \\
&= l_q^2 l_p^2 \mid \dot{\mathbf{\Gamma}}^0 \mid^2.
\end{aligned}$$

(5.36)

The proportionality of these is a remarkable coincidence, and is a unique feature of Hamilton's equations of motion. The proportionality constant $l_q l_p$ has the dimensions of energy × time, which is the same dimensions as Planck's constant h. Since such positive multiplicative constants are immaterial, the phase space weight density of the isolated system is constant. Without loss of generality, it may be set equal to unity,

$$\omega(\mathbf{\Gamma}) = 1.$$

(5.37)

This is obviously a constant on an adiabatic trajectory. In this case this result is consistent with, but stronger than, the result derived in section 5.2.2 from the Fokker–Planck equation and Liouville's theorem: the probability density is uniform along the length

of an adiabatic trajectory, $\wp(\mathbf{\Gamma}^0(t \mid \mathbf{\Gamma}_0)) = \wp_0(\mathbf{\Gamma}_0)$. Neither the trajectory speed nor the energy gradient are constants of the adiabatic motion, and hence it is essential that they cancel as here. This result says that the weight density is not just constant on a trajectory, but that it is a constant throughout phase space itself.

With this result, the total weight of an energy macrostate for the isolated system is just

$$W(E) = \int \mathrm{d}\mathbf{\Gamma} \ \delta(\mathcal{H}(\mathbf{\Gamma}) - E).$$

(5.38)

One can interpret this as the number of phase space points (microstates) in an energy macrostate of an isolated system. This is not the volume of the energy hypersurface, $A(E)$.

5.3 Canonical equilibrium system

5.3.1 Constrained entropy of a phase space point

The canonical equilibrium system consists of a sub-system that can exchange energy with a thermal reservoir. The Hamiltonian of the sub-system is independent of time, $\mathcal{H}(\mathbf{\Gamma})$. The microstates of the system, which are the smallest indivisible states that will be analyzed, will be taken to be the microstates of the sub-system, namely the points in the sub-system phase space $\mathbf{\Gamma}$.

Conversely, it could be argued that the sub-system phase space point $\mathbf{\Gamma}$ is actually a macrostate of the total system, since the microstates of the total system are in fact points in the total phase space of the sub-system and reservoir, $\mathbf{\Gamma}_{\text{total}} = \{\mathbf{\Gamma}, \mathbf{\Gamma}_{\mathrm{r}}\}$. However, since the rationale of the reservoir formalism is to focus on the sub-system, I think it best to regard the sub-system phase space point $\mathbf{\Gamma}$ as a microstate of the total system. The projection of the reservoir phase space points $\mathbf{\Gamma}_{\mathrm{r}}$ out of the formalism contributes to the internal entropy of the microstate $\mathbf{\Gamma}$. This argument over terminology is purely academic because the formalism for entropy is the same for microstates as for macrostates.

To obtain the probability of a microstate $\mathbf{\Gamma}$ we need its total entropy. The result has the same form as the total constrained entropy obtained in section 2.3.1. That analysis was also for the canonical equilibrium system, with the difference being that the sub-system energy macrostate E was used. In the present case we essentially replace E in equation (2.21) by the present sub-system microstate $\mathbf{\Gamma}$.

First we need the temperature of the reservoir. The entropy of an isolated system with energy E is by definition

$$S(E) = k_{\mathrm{B}} \ln W(E).$$

(5.39)

where the total weight is the number of phase space points in the energy hypersurface, as given by the immediately preceding analysis. Here the number and volume arguments have been suppressed.

By the definition (2.11), the temperature is given by

$$\frac{1}{T} \equiv \frac{\partial S(E)}{\partial E} = \frac{k_{\mathrm{B}}}{W(E)} \frac{\partial W(E)}{\partial E}.$$

(5.40)

If the fixed total energy of the total isolated system is E_{total}, then the reservoir energy in the sub-system microstate $\mathbf{\Gamma}$ is $E_{\mathrm{r}}(\mathbf{\Gamma}) = E_{\text{total}} - \mathcal{H}(\mathbf{\Gamma})$, and the reservoir entropy in that

microstate is

$$S_{\mathrm{r}}(E_{\mathrm{r}}(\boldsymbol{\Gamma})) \equiv S_{\mathrm{r}}(E_{\mathrm{total}}) - \mathcal{H}(\boldsymbol{\Gamma})\frac{\partial S_{\mathrm{r}}(E_{\mathrm{total}})}{\partial E_{\mathrm{total}}}$$

$$= \mathrm{const.} - \frac{1}{T}\mathcal{H}(\boldsymbol{\Gamma}),$$

(5.41)

where T is the reservoir temperature. The constant part independent of the sub-system may be dropped.

As in equation (2.21), the constrained total entropy for the canonical equilibrium system in the microstate $\boldsymbol{\Gamma}$ is

$$S_{\mathrm{total}}(\boldsymbol{\Gamma} \mid T) = S_{\mathrm{s}}(\boldsymbol{\Gamma}) + S_{\mathrm{r}}(E_{\mathrm{r}}(\boldsymbol{\Gamma}))$$

$$= \frac{-1}{T}\mathcal{H}(\boldsymbol{\Gamma}),$$

(5.42)

since there is no internal entropy for a point in the sub-system phase space, $S_{\mathrm{s}}(\boldsymbol{\Gamma}) = k_{\mathrm{B}} \ln \omega(\boldsymbol{\Gamma}) = 0$.

5.3.2 Maxwell–Boltzmann probability

With this result, the phase space probability distribution for the canonical equilibrium system is

$$\wp(\boldsymbol{\Gamma} \mid N, V, T) = \frac{1}{Z\prime(T)}\mathrm{e}^{S_{\mathrm{r}}(\boldsymbol{\Gamma}|T)/k_{\mathrm{B}}}$$

$$= \frac{\mathrm{e}^{-\beta\mathcal{H}(\boldsymbol{\Gamma})}}{N!h^{3N}Z(N,V,T)}.$$

(5.43)

This is the Maxwell–Boltzmann distribution. Here and throughout, the inverse temperature is $\beta \equiv 1/k_{\mathrm{B}}T$. The pre-factor of $N!h^{3N}$ will be discussed shortly.

The normalizing partition function is

$$Z(N,V,T) = \frac{1}{N!h^{3N}}\int \mathrm{d}\boldsymbol{\Gamma} \; \mathrm{e}^{-\beta\mathcal{H}(\boldsymbol{\Gamma})}.$$

(5.44)

This is the total weight of the system. As has been mentioned, weight is only defined up to an arbitrary positive constant, and so the pre-factor $N!h^{3N}$ has relatively trivial effects.

Planck's constant $h = 6.626 \times 10^{-34}$ J s has the dimensions of energy × time (equivalently, position × momentum), and including it makes the partition function dimensionless, which is often desirable even though it is not essential. This factor is often interpreted to mean that quantum mechanics gives a phase space volume per microstate of h^{3N}. This is a little loose, but not completely incorrect. Two points can be made: first, within classical mechanics there is nothing to give Planck's constant, and the above formalism would still be valid if it were replaced by something else with the same dimensions or even by nothing at all. Weight is only defined up to a positive scale factor. Second, in chapter 7 this result is derived as the leading order term in an expansion of

quantum statistical mechanics, and in this exact expansion Planck's constant is mandated and appears exactly as here.

The factor of $N!$ was introduced by Gibbs as the resolution to the paradox of the entropy of mixing (Pathria 1972, section 1.5). The way to understand it is to note that the entropy or weight of a state should count only the weight of *distinct* configurations in the state. Since particles are indistinguishable, a permutation of the particle labels does not change the configuration. For example, for two particles, particle 1 at $\mathbf{r}$ and particle 2 at $\mathbf{s}$ is the same configuration as particle 1 at $\mathbf{s}$ and particle 2 at $\mathbf{r}$. The partition function above involves the integral over all of phase space, which means that each distinct configuration is counted $N!$ times. (In classical mechanics configurations with particles at the same phase space point form a set of measure zero and can be neglected.) This over counting is corrected by the pre-factor of $1/N!$ in the partition function. In chapter 7, this factor arises from wave function symmetrization in the zeroth order term in an expansion of quantum statistical mechanics.

These two factors together may be considered to be the uniform 'internal' weight of a phase space of the isolated sub-system,

$$w(\mathbf{\Gamma}) = \frac{1}{N! h^{3N}}.$$

(5.45)

In this case the sub-system entropy of a phase space point is

$$S_{\mathrm{s}}(\mathbf{\Gamma}) = k_{\mathrm{B}} \ln w(\mathbf{\Gamma}) = N k_{\mathrm{B}} - N k_{\mathrm{B}} \ln(N h^3),$$

(5.46)

where Stirling's approximation for $N!$ has been used. In general, one can either show $w(\mathbf{\Gamma}) = 1/N! h^{3N}$ as an explicit pre-factor for the phase space probability and use the reservoir entropy in the exponent, or else one can include it as the sub-system entropy and use the total entropy in the exponent.

The statistical mechanical Helmholtz free energy is essentially the logarithm of this partition function,

$$F(N, V, T) = -k_{\mathrm{B}} T \ln Z(N, V, T).$$

(5.47)

This is a particular form of the generic analysis given in sections 2.3.2 and 2.5.3.

5.3.3 Not Boltzmann distributed

At the beginning of this chapter, the widespread use of the Boltzmann distribution in the form of equation (5.1),

$$\wp_\alpha = \frac{1}{Z(T)} \mathrm{e}^{-U_\alpha / k_{\mathrm{B}} T},$$

(5.48)

was described as questionable. Here $\wp_\alpha$ is meant to be the probability of the state α, U_α is its potential energy, and T is the temperature.

Comparison with the Maxwell–Boltzmann distribution, equation (5.43) shows that this neglects the kinetic energy. This is a relatively minor problem because often one is only interested in static properties and since the kinetic energy and the potential energy are separable, the velocity can be integrated out.

A much more serious problem is that this neglects any internal entropy associated with the state α. The exponent that appears explicitly here is the entropy of the thermal reservoir alone. The point of section 5.2.3 was to rigorously justify that the internal entropy

of a point in phase space could be neglected, $S_{\mathrm{s}}(\mathbf{\Gamma}) = 0$. This was necessary for the Maxwell–Boltzmann phase space form. The problem with the above is that it neglects without justification $S_{\mathrm{s}}(\alpha)$, the internal entropy of the sub-system macrostate α. In general, for an arbitrary static property the entropy cannot simply be assumed constant or negligible. Instead one must use

$$\wp_\alpha = \frac{1}{Z(T)} e^{S_{\mathrm{s}}(\alpha)/k_{\mathrm{B}}} e^{-U_\alpha/k_{\mathrm{B}}T} = \frac{1}{Z(T)} e^{-F_\alpha/k_{\mathrm{B}}T},$$

(5.4
9)

where $F_\alpha = U_\alpha - TS_{\mathrm{s}}(\alpha)$ is the Helmholtz free energy of the macrostate.

For example, if the state α represents the conformation of a macromolecule in a solvent, in addition to the potential energy between the atoms of the macromolecule in this particular configuration, there is also the (weighted) number of possible arrangements of the solvent molecules, which varies with α. As another example, if α represents the location of a defect in a crystal, in addition to the variation of the potential energy of the crystal with defect location, one may also need to account for the change in the spectrum of vibrations of the crystal with defect location, which is another form of entropy. Or if α represents a crystal type, one would need to include the entropy associated with the weighted displacements of the atoms about their mean locations that can occur without destroying the crystal structure.

It may well be that for the specific problem at hand one has reason to believe that any such internal entropy is either constant or its variation is relatively negligible. The important thing is to explicitly justify such an assumption in the given case. Conversely, it is poor practice and likely erroneous to invoke the Boltzmann distribution without explicit justification for neglecting the internal entropy of the state.

5.4 Stochastic, dissipative equations of motion

5.4.1 Equations of motion

We now turn to the equations of motion for the canonical equilibrium system. As above, $\mathbf{\Gamma} = \{\mathbf{q}, \mathbf{p}\}$ is a point in the phase space of the sub-system, $\mathcal{H}(\mathbf{\Gamma})$ is the sub-system energy, and T is the temperature of the reservoir with which the sub-system can exchange energy.

The equations of motion are dominated by the adiabatic part that is internal to the sub-system. The additional external contribution due to the interactions with the reservoir must be a relatively small perturbation because they directly occur only in the region of the boundary of the sub-system, and this is much smaller than the sub-system itself.

These reservoir interactions are treated statistically rather than literally. The stochastic, dissipative terms in the equations of motion that they give rise to are applied equally throughout the sub-system rather than only in the boundary region. This is in the spirit of the reservoir formalism, where the focus is on the sub-system, and the reservoir only enters through certain generic parameters that abstract the microscopic details in any particular case. Of course the functional form and magnitude of the stochastic, dissipative terms are uniquely fixed by the thermodynamic properties of the reservoir.

The Langevin equation for Brownian motion was given in section 4.1.2, and its relationship to the second entropy form for the transition of a thermodynamic fluctuation, equation (3.87), was discussed. One can take the second entropy form of the Langevin equation for Brownian motion, equation (4.21), directly over to the present problem. The position and the velocity of the Brownian particle become the position and momentum coordinates in phase space, and the solvent entropy becomes the reservoir entropy. With these, the stochastic, dissipative equations of motion for the sub-system are

$$\mathbf{q}(t+\tau) = \mathbf{q}(t) + \frac{\tau}{m}\mathbf{p}(t)$$

$$\mathbf{p}(t+\tau) = \mathbf{p}(t) - \tau\nabla_q U(\mathbf{q}) + \frac{|\tau|}{2}\Lambda\nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma}) + \tilde{\mathbf{R}}.$$

<div align="right">(5.50)</div>

The adiabatic parts of these are

$$\mathbf{q}^0(\tau \mid \boldsymbol{\Gamma}) = \mathbf{q} + \tau\frac{\partial\mathcal{H}(\boldsymbol{\Gamma})}{\partial\mathbf{p}} = \mathbf{q} + \frac{\tau}{m}\mathbf{p},$$

<div align="right">(5.51)</div>

and

$$\mathbf{p}^0(\tau \mid \boldsymbol{\Gamma}) = \mathbf{p} - \tau\frac{\partial\mathcal{H}(\boldsymbol{\Gamma})}{\partial\mathbf{q}} = \mathbf{p} - \tau\nabla_q U(\mathbf{q}).$$

<div align="right">(5.52)</div>

Since the reservoir entropy is $S_{\mathrm{r}}(\boldsymbol{\Gamma}) = -\mathcal{H}(\boldsymbol{\Gamma})/T$, the dissipative term in these equations of motion is

$$\frac{|\tau|}{2}\Lambda\nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma}) = \frac{-|\tau|}{2mT}\Lambda\mathbf{p}.$$

<div align="right">(5.53)</div>

The $3N \times 3N$ transport matrix Λ is symmetric, positive definite, and couples only the momentum components. The stochastic 'force' is Gaussian distributed, equation (3.89),

$$\wp(\tilde{\mathbf{R}}) = \frac{1}{[\mathrm{Det}\, 2\pi k_{\mathrm{B}} \mid \tau \mid \Lambda]^{1/2}}\mathrm{e}^{-\Lambda^{-1}:\tilde{\mathbf{R}}\tilde{\mathbf{R}}/2k_{\mathrm{B}}|\tau|}.$$

<div align="right">(5.54)</div>

That Λ appears in these last two equations is the form that the fluctuation–dissipation theorem takes for phase space.

The equilibrium Maxwell–Boltzmann distribution is stationary under these equations of motion for any choice of the symmetric, positive definite, momentum transport matrix Λ. It is simplest, however, to assume that different components of the momenta are uncoupled, which means that Λ is diagonal. In the simplest case that the particles are identical, it is proportional to the identity matrix, $\Lambda = \lambda\mathrm{I}$. Again, the Maxwell–Boltzmann distribution is stationary for any choice of the scalar λ. (The constant $\lambda/2T$ may be called the friction or drag coefficient.)

In this simplest case the stochastic, dissipative equations of motion for the canonical equilibrium system are

$$\mathbf{q}(t+\tau) = \mathbf{q}(t) + \frac{\tau}{m}\mathbf{p}(t)$$

$$\mathbf{p}(t+\tau) = \mathbf{p}(t) - \tau\nabla_q U(\mathbf{q}) - \frac{|\tau|\lambda}{2mT}\mathbf{p} + \tilde{\mathbf{R}},$$

<div align="right">(5.55)</div>

with

$$\wp(\tilde{\mathbf{R}}) = \frac{1}{[2\pi\lambda k_{\mathrm{B}} \mid \tau \mid]^{3N/2}}\mathrm{e}^{-\tilde{\mathbf{R}}\cdot\tilde{\mathbf{R}}/2\lambda k_{\mathrm{B}}|\tau|}.$$

The physical interpretation of these stochastic, dissipative equations of motion is illuminating. The dissipative term, $\hat{\tau}\lambda\nabla_p S_\mathrm{r}$, drives the system up the reservoir entropy gradient, which is to say toward sub-system microstates of higher reservoir entropy. The irreversible factor $\hat{\tau}$ means that going forward in time, $\hat{\tau} = +1$, this term determines that the destination of the transition will have higher reservoir entropy than the present state. If one operates the equations of motion backward to predict where the system came from, $\hat{\tau} = -1$, this term determines that the previous state had higher reservoir entropy than the present state.

The stochastic term $\tilde{\mathbf{R}}$ randomizes the transition without regard to the reservoir entropy. It is, however, sensitive to the sub-system entropy in that there are more sub-system phase space points $\mathbf{\Gamma}$ at higher sub-system energy $\mathcal{H}(\mathbf{\Gamma})$ than at lower, because the sub-system entropy $S_\mathrm{s}(E, N, V)$ is a monotonic increasing function of energy. Since the stochastic term $\tilde{\mathbf{R}}$ chooses destination phase space points uniformly at random, it is more likely to increase the sub-system energy than to decrease it. This stochastic term on average acts like an irreversible deterministic driving force toward macrostates of higher sub-system energy.

There is a competition between the dissipative and the stochastic term. The former drives the sub-system to states of low sub-system energy, because these correspond to high reservoir entropy, $S_\mathrm{r}(\mathbf{\Gamma}) = -\mathcal{H}(\mathbf{\Gamma})/T$. The latter drives the sub-system to states of high sub-system energy, which is to say high sub-system entropy, $S_\mathrm{s}(\mathcal{H}(\mathbf{\Gamma}), N, V)$. These two competing forces balance when the current phase space point belongs to a sub-system energy macrostate corresponding to a sub-system temperature that equals that of the reservoir (see figure 5.4).

Figure 5.4. The reservoir entropy of a sub-system phase space point (full line) and the sub-system entropy of an energy macrostate (dashed curve) as a function of sub-system energy. The circles signify the point at which the tangents are equal and opposite.
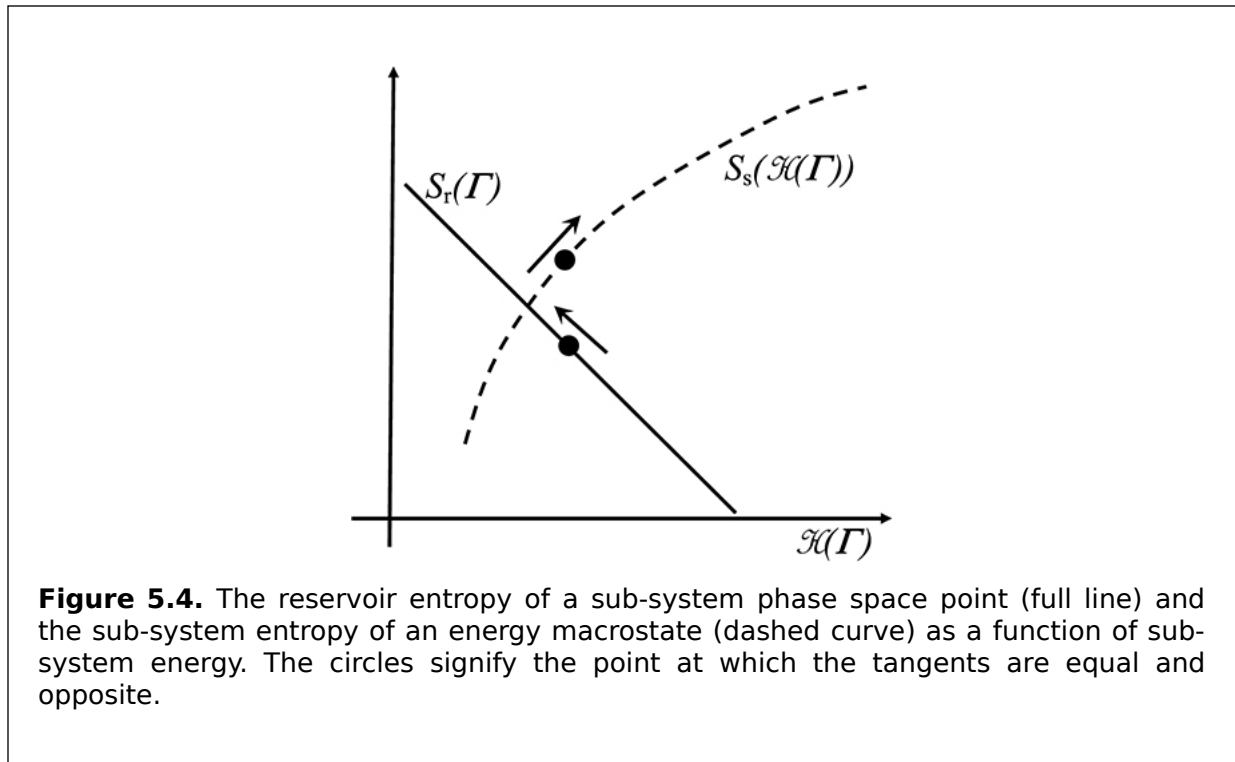
It is worthwhile to reprise the justification for these stochastic, dissipative equations of motions in phase space. It is more or less the same as was given for the Langevin equation

for Brownian motion, equation (4.21).

The position coordinate must evolve adiabatically over an infinitesimal time interval $|\tau|$. It is only over a sequence of consecutive intervals that forces from the sub-system particles and from the reservoir can indirectly effect the position evolution via their direct effect on the momentum evolution. This means that the position is a 'slave' to the momentum. It also means that only the momentum coordinate is to be treated statistically, which requires only the pure parity second entropy theory, equation (3.87).

For the present evolution of phase space over an infinitesimal time interval, the adiabatic evolution of the momentum appears explicitly. This is in contrast to the second entropy treatment of a fluctuation, equation (3.87). Although that case was for an isolated sub-system in which the evolution is purely adiabatic, because the time interval was long on molecular time scales, the adiabatic contributions to it had to be treated *in toto* rather than individually, which is to say statistically rather than explicitly. In the present case of phase space and an infinitesimal time interval, the forces due to the particles of the sub-system must appear explicitly in addition to the statistical forces from the reservoir.

The solvent forces acting on the Brownian particle in equation (4.21) are replaced here by the reservoir forces acting on the sub-system. The statistical treatment of the latter is identical to the pure parity second entropy analysis: the reservoir-induced transitions are treated as Gaussian fluctuations, with the coefficients being expanded for small time intervals in accord with general rules for the time symmetry of the transition in an equilibrium system. The result is that in the stochastic, dissipative equations of motion, the adiabatic forces are reversible, $\propto \tau$, whereas the dissipative and stochastic forces are irreversible, $\propto |\tau|$.

There are three reasons why it is valid to apply the second entropy analysis over infinitesimal time scales. First, the present momentum macrostate is the end point of a continuous evolution, and there is no need for an inertial period for it to become organized. Second, the relaxation time for individual particle momenta can be expected to be relatively short compared to that for a macroscopic fluctuation comprising Avogadro's number of particles. And third, any term linearly proportional to $|\tau_{\mathrm{macro}}|$ can be broken down into a series of infinitesimal transitions each proportional to $|\tau_{\mathrm{molec}}| \ll |\tau_{\mathrm{macro}}|$, and over the longer term these in series return the original expression.

It is important to note that these stochastic, dissipative equations of motions in phase space are justified by the second entropy analysis, and only by it. They cannot be justified by the Langevin equation for Brownian motion because the hydrodynamic drag term that is invoked therein has no justification on molecular length scales. The point is that the second entropy analysis shows that the dissipative force $\hat{\tau}\lambda\nabla_p S_{\mathrm{r}}$ is a direct consequence of the time symmetry of the second entropy. The hydrodynamic drag force is an effect, not a cause, of the second entropy dissipative force. It would be wrong to interpret the transport scalar λ as the product of the radius of the molecule and the viscosity of the sub-system (i.e. Stoke's drag coefficient) since it is not determined by hydrodynamics. All that is essential is that the fluctuation–dissipation theorem be satisfied: the magnitude of the dissipation must be equal to the strength of the fluctuations. Since both are linearly proportional to λ, the actual value of λ is immaterial (except, of course, that the reservoir forces should be negligible compared to the adiabatic forces.)

5.4.2 Reversibility of the stochastic, dissipative equations of motion

As was shown in section 5.1.2.2, the adiabatic equations of motion are time reversible,

$$\mathbf{\Gamma}_2 = \mathbf{\Gamma}^0(t \mid \mathbf{\Gamma}_1) \Leftrightarrow \mathbf{\Gamma}_1 = \mathbf{\Gamma}^0(-t \mid \mathbf{\Gamma}_2).$$

(5.57)

This may be called temporal reversibility. It follows because a unique adiabatic trajectory passes through each phase space point. The adiabatic equations of motion also

display what might be called conjugate reversibility,

$$\boldsymbol{\Gamma}_2 = \boldsymbol{\Gamma}^0(t \mid \boldsymbol{\Gamma}_1) \Leftrightarrow \boldsymbol{\Gamma}_1^\dagger = \boldsymbol{\Gamma}^0(t \mid \boldsymbol{\Gamma}_2^\dagger).$$

Recall that each phase space point $\boldsymbol{\Gamma} = \{\mathbf{q}, \mathbf{p}\}$ has conjugate point with all the momenta reversed, $\boldsymbol{\Gamma}^\dagger = \{\mathbf{q}, -\mathbf{p}\}$. The combination of temporal reversibility and conjugate reversibility can be called microscopic reversibility,

$$\boldsymbol{\Gamma}_2 = \boldsymbol{\Gamma}^0(t \mid \boldsymbol{\Gamma}_1) \Leftrightarrow \boldsymbol{\Gamma}_2^\dagger = \boldsymbol{\Gamma}^0(-t \mid \boldsymbol{\Gamma}_1^\dagger).$$

The distinction between conjugate reversibility and microscopic reversibility is usually overlooked because in the adiabatic case temporal reversibility is both obvious and trivial. For stochastic dissipative equations it is necessary to distinguish the two.

Microscopic reversibility says that if both the velocities and the time interval are reversed, then the end point of this transition is the conjugate of the end point of the original transition. The physical origin of microscopic reversibility is that reversing the velocities causes the particles to retrace their previous paths, and hence reversing both the velocities and time is equivalent to the original transition. Mathematically, for a single time step,

$$\Delta_t \dot{\boldsymbol{\Gamma}}_q^0(\boldsymbol{\Gamma}^\dagger) = \Delta_t \left\{ \nabla_{p^\dagger} \mathcal{K}(\mathbf{p}^\dagger) \right\}_q = -\Delta_t \{\nabla_p \mathcal{K}(\mathbf{p})\}_q = -\Delta_t \{\nabla_p \mathcal{K}(\mathbf{p})\}_q^\dagger$$

$$\Delta_t \dot{\boldsymbol{\Gamma}}_p^0(\boldsymbol{\Gamma}^\dagger) = \Delta_t \left\{ \nabla_{q^\dagger} U(\mathbf{q}^\dagger) \right\}_p = \Delta_t \{\nabla_q U(\mathbf{q})\}_p = -\Delta_t \{\nabla_q U(\mathbf{q})\}_p^\dagger.$$

These are a more explicit version of equation (5.13). One sees that the sign of the time step has been reversed in the conjugate system. Temporal, conjugate, and microscopic reversibility for an adiabatic transition are shown in table 5.1.

Table 5.1. Time symmetries of equations of motion.

	Adiabatic	Dissipative
	$\boldsymbol{\Gamma}_2 = \boldsymbol{\Gamma}^0(\Delta_t \mid \boldsymbol{\Gamma}_1)$	$\boldsymbol{\Gamma}_2 = \bar{\boldsymbol{\Gamma}}(\Delta_t \mid \boldsymbol{\Gamma}_1)$
	$= \boldsymbol{\Gamma}_1 + \Delta_t \dot{\boldsymbol{\Gamma}}^0(\boldsymbol{\Gamma})$	$= \boldsymbol{\Gamma}_1 + \Delta_t \dot{\boldsymbol{\Gamma}}^0(\boldsymbol{\Gamma}) + \frac{|\Delta_t|}{2} \lambda \nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma})$
Temporal	$\boldsymbol{\Gamma}_1 = \boldsymbol{\Gamma}^0(-\Delta_t \mid \boldsymbol{\Gamma}_2)$	$\boldsymbol{\Gamma}_3 = \bar{\boldsymbol{\Gamma}}(-\Delta_t \mid \boldsymbol{\Gamma}_2)$
	$= \boldsymbol{\Gamma}_2 - \Delta_t \dot{\boldsymbol{\Gamma}}^0(\boldsymbol{\Gamma})$	$= \boldsymbol{\Gamma}_2 - \Delta_t \dot{\boldsymbol{\Gamma}}^0(\boldsymbol{\Gamma}) + \frac{|\Delta_t|}{2} \lambda \nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma})$
Conjugate	$\boldsymbol{\Gamma}_1^\dagger = \boldsymbol{\Gamma}^0(\Delta_t \mid \boldsymbol{\Gamma}_2^\dagger)$	$\boldsymbol{\Gamma}_4^\dagger = \bar{\boldsymbol{\Gamma}}(\Delta_t \mid \boldsymbol{\Gamma}_2)^\dagger$
	$= \boldsymbol{\Gamma}_2^\dagger + \Delta_t \dot{\boldsymbol{\Gamma}}^0(\boldsymbol{\Gamma}^\dagger)$	$= \boldsymbol{\Gamma}_2^\dagger + \Delta_t \dot{\boldsymbol{\Gamma}}^0(\boldsymbol{\Gamma}^\dagger) + \frac{|\Delta_t|}{2} \lambda \nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma}^\dagger)$
Microscopic	$\boldsymbol{\Gamma}_2^\dagger = \boldsymbol{\Gamma}^0(-\Delta_t \mid \boldsymbol{\Gamma}_1^\dagger)$	$\boldsymbol{\Gamma}_2^\dagger = \bar{\boldsymbol{\Gamma}}(-\Delta_t \mid \boldsymbol{\Gamma}_1)^\dagger$
	$= \boldsymbol{\Gamma}_1^\dagger - \Delta_t \dot{\boldsymbol{\Gamma}}^0(\boldsymbol{\Gamma}^\dagger)$	$= \boldsymbol{\Gamma}_1^\dagger - \Delta_t \dot{\boldsymbol{\Gamma}}^0(\boldsymbol{\Gamma}^\dagger) + \frac{|\Delta_t|}{2} \lambda \nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma}^\dagger)$

The stochastic part of the stochastic, dissipative equations of motion is dependent on the magnitude of the time interval, and is as likely positive as negative. Hence it displays each of temporal, conjugate, and microscopic reversibility.

The dissipative part depends on the magnitude of the time interval, and hence it is irreversible in a temporal sense. The difference upon temporarily reversing a dissipative transition is

$$\bar{\Gamma}(-\Delta_t \mid \bar{\Gamma}(\Delta_t \mid \Gamma_1)) - \Gamma_1 = \bar{\Gamma}(\Delta_t \mid \Gamma_1) - \Delta_t \dot{\Gamma}^0(\Gamma) + \frac{\mid \Delta_t \mid}{2}\lambda \nabla_p S_{\mathrm{r}}(\Gamma) - \Gamma_1$$

$$= \mid \Delta_t \mid \lambda \nabla_p S_{\mathrm{r}}(\Gamma) + \mathcal{O}(\Delta_t^2).$$

(5.6
1)

The dissipative force increases the entropy at both stages of the cyclic transition, which causes the stochastic, dissipative equations of motion to be irreversible in a temporal sense.

Similarly, reversing the momenta at the end of the first transition and continuing forward another time step gives a difference of

$$\bar{\Gamma}(\Delta_t \mid \bar{\Gamma}(\Delta_t \mid \Gamma_1)^\dagger) - \Gamma_1^\dagger = \bar{\Gamma}(\Delta_t \mid \Gamma_1)^\dagger + \Delta_t \dot{\Gamma}^0(\Gamma^\dagger) + \frac{\mid \Delta_t \mid}{2}\lambda \left\{ \nabla_p S_{\mathrm{r}}(\Gamma^\dagger) \right\}_p - \Gamma_1^\dagger$$

$$= - \mid \Delta_t \mid \lambda \nabla_p S_{\mathrm{r}}(\Gamma) + \mathcal{O}(\Delta_t^2).$$

(5.6
2)

The second equality follows because for an equilibrium system $S_{\mathrm{r}}(\Gamma)$ is an even function of the momenta, which means that the dissipative force, $\propto \nabla_p S_{\mathrm{r}}(\Gamma)$, is an odd function of momenta. One sees that over this particular cycle the entropy has decreased, and that the dissipative equations of motion are irreversible in a conjugate sense.

Figure 5.5 shows several dissipative trajectories, both forward and backward in time from their initial point. It can be seen that a point in phase space does not have a unique trajectory through it. This is one reason for the temporal and conjugate irreversibility just mentioned. Although in figure 5.5 the square lies on the most likely backward trajectory from the circle, the circle does not lie on the most likely forward trajectory from the square. This is another reason for the temporal irreversibility of a dissipative trajectory. The discontinuity in the derivative of a trajectory at its initial point was discussed in connection with figure 4.2, where it was attributed to the fact that the dissipative contribution is proportional to the absolute value of the time interval. The essential physical reason for the kink in the trajectory at the initial point, and for the irreversibility, is that most likely the sub-system will head *toward* an equilibrium state, and most likely the sub-system came *from* an equilibrium state. The stochastic contributions will perturb the most likely trajectories shown in figure 5.5, increasingly so as time progresses or regresses from the initial time (cf figure 4.2), but the qualitative picture remains as sketched.

Figure 5.5. Six dissipative trajectories. The circle on the solid curve marks the initial point at time t_1 of a forward and a backward trajectory. The triangle on the dashed
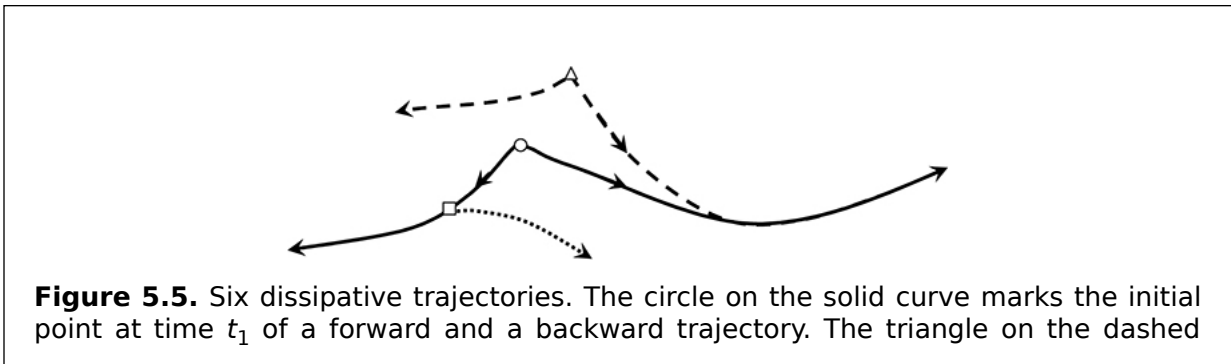
curve at t_2 marks the initial point of a forward and a backward trajectory. The square on the dotted curve at $t_3 < t_1$ marks the initial point of a forward and a backward (obscured) trajectory.

As was mentioned in the discussion of figure 4.2 and equation (4.22), the discontinuity in the derivative of the trajectory means that the forward or backward time derivative of a function on the trajectory has to be specified. The irreversibility embodied by the dissipative and stochastic terms means that the forward and backward time derivatives are not equal to each other.

Microscopic reversibility for the dissipative transition is equivalent to

$$
\begin{aligned}
\bar{\mathbf{\Gamma}}(-\Delta_t \mid \mathbf{\Gamma}_1^\dagger) &- \bar{\mathbf{\Gamma}}(\Delta_t \mid \mathbf{\Gamma}_1)^\dagger \\
&= \mathbf{\Gamma}_1^\dagger - \Delta_t \dot{\mathbf{\Gamma}}^0(\mathbf{\Gamma}^\dagger) + \frac{\mid \Delta_t \mid}{2} \lambda \left\{ \nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}^\dagger) \right\}_p \\
&\quad - \mathbf{\Gamma}_1^\dagger - \Delta_t \dot{\mathbf{\Gamma}}^0(\mathbf{\Gamma})^\dagger - \frac{\mid \Delta_t \mid}{2} \lambda \{ \nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}) \}_p^\dagger \\
&= \frac{- \mid \Delta_t \mid}{2} \lambda \{ \nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}) \}_p + \frac{\mid \Delta_t \mid}{2} \lambda \{ \nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}) \}_p + \mathcal{O}(\Delta_t^2) \\
&= 0.
\end{aligned}
$$

(5.63)

This says that the original transition forward in time is equivalent to reversing the momenta, making a transition backward in time, and reversing the momenta again. Temporal and conjugate irreversibility, and microscopic reversibility for a dissipative transition are shown in table 5.1.

For the stochastic, dissipative equations of motion for the canonical equilibrium system equation (5.55), one can exhibit microscopic reversibility explicitly. The position evolution in a time step is purely adiabatic and hence microscopic reversible. The most likely momentum evolution is

$$
\mathbf{p}_2 \equiv \bar{\mathbf{p}}(\Delta_t \mid \mathbf{\Gamma}_1) = \mathbf{p}_1 - \Delta_t \nabla_q U(\mathbf{q}_1) - \frac{\mid \Delta_t \mid \lambda}{2mT} \mathbf{p}_1.
$$

(5.64)

Negating both sides this can be written as

$$
\begin{aligned}
\mathbf{p}_2^\dagger &= \mathbf{p}_1^\dagger + \Delta_t \nabla_q U(\mathbf{q}_1^\dagger) - \frac{\mid \Delta_t \mid \lambda}{2mT} \mathbf{p}_1^\dagger \\
&= \bar{\mathbf{p}}(-\Delta_t \mid \mathbf{\Gamma}_1^\dagger).
\end{aligned}
$$

(5.65)

This says explicitly that the most likely part of the stochastic, dissipative equations of motion for the momenta for the canonical equilibrium system are microscopically reversible. (The dissipative equations for the position are purely adiabatic, and so they are automatically microscopically reversible.)

Microscopic reversibility for the dissipative equations of motion may in general be written as

$$
\mathbf{\Gamma}_2 = \bar{\mathbf{\Gamma}}(t \mid \mathbf{\Gamma}_1) \Leftrightarrow \mathbf{\Gamma}_2^\dagger = \bar{\mathbf{\Gamma}}(-t \mid \mathbf{\Gamma}_1^\dagger).
$$

This is identical to microscopic reversibility for adiabatic equations of motion, which was sketched in figure 5.2. Since the stochastic part of the transition is insensitive to the direction of time, and is as likely positive as negative, $\wp(\tilde{\mathbf{R}}) = \wp(\tilde{\mathbf{R}}^\dagger)$, the stochastic, dissipative equations of motion for an equilibrium system themselves are microscopically reversible

$$\mathbf{\Gamma}_2 = \mathbf{\Gamma}(t \mid \mathbf{\Gamma}_1) \Leftrightarrow \mathbf{\Gamma}_2^\dagger = \mathbf{\Gamma}(-t \mid \mathbf{\Gamma}_2^\dagger).$$

(5.66)

(5.67)

This is to be understood in a probabilistic sense: the conditional probability of the transition $\mathbf{\Gamma}_1 \xrightarrow{t} \mathbf{\Gamma}_2$ is equal to that of $\mathbf{\Gamma}_1^\dagger \xrightarrow{-t} \mathbf{\Gamma}_2^\dagger$. Since for an equilibrium system the probability of a point in phase space is equal to that of the conjugate point, $\wp(\mathbf{\Gamma}_1) = \wp(\mathbf{\Gamma}_1^\dagger)$, the unconditional probability of the forward and the conjugate backward transition must also be equal.

5.4.3 Stationarity of the Maxwell–Boltzmann distribution

The Fokker–Planck equation for the evolution of an arbitrary phase space probability distribution under stochastic, dissipative equations of motion was derived above as equation (4.36). Replacing $X \Rightarrow \mathbf{\Gamma}$ it is

$$\frac{\partial \wp(\mathbf{\Gamma}, t)}{\partial t} = -[\nabla \cdot \dot{\mathbf{\Gamma}}^{\text{det}}] \wp(\mathbf{\Gamma}, t) - \dot{\mathbf{\Gamma}}^{\text{det}} \cdot \nabla \wp(\mathbf{\Gamma}, t) + \frac{\hat{\tau} k_\text{B}}{2} \Lambda : \nabla \nabla \wp(\mathbf{\Gamma}, t)$$

$$= \frac{3N\hat{\tau}\lambda}{2mT} \wp(\mathbf{\Gamma}, t) - \dot{\mathbf{\Gamma}}^{\text{det}} \cdot \nabla \wp(\mathbf{\Gamma}, t) + \frac{\hat{\tau} k_\text{B}}{2} \lambda \nabla_p^2 \wp(\mathbf{\Gamma}, t).$$

(5.68)

Since phase space is incompressible under the adiabatic equations of motion, $\nabla \cdot \dot{\mathbf{\Gamma}}^{\text{det}} = -\hat{\tau}\lambda \nabla_p \cdot \mathbf{p}/2mT$.

With the Maxwell–Boltzmann distribution, $\wp_{\text{MB}}(\mathbf{\Gamma}) = Z^{-1} \mathrm{e}^{-\beta \mathcal{H}(\mathbf{\Gamma})}$, and the fact that energy is conserved on an adiabatic trajectory, $\dot{\mathbf{\Gamma}}^0 \cdot \nabla \mathcal{H}(\mathbf{\Gamma}) = 0$, the right-hand side of this is

$$\begin{aligned}
\text{RHS} &= \frac{3N\hat{\tau}\lambda}{2mT} \wp_{\text{MB}}(\mathbf{\Gamma}) + \frac{\hat{\tau}\lambda}{2mT} \mathbf{p} \cdot \nabla_p [-\beta \mathcal{H}(\mathbf{\Gamma})] \wp_{\text{MB}}(\mathbf{\Gamma}) \\
&\quad + \frac{\hat{\tau} k_\text{B}\lambda}{2} \left[-\beta \nabla_p^2 \mathcal{H}(\mathbf{\Gamma}) + \beta^2 (\nabla_p \mathcal{H}(\mathbf{\Gamma}))^2 \right] \wp_{\text{MB}}(\mathbf{\Gamma}) \\
&= \left\{ \frac{3N\hat{\tau}\lambda}{2mT} - \frac{\hat{\tau}\lambda}{2m^2 k_\text{B} T^2} \mathbf{p} \cdot \mathbf{p} - \frac{\hat{\tau}3N\lambda}{2mT} + \frac{\hat{\tau}\lambda}{2m^2 k_\text{B} T^2} \mathbf{p} \cdot \mathbf{p} \right\} \wp_{\text{MB}}(\mathbf{\Gamma}) \\
&= 0.
\end{aligned}$$

(5.69)

This shows that the Maxwell–Boltzmann distribution is stationary under the stochastic, dissipative equations of motion. Of course these equations of motion simply implement the second entropy transition probability for the canonical equilibrium system, and so they are consistent with the Maxwell–Boltzmann distribution by design.

5.4.4 Grand canonical system

A grand canonical system can exchange particles and energy with a reservoir of constant chemical potential μ and temperature T. As was discussed in section 2.4.1, the total constrained entropy for an energy-number macrostate is $S_{\text{total}}(N, E \mid \mu, V, T) = S_s(E, V, N) - E/T + \mu N/T$. The reservoir entropy for a point in the sub-system phase space $\boldsymbol{\Gamma}$ is therefore

$$S_{\text{r}}(\boldsymbol{\Gamma}, N \mid \mu, V, T) = \frac{-\mathcal{H}(\boldsymbol{\Gamma})}{T} + \frac{\mu}{T} N.$$

(5.70)

In the derivation of the Maxwell–Boltzmann distribution above, the sub-system entropy of a phase space point was given as equation (5.46), $S_s(\boldsymbol{\Gamma}) = Nk_{\text{B}} - Nk_{\text{B}} \ln(Nh^3)$. With this the total entropy for the microstate $\boldsymbol{\Gamma}$ is

$$S_{\text{total}}(\boldsymbol{\Gamma}, N \mid \mu, V, T) = S_s(\boldsymbol{\Gamma}) + S_{\text{r}}(\boldsymbol{\Gamma}, N \mid \mu, V, T)$$
$$= Nk_{\text{B}} - Nk_{\text{B}} \ln(Nh^3) - \frac{\mathcal{H}(\boldsymbol{\Gamma})}{T} + \frac{\mu}{T} N.$$

(5.71)

Hence the probability density for the grand canonical system can be written explicitly either in terms of the reservoir entropy,

$$\wp(\boldsymbol{\Gamma}, N \mid \mu, V, T) = \frac{\mathrm{e}^{S_{\text{r}}(\boldsymbol{\Gamma}, N \mid \mu, V, T)/k_{\text{B}}}}{N! h^{3N} \Xi(\mu, V, T)} = \frac{\mathrm{e}^{-\beta \mathcal{H}(\boldsymbol{\Gamma}) + \beta \mu N}}{N! h^{3N} \Xi(\mu, V, T)},$$

(5.72)

or else in terms of the total entropy,

$$\wp(\boldsymbol{\Gamma}, N \mid \mu, V, T) = \frac{\mathrm{e}^{S_{\text{total}}(\boldsymbol{\Gamma}, N \mid \mu, V, T)/k_{\text{B}}}}{\Xi(\mu, V, T)} = \frac{\mathrm{e}^{N - N \ln(Nh^3) - \beta \mathcal{H}(\boldsymbol{\Gamma}) + \beta \mu N}}{\Xi(\mu, V, T)}.$$

(5.73)

In the thermodynamic limit these are the same, although the former is the more conventional expression. In these the grand partition function is

$$\Xi(\mu, V, T) = \sum_{N=0}^{\infty} \frac{\mathrm{e}^{\beta \mu N}}{N! h^{3N}} \int \mathrm{d}\boldsymbol{\Gamma} \ \mathrm{e}^{-\beta \mathcal{H}(\boldsymbol{\Gamma})}.$$

(5.74)

Sometimes the fugacity, $z \equiv \mathrm{e}^{\beta \mu}$, is used instead of the chemical potential as the independent variable of the grand canonical system.

5.4.4.1 Grand canonical stochastic molecular dynamics

Let the number of particles in the sub-system be described by a positive real number $\mathcal{N}$, with the integer part of this $N = \lfloor \mathcal{N} \rfloor$ being the number of fully coupled particles, and $\xi = \mathcal{N} - N$ being the extent of coupling of the $(N + 1)$st particle. The Hamiltonian of the system is

$$\mathcal{H}(\boldsymbol{\Gamma}^{\mathcal{N}}) = \mathcal{H}(\boldsymbol{\Gamma}^N) + \xi \frac{p_{N+1}^2}{2m} + \xi U(\mathbf{q}^N, \mathbf{q}_{N+1}).$$

(5.75)

Here a linear coupling has been chosen for the partial particle, which is simplest but not essential. The reservoir entropy is

$$S_r(\mathbf{\Gamma}^{\mathcal{N}}, \mathcal{N} \mid \mu, V, T) = \frac{-1}{T} \mathcal{H}(\mathbf{\Gamma}^{\mathcal{N}}) + \frac{\mu}{T} \mathcal{N}.$$

(5.76)

The internal entropy of a sub-system phase point is

$$S_s(\mathbf{\Gamma}) = \mathcal{N} k_B - \mathcal{N} k_B \ln(\mathcal{N} h^3),$$

(5.77)

and so the total entropy is

$$\begin{aligned}
S_{tot}(\mathbf{\Gamma}, \mathcal{N} \mid \mu, V, T) &= S_s(\mathbf{\Gamma}) + S_r(\mathbf{\Gamma}, \mathcal{N} \mid \mu, V, T) \\
&= \mathcal{N} k_B - \mathcal{N} k_B \ln(\mathcal{N} h^3) - \frac{1}{T} \mathcal{H}(\mathbf{\Gamma}^{\mathcal{N}}) + \frac{\mu}{T} \mathcal{N}.
\end{aligned}$$

(5.78)

The equations of motion for the particles are

$$\mathbf{q}^N(t + \Delta_t) = \mathbf{q}^N(t) + \frac{\Delta_t}{m} \mathbf{p}^N(t)$$

$$\mathbf{q}_{N+1}(t + \Delta_t) = \mathbf{q}_{N+1}(t) + \frac{\Delta_t \xi}{m} \mathbf{p}_{N+1}(t)$$

$$\mathbf{p}^{N+1}(t + \Delta_t) = \mathbf{p}^{N+1}(t) - \Delta_t \frac{\partial \mathcal{H}(\mathbf{\Gamma}^{\mathcal{N}})}{\partial \mathbf{q}^{N+1}} + \frac{|\Delta_t|}{2} \lambda \frac{\partial S_{tot}(\mathbf{\Gamma}^{\mathcal{N}})}{\partial \mathbf{p}^{N+1}} + \tilde{R}^{N+1}.$$

(5.79)

Here $-\partial \mathcal{H}(\mathbf{\Gamma}^{\mathcal{N}})/\partial \mathbf{q}^{N+1}$ is the force for *N* fully coupled particles and the (*N* + 1)st partially coupled particle, and the gradient of the total entropy is

$$\frac{\partial S_{tot}(\mathbf{\Gamma}^{\mathcal{N}}, \mathcal{N} \mid \mu, V, T)}{\partial \mathbf{p}^{N+1}} = \frac{-1}{mT} \{\mathbf{p}^N, \xi \mathbf{p}_{N+1}\}.$$

(5.80)

The transport matrix has been chosen to be diagonal with all elements equal to λ. The probability distribution for the stochastic force is

$$\wp(\tilde{R}^{N+1}) = \frac{1}{Z} e^{-\tilde{R}^{N+1} \cdot \tilde{R}^{N+1} / 2\lambda k_B |\Delta_t|}.$$

(5.81)

The equation of motion for the number is

$$\begin{aligned}
\mathcal{N}(t + \Delta_t) &= \mathcal{N}(t) + \frac{|\Delta_t|}{2} \lambda\prime \frac{\partial S_{tot}(\mathbf{\Gamma}^{\mathcal{N}}(t), \mathcal{N} \mid \mu, V, T)}{\partial \mathcal{N}} + \tilde{R}\prime \\
&= \mathcal{N}(t) - \frac{|\Delta_t|}{2T} \lambda\prime \left\{ \frac{p_{N+1}^2}{2m} + U(\mathbf{q}^N, \mathbf{q}_{N+1}) - \mu + k_B T \ln(\mathcal{N} h^3) \right\} + \tilde{R}\prime.
\end{aligned}$$

(5.82)

Notice how the sub-system entropy contributes to this. The probability distribution for the stochastic force in this case is

$$\wp(\tilde{R}\prime) = \frac{1}{Z\prime} e^{-\tilde{R}\prime^2/2\lambda\prime k_{\mathrm{B}}|\Delta_t|}.$$

The positive scalar $\lambda\prime$ is the 'drag' or 'friction' coefficient for number.

When $\xi(t + \Delta_t)$ exceeds 1, then the $(N + 1)$st particle becomes an ordinary particle in the sub-system, and a new partially coupled particle is created, coupled to the extent $\xi(t + \Delta_t) - 1$. This is randomly placed in the sub-system phase space according to the probability distribution $\propto \exp - \beta(\xi - 1)[p_{N+2}^2/2\,m + U(\mathbf{q}^{N+1}, \mathbf{q}_{N+2})]$. (The momentum can be chosen using the Gaussian; the position using several trial steps of the Metropolis algorithm, for example.) When $\xi(t + \Delta_t)$ drops below 0, the existing partially coupled particle is eliminated, and one of the remaining N fully coupled particles is randomly chosen as the new partially coupled particle, and it is coupled to the extent $\xi(t + \Delta_t) + 1$.

5.4.5 Isobaric system

An isobaric system can exchange volume and energy with a reservoir of constant pressure p and temperature T. As was discussed in section 2.4.1, the total constrained entropy for an energy–volume macrostate is $S_{\mathrm{total}}(E, V \mid N, p, T) = S_{\mathrm{s}}(E, V, N) - E/T - pV/T$. Hence the reservoir entropy for the sub-system phase space $\boldsymbol{\Gamma}$ is $S_{\mathrm{r}}(\boldsymbol{\Gamma}, V \mid N, p, T) = -\mathcal{H}(\boldsymbol{\Gamma})/T - pV/T$. The sub-system entropy for $\boldsymbol{\Gamma}$ is constant, $S_{\mathrm{s}}(\boldsymbol{\Gamma}) = Nk_{\mathrm{B}} - Nk_{\mathrm{B}}\,\ln(Nh^3)$. Hence the total entropy is

$$S_{\mathrm{total}}(\boldsymbol{\Gamma}, V \mid N, p, T) = Nk_{\mathrm{B}} - Nk_{\mathrm{B}}\,\ln(Nh^3) - \frac{\mathcal{H}(\boldsymbol{\Gamma})}{T} - \frac{p}{T}V.$$

This gives the probability density for the grand canonical system as

$$\wp(\boldsymbol{\Gamma}, V \mid N, p, T) = \frac{e^{S_{\mathrm{r}}(\boldsymbol{\Gamma}, V \mid N, p, T)/k_{\mathrm{B}}}}{N!h^{3N}\Delta_V Z(N, p, T)} = \frac{e^{-\beta\mathcal{H}(\boldsymbol{\Gamma})-\beta pV}}{N!h^{3N}\Delta_V Z(N, p, T)}.$$

Here the isobaric partition function is

$$Z(N, p, T) = \frac{1}{N!h^{3N}\Delta_V} \int dV\, e^{-\beta pV} \int d\boldsymbol{\Gamma}\, e^{-\beta\mathcal{H}(\boldsymbol{\Gamma})}.$$

The volume width Δ_V is used to make the partition function dimensionless and its actual value has no physical consequences.

5.4.5.1 Isobaric stochastic molecular dynamics

Let the sub-system volume be cubic, $V = L^3$, and use the edge length to scale the positions, $\mathbf{q} = L\mathbf{x}$. The potential energy may be written $U(\mathbf{q}) = U(\mathbf{x}; L)$. The reservoir entropy is

$$S_{\mathrm{r}}(\mathbf{p}, \mathbf{x}, L \mid N, p, T) = \frac{-1}{T}\mathcal{H}(\mathbf{p}, \mathbf{x}; L) - \frac{p}{T}L^3.$$

The internal entropy of a sub-system phase point is $S_s(\mathbf{\Gamma}) = Nk_B - Nk_B \ln(Nh^3)$, and so the total entropy is

$$S_{\text{tot}}(\mathbf{p}, \mathbf{x}, L \mid N, p, T) = S_s(\mathbf{\Gamma}) + S_r(\mathbf{p}, \mathbf{x}, L \mid N, p, T)$$
$$= Nk_B - Nk_B \ln(Nh^3) - \frac{1}{T}\mathcal{H}(\mathbf{p}, \mathbf{x}; L) - \frac{p}{T}L^3.$$

(5.88)

The equations of motion for the particles are

$$\mathbf{x}(t + \Delta_t) = \mathbf{x}(t) + \frac{\Delta_t}{mL}\mathbf{p}(t)$$
$$\mathbf{p}(t + \Delta_t) = \mathbf{p}(t) - \Delta_t\frac{\partial U(\mathbf{q})}{\partial \mathbf{q}} + \frac{|\Delta_t|}{2}\lambda\frac{\partial S_{\text{tot}}(\mathbf{p}, \mathbf{x}, L \mid N, p, T)}{\partial \mathbf{p}} + \tilde{R}.$$

(5.89)

Here the adiabatic force is $-\partial U(\mathbf{q})/\partial\mathbf{q} = -\partial U(\mathbf{x}; L)/L\partial\mathbf{x}$, and the gradient of the reservoir entropy is

$$\frac{\partial S_{\text{tot}}(\mathbf{p}, \mathbf{x}, L \mid N, p, T)}{\partial \mathbf{p}} = \frac{-1}{mT}\mathbf{p}.$$

(5.90)

The $N \times N$ transport matrix has been chosen to be diagonal with all elements equal to λ. The probability distribution for the stochastic force is

$$\wp(\tilde{R}) = \frac{1}{Z}e^{-\tilde{R}^2/2\lambda k_B|\Delta_t|}.$$

(5.91)

The equation of motion for the edge length is

$$L(t + \Delta_t) = L(t) + \frac{|\Delta_t|\lambda\prime}{2}\frac{\partial S_{\text{tot}}(\mathbf{p}, \mathbf{x}, L \mid N, p, T)}{\partial L} + \tilde{R}\prime$$
$$= L(t) - \frac{|\Delta_t|\lambda\prime}{2TL}\left\{\mathbf{q}\cdot\frac{\partial U(\mathbf{q})}{\partial\mathbf{q}} + 3pL^3\right\} + \tilde{R}\prime.$$

(5.92)

Again the adiabatic force is $-\partial U(\mathbf{q})/\partial\mathbf{q} = -\partial U(\mathbf{x}; L)/L\partial\mathbf{x}$. The probability distribution for the stochastic force is

$$\wp(\tilde{R}\prime) = \frac{1}{Z\prime}e^{-\tilde{R}\prime^2/2\lambda\prime k_B|\Delta_t|}.$$

(5.93)

5.5 Equilibrium phase space averages

5.5.1 Probability densities and averages

The stochastic, dissipative equations of motion derived above may be used to generate a trajectory through the sub-system phase space, $\mathbf{\Gamma}(t \mid \mathbf{\Gamma}_0)$. The canonical equilibrium statistical average of a phase function $f(\mathbf{\Gamma})$ can then be written as a time average,

$$\langle f \rangle_{N,V,T} = \frac{1}{t} \int_0^t \mathrm{d}t\prime \; f(\mathbf{\Gamma}(t\prime | \; \mathbf{\Gamma}_0)).$$

For a long enough trajectory, the results can be expected to be independent of the starting position $\mathbf{\Gamma}_0$.

The statistical average can also be written as an integral over phase space,

$$\langle f \rangle_{N,V,T} = \int \mathrm{d}\mathbf{\Gamma} \; \wp(\mathbf{\Gamma} \mid N,V,T) f(\mathbf{\Gamma}),$$

where the Maxwell–Boltzmann probability density is

$$\wp(\mathbf{\Gamma} \mid N,V,T) = \frac{\mathrm{e}^{-\beta \mathcal{H}(\mathbf{\Gamma})}}{N! h^{3N} Z(N,V,T)},$$

and the normalizing partition function is $Z(N,V,T) = (N! h^{3N})^{-1} \int \mathrm{d}\mathbf{\Gamma} \; \mathrm{e}^{-\beta \mathcal{H}(\mathbf{\Gamma})}$. Where there is no ambiguity the Maxwell–Boltzmann probability density will be written simply as $\wp(\mathbf{\Gamma})$.

Results will be presented in this section for averages of various phase functions. We shall focus on the formula that gives these as an integral over phase space. In general the average is the same for all systems (canonical, grand canonical, isobaric, etc). Fluctuations about the average vary from system to system (e.g. the fluctuations in number are zero for a canonical system, but nonzero for a grand canonical system). For brevity most of the following results will be given for a canonical equilibrium system.

5.5.2 Helmholtz free energy and entropy

Since the partition function is the integral over phase space of the weight of phase space points, it is the total weight of the total system. Hence its logarithm is the total entropy of the total system,

$$\begin{aligned} S_{\text{total}}(N,V,T) &= k_{\text{B}} \ln Z(N,V,T) \\ &\equiv -F(N,V,T)/T \\ &\approx -\bar{F}(N,V,T)/T \\ &\equiv -F(\bar{E}(N,V,T) \mid N,V,T)/T. \end{aligned}$$

Here $F(N,V,T)$ is the statistical mechanical and $\bar{F}(N,V,T)$ is the thermodynamic Helmholtz free energy (see chapter 2). Also, $\bar{E}(N,V,T)$ is the most likely sub-system energy, which is equal to the average sub-system energy.

Recall from chapter 2 that the constrained total entropy is $S_{\text{total}}(E \mid N,V,T) = S_{\text{s}}(E,N,V) - E/T$, and therefore that the Helmholtz free energy is

$$F(\bar{E} \mid N,V,T) = \bar{E} - T S_{\text{s}}(\bar{E},N,V).$$

Re-arranging these and equating the most likely energy to the average energy, the entropy of the sub-system may be written as

$$S_{\mathrm{s}}(\bar{E}, N, V) = \frac{\bar{E}}{T} + S_{\mathrm{total}}(N, V, T)$$

$$= k_{\mathrm{B}} \int d\mathbf{\Gamma}\ \wp(\mathbf{\Gamma})\{\beta\mathcal{H}(\mathbf{\Gamma}) + \ln\ Z(N, V, T)\}$$

$$= -k_{\mathrm{B}} \int d\mathbf{\Gamma}\ \wp(\mathbf{\Gamma})\ln\left[h^{3N}N!\wp(\mathbf{\Gamma})\right]$$

$$= -k_{\mathrm{B}} \int d\mathbf{\Gamma}\ \wp(\mathbf{\Gamma})\ln\ \wp(\mathbf{\Gamma}).$$

(5.99)

The third equality follows because the probability density is normalized to unity. The penultimate equality follows because $\wp(\mathbf{\Gamma}) = e^{-\beta\mathcal{H}(\mathbf{\Gamma})}/h^{3N}N!Z(N, V, T)$. The final equality neglects an immaterial constant, $-k_{\mathrm{B}} \ln\ h^{3N}N!$, which can be interpreted as the internal entropy of each phase space point. One sees from this that the so-called Gibbs or Shannon entropy is only the sub-system entropy, which is just part of the total entropy (cf the information entropy, equation (1.1), and the discussion of it in reference to the full result equation (1.19)).

5.5.3 Energy and heat capacity

Multiplying both sides of the Helmholtz free energy, equation (5.98), by the inverse temperature and differentiating yields the most likely energy of the sub-system,

$$\frac{\partial(\beta\bar{F})}{\partial\beta} = \bar{E}.$$

(5.100)

Here $\bar{E}(N, V, T)$ has been held constant during the differentiating because the constrained Helmholtz free energy, $F(E \mid N, V, T)$, is a variational principle for the energy. This was given in chapter 2 as equation (2.35). From the above, the left-hand side is the negative logarithmic derivative of the partition function,

$$\frac{-\partial\ \ln\ Z(N, V, T)}{\partial\beta} = \frac{1}{N!h^{3N}Z} \int d\mathbf{\Gamma}\ \frac{-\partial}{\partial\beta}e^{-\beta\mathcal{H}(\mathbf{\Gamma})}$$

$$= \frac{1}{N!h^{3N}Z} \int d\mathbf{\Gamma}\ e^{-\beta\mathcal{H}(\mathbf{\Gamma})}\mathcal{H}(\mathbf{\Gamma})$$

$$= \langle\mathcal{H}\rangle_{N,V,T}.$$

(5.101)

This is equation (2.141) with $x = 1$ and $X = E$.

The energy of the sub-system is extensive, and its fluctuations are Gaussian. Hence the average value equals the most likely value,

$$\langle\mathcal{H}\rangle_{N,V,T} = \bar{E}(N, V, T).$$

(5.102)

With this one sees that the thermodynamic derivative equals the statistical mechanical derivative, which shows the consistency of the two.

The second derivative of the Helmholtz free energy gives the heat capacity at constant volume, equation (2.110),

$$C_V = \frac{-1}{T^2} \left(\frac{\partial^2 (\bar{F}/T)}{\partial (1/T)^2} \right)_{V,N} = -k_{\mathrm{B}} \beta^2 \frac{\partial}{\partial \beta} \frac{\partial (\beta \bar{F})}{\partial \beta}.$$

In terms of the partition function this is

$$\begin{aligned}
C_V &= k_{\mathrm{B}} \beta^2 \frac{\partial}{\partial \beta} \frac{Z\prime(\beta)}{Z(\beta)} \\
&= k_{\mathrm{B}} \beta^2 \left\{ \frac{Z\prime\prime(\beta)}{Z(\beta)} - \left(\frac{Z\prime(\beta)}{Z(\beta)} \right)^2 \right\} \\
&= k_{\mathrm{B}} \beta^2 \left\{ \langle \mathcal{H}^2 \rangle_{N,V,T} - \langle \mathcal{H} \rangle^2_{N,V,T} \right\} \\
&= k_{\mathrm{B}} \beta^2 \left\langle \left[\mathcal{H} - \langle \mathcal{H} \rangle_{N,V,T} \right]^2 \right\rangle_{N,V,T}.
\end{aligned}$$

This is just equation (2.144) with $x = 1$ and $X = E$. This is evidently positive and extensive (because the free energy is extensive). The square root of the right-hand side tells how likely an individual measurement of the energy is to depart from the average energy due to statistical variations. The relative root mean square fluctuation, $\langle [\mathcal{H} - \langle \mathcal{H} \rangle]^2 \rangle^{1/2} / \langle \mathcal{H} \rangle \sim V^{-1/2}$, vanishes in the thermodynamic limit, which is to say that the statistical measurement error is relatively negligible.

5.5.4 Virial pressure

The volume derivative of the Helmholtz free energy gives the pressure, equation (2.110),

$$\left(\frac{\partial \bar{F}(N,V,T)}{\partial V} \right)_{T,N} = -\bar{p}.$$

Using $y = p$ and $Y = V$ in equation (2.141) gives the equivalent expression for the logarithmic derivative of the partition function,

$$\frac{\partial \ln Z(N,V,T)}{\partial V} = \langle \beta_{\mathrm{s}} p \rangle_{N,V,T}.$$

Most commonly one regards the fluctuations in the sub-system temperature as negligible and one takes $\beta_{\mathrm{s}} = \beta$ and takes it outside of the average.

That part of the partition function that depends upon the volume is called the configuration integral. It is

$$\begin{aligned}
Q(N,V,T) &= \int_V d\mathbf{q} \; e^{-\beta U(\mathbf{q})} \\
&= L^{3N} \int_0^1 d\mathbf{x} \; e^{-\beta U(\mathbf{x};L)}.
\end{aligned}$$

In the second equality L has been used to scale all of the coordinates, $q_{j\alpha} = Lx_{j\alpha}$. Note that some authors use the opposite notation convention to here, denoting the partition function by Q and the configuration integral by Z.

Note that the potential energy that appears here is solely that due to the intermolecular contributions. The external contributions, which provide the external forces that balance the internal pressure being obtained here, implicitly appear as the limit on the configuration integral.

The derivative of the potential energy is

$$
\frac{\partial U(\mathbf{x}; L)}{\partial L} = \frac{\partial U(\mathbf{q})}{\partial \mathbf{q}} \cdot \frac{\partial \mathbf{q}}{\partial L}
$$
$$
= L^{-1} \mathbf{q} \cdot \nabla U(\mathbf{q})
$$
$$
\equiv -L^{-1} \mathcal{V}.
$$

(5.1 08)

In the final equality has been defined the virial of Clausius, $\mathcal{V}$. This is essentially the sum of the force acting on each particle due to the other particles. (We assume that only the internal potential energy between the particles alone appears here. Any external energy such as that due to the walls of the sub-system is incorporated into the limits on the integrals.) With this the logarithmic derivative of the partition function is

$$
\langle \beta_{\mathrm{s}} p \rangle = \frac{\partial \ln Z(N, V, T)}{\partial V}
$$
$$
= \frac{1}{Q} \frac{\partial Q(N, V, T)}{3L^2 \partial L}
$$
$$
= \frac{1}{3L^2 Q} \left[3NL^{3N-1} \int_0^1 \mathrm{d}\mathbf{x} \; \mathrm{e}^{-\beta U(\mathbf{x}; L)} - \beta L^{3N} \int_0^1 \mathrm{d}\mathbf{x} \; \mathrm{e}^{-\beta U(\mathbf{x}; L)} \frac{\partial U(\mathbf{x}; L)}{\partial L} \right]
$$
$$
= \frac{N}{V} + \frac{\beta \langle \mathcal{V} \rangle_{N, V, T}}{3V}.
$$

(5.1 09)

This is the virial equation for the pressure. The first term is the pressure of an ideal gas (see below), and the second is the average of the intermolecular forces. For a dilute real gas, this is dominated by the long-ranged attractions between the particles, which makes the pressure less than that of the corresponding ideal gas. For solids and dense fluids, the short-ranged core repulsion that defines the size of the particles typically dominates, which makes the pressure greater than that of the corresponding ideal gas. Because the core repulsion is short-ranged, it is steep, and therefore its contribution to the pressure increases rapidly with increasing density as the particles sample more of it. This means that dense fluids and solids have a pressure that increases more rapidly with density than an ideal gas, which means that they have a lower compressibility.

5.5.5 Ideal gas

The Hamiltonian separates into momentum and position parts, $\mathcal{H}(\mathbf{\Gamma}) = \mathcal{K}(\mathbf{p}) + U(\mathbf{q})$. For an equilibrium system one is generally interested in the average of a function of the velocity alone or of the position alone. Accordingly, the canonical equilibrium partition function can be written

$$Z(N, V, T) = \frac{1}{N! h^{3N}} \int d\mathbf{\Gamma} \ e^{-\beta \mathcal{H}(\mathbf{\Gamma})}$$

$$= \frac{1}{N! h^{3N}} \int d\mathbf{p} \ e^{-\beta p^2 / 2m} \int_V d\mathbf{q} \ e^{-\beta U(\mathbf{q})}$$

$$= \frac{\Lambda^{-3N}}{N!} \int_V d\mathbf{q} \ e^{-\beta U(\mathbf{q})}$$

$$= Z^{\text{id}}(N, V, T) V^{-N} Q(N, V, T).$$

(5.1 10)

The configuration integral was defined above as $Q(N, V, T) \equiv \int_V d\mathbf{q} \ e^{-\beta U(\mathbf{q})}$.

The ideal gas has $U(\mathbf{q}) = 0$, and the corresponding partition function is

$$Z^{\text{id}}(N, V, T) = \frac{1}{N! h^{3N}} \int d\mathbf{p} \ e^{-\beta p^2 / 2m} \int_V d\mathbf{q}$$

$$= \frac{1}{N! h^{3N}} [2\pi m k_{\text{B}} T]^{3N/2} V^N$$

$$\equiv \frac{V^N}{N! \Lambda^{3N}}.$$

(5.1 11)

The thermal wave length here and below is

$$\Lambda \equiv \sqrt{\frac{2\pi \hbar^2}{m k_{\text{B}} T}},$$

(5.1 12)

where $\hbar \equiv h/2\pi$. This quantity will recur in the quantum analysis of chapter 7, where it will be shown that it is the lower limit of the length scales on which wave function symmetrization effects are negligible. One must glean from the context in each case whether Λ means the thermal wave length or the transport coefficient matrix.

The average energy for the ideal gas is

$$\langle \mathcal{H}^{\text{id}} \rangle_{N,V,T} = \frac{-\partial \ln Z^{\text{id}}(N, V, T)}{\partial \beta}$$

$$= \frac{-\partial \ln \Lambda^{-3N}}{\partial \beta}$$

$$= \frac{3N}{2} k_{\text{B}} T.$$

(5.1 13)

The total unconstrained entropy of the ideal gas for this canonical equilibrium system is just the logarithm of the partition function,

$$S_{\text{tot}}^{\text{id}}(N, V, T) = k_{\text{B}} \ln Z^{\text{id}}(N, V, T) = N k_{\text{B}} [1 - \ln \rho \Lambda^3],$$

(5.1 14)

where the density is $\rho = N/V$. The reservoir part of this is

$$S_r^{id}(N,V,T) = \frac{-\langle \mathcal{H}^{id} \rangle_{N,V,T}}{T} = \frac{-3Nk_B}{2},$$

and the sub-system part is

$$S_s^{id}(N,V,T) = S_{tot}^{id}(N,V,T) - S_r^{id}(N,V,T) = Nk_B \left[\frac{5}{2} - \ln \rho \Lambda^3 \right].$$

The ideal gas heat capacity at constant volume is

$$\begin{aligned}
C_V^{id} &= -k_B \beta^2 \frac{\partial}{\partial \beta} \frac{\partial(\beta \bar{F}^{id})}{\partial \beta} \\
&= -k_B \beta^2 \frac{\partial \langle \mathcal{H}^{id} \rangle_{N,V,T}}{\partial \beta} \\
&= \frac{3N}{2} k_B.
\end{aligned}$$

This is evidently independent of temperature.

The pressure of the ideal gas is given by

$$\langle \beta_s p \rangle_{N,V,T}^{id} = \frac{\partial \ln Z^{id}(N,V,T)}{\partial V} = \frac{N}{V}.$$

Setting the sub-system temperature to that of the reservoir, $\beta_s = \beta = 1/k_B T$, and defining the number density as $\rho \equiv N/V$, this is $\bar{p}^{id} = \rho k_B T$, which is the well-known result. Rewriting equation (2.115) for a canonical equilibrium system, the ideal gas isothermal compressibility is

$$\chi_T^{id} = \frac{-1}{V} \left(\frac{\partial \bar{p}^{id}}{\partial V} \right)_{T,N}^{-1} = (\rho k_B T)^{-1}.$$

This says that as the number density and temperature increase, the ideal gas becomes less compressible.

The chemical potential of the ideal gas is given by

$$\langle \beta_s \mu \rangle_{N,V,T}^{id} = \frac{-\partial \ln Z^{id}(N,V,T)}{\partial N} = \ln \rho \Lambda^3,$$

or $\rho \Lambda^3 = e^{\beta \mu^{id}}$.

5.5.6 Equipartition theorem

The generalized equipartition theorem, equation (4.62), in the case of the canonical equilibrium system is

$$k_{\mathrm{B}}^{-2}\langle[\nabla S_{\mathrm{r}}(\boldsymbol{\Gamma})][\nabla S_{\mathrm{r}}(\boldsymbol{\Gamma})]\rangle_{N,V,T} = k_{\mathrm{B}}^{-2}\int\mathrm{d}\boldsymbol{\Gamma}\ \wp(\boldsymbol{\Gamma})[\nabla S_{\mathrm{r}}(\boldsymbol{\Gamma})][\nabla S_{\mathrm{r}}(\boldsymbol{\Gamma})]$$

$$= k_{\mathrm{B}}^{-1}\int\mathrm{d}\boldsymbol{\Gamma}\ [\nabla\wp(\boldsymbol{\Gamma})][\nabla S_{\mathrm{r}}(\boldsymbol{\Gamma})]$$

$$= -k_{\mathrm{B}}^{-1}\int\mathrm{d}\boldsymbol{\Gamma}\ \wp(\boldsymbol{\Gamma})\nabla\nabla S_{\mathrm{r}}(\boldsymbol{\Gamma})$$

$$= -k_{\mathrm{B}}^{-1}\langle\nabla\nabla S_{\mathrm{r}}(\boldsymbol{\Gamma})\rangle_{N,V,T}.$$

(5.121)

If, for example, this is evaluated for the momentum components of the gradient operator, the averand on the left-hand side is

$$k_{\mathrm{B}}^{-2}[\nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma})][\nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma})] = \frac{\beta^2}{m^2}\mathbf{p}\mathbf{p},$$

(5.122)

and that on the right-hand side is

$$-k_{\mathrm{B}}^{-1}\nabla_p\nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma}) = \frac{\beta}{m}\mathrm{I}.$$

(5.123)

If one actually evaluates the average of the left-hand side, since the momentum probability is a Gaussian, $\wp(\boldsymbol{\Gamma})\propto\mathrm{e}^{-\beta\mathbf{p}\cdot\mathbf{p}/2m}$, one sees indeed that

$$\frac{\beta^2}{m^2}\langle\mathbf{p}\mathbf{p}\rangle_{N,V,T} = \frac{\beta}{m}\mathrm{I}.$$

(5.124)

This is a general form of the traditional equipartition theorem.

More interesting is to consider whether or not this generalized equipartition theorem holds locally. That is, can these be equated without averaging,

$$\mathbf{p}\mathbf{p}\ \wp(\boldsymbol{\Gamma}) = mk_{\mathrm{B}}T\wp(\boldsymbol{\Gamma})\ \mathrm{I}?$$

(5.125)

We have included the probability density in this since it also appears in the integrand, and we wish to emphasize the fact that if this holds anywhere it can only be on the likely points of space. This is probably not such a good approximation, as one could choose values of the momentum components that are not individually unlikely but where the two matrices disagree in individual terms. (In fact, no dyadic matrix can be diagonal, a point which will prove of some significance in chapter 7.)

On the other hand, if one takes the trace of both sides this becomes

$$p^2\wp(\boldsymbol{\Gamma}) = 3Nmk_{\mathrm{B}}T\wp(\boldsymbol{\Gamma}).$$

(5.126)

This is actually quite a good approximation. Since the kinetic energy is extensive, on the likely points of phase space the fluctuations in it are relatively negligible, and the left-hand side can indeed be expected to equal the right-hand side. Although in a formal mathematical sense one can cancel the probability density on both sides of this, we prefer to leave it in to again emphasize that this only holds on the likely points of phase space. If one chose a set of momenta $p^2\neq 3Nmk_{\mathrm{B}}T$, then for this set $\wp(\boldsymbol{\Gamma})\approx 0$ and both sides of the equation are zero.

The generalized equipartition theorem is not restricted to momenta. For example, taking the position coordinates of the gradient operators yields

$$\beta^2 \langle [\nabla_q U(\mathbf{q})][\nabla_q U(\mathbf{q})] \rangle_{N,V,T} = \beta \langle \nabla_q \nabla_q U(\mathbf{q}) \rangle_{N,V,T}.$$

(5.127)

Taking the trace of this and assuming that it holds locally allows it to be written

$$\beta \wp(\mathbf{\Gamma}) \, \nabla_q U(\mathbf{q}) \cdot \nabla_q U(\mathbf{q}) = \wp(\mathbf{\Gamma}) \, \nabla_q^2 U(\mathbf{q}).$$

(5.128)

This takes on a rather simple form if the only potential energy is a one-body, parabolic, potential trap.

There is another result that can be called a generalized equipartition theorem. Consider the canonical equilibrium average of the $6N \times 6N$ dyadic,

$$
\begin{aligned}
\left\langle \mathbf{\Gamma} \frac{\partial \mathcal{H}(\mathbf{\Gamma})}{\partial \mathbf{\Gamma}} \right\rangle_{N,V,T} &= \frac{1}{h^{3N} N! Z} \int d\mathbf{\Gamma} \, e^{-\beta \mathcal{H}(\mathbf{\Gamma})} \mathbf{\Gamma} \frac{\partial \mathcal{H}(\mathbf{\Gamma})}{\partial \mathbf{\Gamma}} \\
&= \frac{-k_{\mathrm{B}} T}{h^{3N} N! Z} \int d\mathbf{\Gamma} \, \mathbf{\Gamma} \frac{\partial e^{-\beta \mathcal{H}(\mathbf{\Gamma})}}{\partial \mathbf{\Gamma}} \\
&= \frac{k_{\mathrm{B}} T}{h^{3N} N! Z} \int d\mathbf{\Gamma} \, e^{-\beta \mathcal{H}(\mathbf{\Gamma})} \frac{\partial \mathbf{\Gamma}}{\partial \mathbf{\Gamma}} \\
&= k_{\mathrm{B}} T \, \mathrm{I}.
\end{aligned}
$$

(5.129)

The third equality follows from an integration by parts. Choosing a single momentum component this gives the conventional equipartition theorem,

$$\left\langle \frac{1}{m} p_{j\alpha}^2 \right\rangle_{N,V,T} = k_{\mathrm{B}} T.$$

(5.130)

More generally, it gives

$$\left\langle \frac{1}{m} p_{j\alpha} p_{i\gamma} \right\rangle_{N,V,T} = \delta_{j,i} \delta_{\alpha,\gamma} k_{\mathrm{B}} T,$$

(5.131)

$$\left\langle q_{j\alpha} \frac{\partial U(\mathbf{q})}{\partial q_{i\gamma}} \right\rangle_{N,V,T} = \delta_{j,i} \delta_{\alpha,\gamma} k_{\mathrm{B}} T,$$

(5.132)

and

$$\left\langle p_{j\alpha} \frac{\partial U(\mathbf{q})}{\partial q_{i\gamma}} \right\rangle_{N,V,T} = \left\langle q_{j\alpha} \frac{1}{m} p_{i\gamma} \right\rangle_{N,V,T} = 0.$$

(5.133)

In order to connect the second of these with the virial of Clausius, one should ensure that the total potential energy is used, including the singlet potential due to any containing walls.

5.5.7 Particle densities and distributions

The one-particle density $\rho^{(1)}(\mathbf{r})$, usually denoted $\rho(\mathbf{r})$, or ρ for a homogeneous system, gives the number of particles per unit volume at $\mathbf{r}$. For a canonical equilibrium system it is

$$
\begin{aligned}
\rho^{(1)}(\mathbf{r}) &= \left\langle \sum_{j=1}^{N} \delta(\mathbf{r} - \mathbf{q}_j) \right\rangle_{N,V,T} \\
&= \frac{1}{h^{3N} N! Z(N,V,T)} \int d\mathbf{\Gamma} \; e^{-\beta \mathcal{H}(\mathbf{\Gamma})} \sum_{j=1}^{N} \delta(\mathbf{r} - \mathbf{q}_j) \\
&= \frac{1}{Q(N,V,T)} \int_V d\mathbf{q} \; e^{-\beta U(\mathbf{q})} \sum_{j=1}^{N} \delta(\mathbf{r} - \mathbf{q}_j) \\
&= \frac{N}{Q(N,V,T)} \int_V d\mathbf{q} \; e^{-\beta U(\mathbf{q})} \delta(\mathbf{r} - \mathbf{q}_N).
\end{aligned}
$$

(5.134)

The final equality follows because all the particles are identical. Equivalently,

$$
\rho^{(1)}(\mathbf{q}_N) = \frac{N}{Q(N,V,T)} \int_V d\mathbf{q}^{N-1} \; e^{-\beta U(\mathbf{q}^N)},
$$

(5.135)

where the integration is over the first $N-1$ particles, and the Nth particle is at $\mathbf{q}_N$.

The two-particle density is proportional to the probability of finding two different particles simultaneously at two positions. It is

$$
\rho^{(2)}(\mathbf{r}, \mathbf{s}) = \left\langle \sum_{j=1}^{N} \sum_{k=1}^{N} {}^{(k \neq j)} \; \delta(\mathbf{r} - \mathbf{q}_j) \delta(\mathbf{s} - \mathbf{q}_k) \right\rangle_{N,V,T},
$$

(5.136)

or

$$
\rho^{(2)}(\mathbf{q}_N, \mathbf{q}_{N-1}) = \frac{N(N-1)}{Q(N,V,T)} \int_V d\mathbf{q}^{N-2} \; e^{-\beta U(\mathbf{q}^N)}.
$$

(5.137)

Similarly, the n-particle density is

$$
\rho^{(n)}(\mathbf{q}_N, \mathbf{q}_{N-1}, \ldots, \mathbf{q}_{N-n+1}) = \frac{N!/(N-n)!}{Q(N,V,T)} \int_V d\mathbf{q}^{N-n} \; e^{-\beta U(\mathbf{q}^N)}.
$$

(5.138)

Appending the subscript N to make it clear that this is a closed or canonical density, this is evidently normalized such that

$$
\int_V d\mathbf{q}^n \; \rho_N^{(n)}(\mathbf{q}^n) = \frac{N!}{(N-n)!}.
$$

(5.139)

For an open or grand canonical system, the normalization is

$$\int_V \mathrm{d}\mathbf{q}^n \; \rho_\mu^{(n)}(\mathbf{q}^n) = \left\langle \frac{N!}{(N-n)!} \right\rangle_{\mu,V,T}.$$

One can confirm that the fluctuations in particle number are given by

$$\left\langle \left[N - \langle N \rangle_{\mu,V,T} \right]^2 \right\rangle_{\mu,V,T}$$
$$= \int_V \mathrm{d}\mathbf{r} \, \mathrm{d}\mathbf{s} \; \left\{ \rho_\mu^{(2)}(\mathbf{r},\mathbf{s}) - \rho_\mu^{(1)}(\mathbf{r})\rho_\mu^{(1)}(\mathbf{s}) + \rho_\mu^{(1)}(\mathbf{r})\delta(\mathbf{r}-\mathbf{s}) \right\}.$$

This is also the second derivative of the grand potential, (cf equation (2.144)), and it is related to the isothermal compressibility, (cf equation (2.115)), as is shown in equation (5.158).

For an ideal gas, $Q^{\mathrm{id}} = V^N$. Hence in the uniform case of no external one-body potential, the n-particle canonical density is

$$\rho_N^{(n),\mathrm{id}}(\mathbf{q}^n) = \frac{N! V^{N-n}}{(N-n)! V^N} = \frac{N}{V} \frac{N-1}{V} \cdots \frac{N-n+1}{V}.$$

For $n \ll N$ this is just ρ^n, where the singlet density is $\rho = N/V$.

For the case of an ideal gas in an external field $U(\mathbf{r})$, this is

$$\rho_N^{(n),\mathrm{id}}(\mathbf{q}^n) = \frac{N!}{(N-n)! \tilde{V}^n} \prod_{j=1}^n \mathrm{e}^{-\beta U(\mathbf{q}_j)}, \quad \tilde{V} \equiv \int_V \mathrm{d}\mathbf{r} \; \mathrm{e}^{-\beta U(\mathbf{r})}.$$

In general, particles are only correlated over short-ranges. At large separations between clusters of particles, the multi-particle density is a product form

$$\rho^{(m+n)}(\mathbf{r}^m, \mathbf{s}^n) \to \rho^{(m)}(\mathbf{r}^m) \, \rho^{(n)}(\mathbf{s}^n), \quad \text{all } \mid \mathbf{r}_j - \mathbf{s}_k \mid \to \infty.$$

In the opposite limit of small separations, real particles have finite size and a repulsive core that prevents overlap. Hence

$$\rho^{(n)}(\mathbf{r}^n) \to 0, \quad \text{any } \mid \mathbf{r}_j - \mathbf{r}_k \mid \to 0.$$

The n-particle distribution function is defined to be

$$g^{(n)}(\mathbf{r}^n) = \frac{\rho^{(n)}(\mathbf{r}^n)}{\prod_{j=1}^n \rho^{(1)}(\mathbf{r}_j)}.$$

The departure of this from unity is a measure of the correlation between the particles. Asymptotically,

$$g^{(n)}(\mathbf{r}^n) \to 1, \quad \text{all } r_{jk} \to \infty.$$

(5.147)

The most important distribution function is the pair one. For a uniform fluid (no one-body potential) in which the particles interact with a radially symmetric pair potential, this is the radial distribution function,

$$g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = g(r_{12}).$$

(5.148)

Here $r_{12} = | \mathbf{r}_1 - \mathbf{r}_2 |$ is the separation between the particles. One has

$$g(r) \to \begin{cases} 1, & r \to \infty, \\ 0, & r \to 0. \end{cases}$$

(5.149)

One also has

$$g(r) \to e^{-\beta u(r)}, \quad \rho \to 0.$$

(5.150)

Consider a uniform fluid, $U^{(1)}(\mathbf{r}) = 0$, with a radially symmetric pair potential,

$$U(\mathbf{q}) = \sum_{j=1}^{N} \sum_{k=j+1}^{N} u(q_{jk}) \equiv \sum_{k<j}^{N} u(q_{jk}).$$

(5.151)

The excess energy is the average potential energy, which is to say the energy without the kinetic energy. In this case for a canonical equilibrium system it is

$$\begin{aligned}
E^{\mathrm{ex}} &= \left\langle \sum_{k<j}^{N} u(q_{jk}) \right\rangle_{N,V,T} \\
&= \frac{N(N-1)}{2Q(N,V,T)} \int_V d\mathbf{q}^N \ e^{-\beta U(\mathbf{q}^N)} u(q_{N,N-1}) \\
&= \frac{1}{2} \int_V d\mathbf{q}_N \, d\mathbf{q}_{N-1} \ \rho^{(2)}(\mathbf{q}_N, \mathbf{q}_{N-1}) u(q_{N,N-1}) \\
&= \frac{\rho^2 V}{2} \int_V d\mathbf{r} \ g(r) u(r).
\end{aligned}$$

(5.152)

The integral is convergent and independent of V because generally $r^3 u(r) \to 0$, $r \to \infty$. (Coulomb and dipole interactions require special treatment.)

The virial of Clausius, equation (5.108), in this case is

$$\mathcal{V} = -\mathbf{q} \cdot \nabla U(\mathbf{q})$$

$$= -\sum_{i=1}^{N} \mathbf{q}_i \cdot \sum_{j<k}^{N} \frac{\partial u(q_{jk})}{\partial \mathbf{q}_i} = -\sum_{i=1}^{N} \mathbf{q}_i \cdot \sum_{k=1}^{N} {}^{(k \neq i)} \frac{\partial u(q_{ik})}{\partial \mathbf{q}_i}$$

$$= -\sum_{i \neq k}^{N} u\prime(q_{ik})\mathbf{q}_i \cdot \frac{\mathbf{q}_{ik}}{q_{ik}} = -\sum_{i<k}^{N} u\prime(q_{ik})\mathbf{q}_{ik} \cdot \frac{\mathbf{q}_{ik}}{q_{ik}}$$

$$= \frac{-1}{2} \sum_{i \neq k}^{N} q_{ik} u\prime(q_{ik}).$$

The canonical equilibrium average of this is

$$\langle \mathcal{V} \rangle_{N,V,T} = \left\langle -\sum_{k<j}^{N} q_{kj} u\prime(q_{kj}) \right\rangle_{N,V,T}$$

$$= \frac{-\rho^2 V}{2} \int_V \mathrm{d}\mathbf{r} \; g(r) r u\prime(r).$$

The analysis is identical to that for the excess energy with the replacement $u(r) \Rightarrow -r u\prime(r)$. From the virial equation (5.109), the average pressure is

$$\langle \beta_{\mathrm{s}} p \rangle = \frac{N}{V} + \frac{\beta \langle \mathcal{V} \rangle_{N,V,T}}{3V}$$

$$= \rho - \frac{\beta \rho^2}{6} \int_V \mathrm{d}\mathbf{r} \; g(r) r u\prime(r).$$

The first term is the ideal gas result. At low densities, $g(r) \sim 1$, the second term is dominated by the long-range tail, which is generally attractive (i.e. a negative potential decreasing with increasing separation, which is to say a positive slope). Hence the second term is negative in this range, and so the pressure is less than that of an ideal gas at the same density and temperature.

In section 2.5.3, the isothermal compressibility was given as the derivative of the volume with pressure at constant number and temperature, equation (2.115),

$$\chi_T = \frac{-1}{\bar{V}} \left(\frac{\partial \bar{V}}{\partial p} \right)_{T,N} = \frac{1}{\rho} \left(\frac{\partial \rho}{\partial p} \right)_T.$$

In the second equality the number density $\rho \equiv N/V$ has been used. In section 2.5.3 this was written as the second pressure derivative of the Gibbs free energy.

This can also be written as the second chemical potential derivative of the grand potential as follows. Above, the density was considered as a function of pressure and temperature, $\rho(p, T)$. But because of the uniqueness of the thermodynamic state, one can instead write $\rho(\mu, T)$ and $p(\mu, T)$, with which the compressibility becomes

$$\chi_T = \frac{1}{\rho}\left(\frac{\partial \rho}{\partial \mu}\right)_T \left(\frac{\partial \mu}{\partial p}\right)_T$$

$$= \frac{1}{\rho}\left(\frac{-\partial^2 \Omega}{V \partial \mu^2}\right)_{T,V} \left(\frac{N^{-1}\partial G}{\partial p}\right)_{T,N}$$

$$= \frac{-1}{\rho^2 V}\left(\frac{\partial^2 \Omega}{\partial \mu^2}\right)_{T,V}.$$

Writing $\Omega(\mu, V, T) = -k_{\mathrm{B}}T \ln \Xi(\mu, V, T)$, this is

$$\chi_T = \frac{1}{k_{\mathrm{B}}T\rho^2 V} \frac{\partial^2 \ln \Xi(\mu, V, T)}{\partial(\beta\mu)^2}$$

$$= \frac{1}{k_{\mathrm{B}}T\rho^2 V}\left\{\frac{\Xi''(\beta\mu)}{\Xi(\beta\mu)} - \left(\frac{\Xi'(\beta\mu)}{\Xi(\beta\mu)}\right)^2\right\}$$

$$= \frac{1}{k_{\mathrm{B}}T\rho^2 V}\left\{\langle N^2\rangle_{\mu,V,T} - \langle N\rangle^2_{\mu,V,T}\right\}$$

$$= \frac{1}{k_{\mathrm{B}}T\rho^2 V}\left\langle\left[N - \langle N\rangle_{\mu,V,T}\right]^2\right\rangle_{\mu,V,T}.$$

This is just equation (2.144) with $x = -\mu$ and $X = N$.

In equation (5.141), the fluctuations in number were related to the integral over the pair density. For a homogeneous system, $\rho_\mu^{(1)}(\mathbf{r}) = \rho$, with spherically symmetric pair potential, $\rho_\mu^{(2)}(\mathbf{r}, \mathbf{s}) = \rho^2 g(|\mathbf{r} - \mathbf{s}|)$ this is

$$\left\langle\left[N - \langle N\rangle_{\mu,V,T}\right]^2\right\rangle_{\mu,V,T} = \int_V d\mathbf{r}\, d\mathbf{s}\,\left\{\rho_\mu^{(2)}(\mathbf{r}, \mathbf{s}) - \rho_\mu^{(1)}(\mathbf{r})\rho_\mu^{(1)}(\mathbf{s}) + \rho_\mu^{(1)}(\mathbf{r})\delta(\mathbf{r} - \mathbf{s})\right\}$$

$$= \int_V d\mathbf{r}\, d\mathbf{s}\,\left\{\rho^2 g(|\mathbf{r} - \mathbf{s}|) - \rho^2 + \rho\delta(\mathbf{r} - \mathbf{s})\right\}$$

$$= \rho V + \rho^2 V \int d\mathbf{r}\,\{g(r) - 1\}.$$

The quantity in the integrand is called the total correlation function, $h(r) \equiv g(r) - 1$. It evidently goes to zero at large separations.

The isothermal compressibility diverges at spinodal points where the system is infinitely compressible. The curve of spinodal points lies inside the coexistence curve and it marks the limit of absolute thermodynamic stability of a phase. The critical point is the extremum of the spinodal curve. It is also the point at which the spinodal and coexistence curves coincide. The divergence of the density fluctuations due to the infinitely compressibility gives rise to the phenomenon known as critical opalescence.

5.6 Sub-system entropy as a functional of particle densities

5.6.1 Gibbs–Shannon information entropy

As was discussed in section 2.4.1, the Gibbs–Shannon information entropy, $S_{\mathrm{GS}}[\wp_\alpha] = -k_{\mathrm{B}} \sum_\alpha \wp_\alpha \ln \wp_\alpha$, gives only part of the total entropy, because it lacks the entropy internal to each macrostate α, $\sum_\alpha \wp_\alpha S_\alpha$. In thermodynamics and statistical mechanics, the 'external' entropy that is retained is the sub-system entropy, and the 'internal' entropy that is lacking is the reservoir entropy.

For example, for a canonical equilibrium system, the total entropy is

$$S_{\mathrm{tot}}(N, V, T) = k_{\mathrm{B}} \ln Z(N, V, T)$$
$$= S_{\mathrm{s}}(\bar{E}) + S_{\mathrm{r}}(\bar{E}),$$

(5.160)

the second equality holding for the most likely energy macrostate in the thermodynamic limit. The reservoir entropy is

$$S_{\mathrm{r}}(\bar{E}) = \frac{-\bar{E}(N, V, T)}{T} = \frac{-1}{T} \langle \mathcal{H} \rangle_{N,V,T},$$

(5.161)

and, as shown in equation (5.99), the sub-system entropy is

$$S_{\mathrm{s}}(\bar{E}, N, V) = -k_{\mathrm{B}} \int d\boldsymbol{\Gamma} \, \wp(\boldsymbol{\Gamma}) \ln \wp(\boldsymbol{\Gamma}).$$

(5.162)

5.6.2 Lattice models

We now give a rather useful approximation for calculating the sub-system entropy using the Gibbs–Shannon form and the multi-particle probability density. The scheme is most transparent for a lattice model.

Consider a lattice of N sites labeled $j = 1, 2, \ldots, N$, with a discrete variable σ_i at each site. This variable can represent whether or not a site is occupied by a particle, or the species of particle, or the magnetic spin state etc. The microstate of the system is signified σ^N, and it is assumed that for an isolated system all microstates have equal weight. The energy of the system is $\mathcal{H}(\sigma^N)$. At this stage we do not need to specify this or the dimensionality of the system.

For the canonical equilibrium case, the probability of a configuration is

$$\wp(\sigma^N) = \frac{1}{Z(N, T)} e^{-\beta \mathcal{H}(\sigma^N)},$$

(5.163)

the partition function being

$$Z(N, T) = \sum_{\sigma^N} e^{-\beta \mathcal{H}(\sigma^N)}.$$

(5.164)

There is no $N!$ here because the sites are regarded as distinguishable. The total unconstrained entropy is $S_{\mathrm{tot}}(N, T) = k_{\mathrm{B}} \ln Z(N, T)$. This is the sum of the reservoir entropy,

$$S_{\mathrm{r}}(\bar{E}) = \frac{-\bar{E}}{T} = \frac{-1}{T} \sum_{\sigma^N} \wp(\sigma^N) \, \mathcal{H}(\sigma^N),$$

and the sub-system entropy, which is the Gibbs–Shannon information entropy,

$$S_{\rm s}(\bar{E}) = S_{\rm GS}\left[\wp\right] = -k_{\rm B} \sum_{\sigma^N} \wp(\sigma^N) \ln\,\wp(\sigma^N).$$

A naive approximation to the sub-system entropy can be made based on its extensivity. The entropy per site may be expected to be independent of the size of the system provided that the system is not too small,

$$\frac{1}{N} S_{\rm GS}\left[\wp\right] = \frac{-k_{\rm B}}{N} \sum_{\sigma^N} \wp(\sigma^N) \ln\,\wp(\sigma^N)$$

$$\approx \frac{-k_{\rm B}}{n} \sum_{\sigma^n} \wp(\sigma^n) \ln\,\wp(\sigma^n).$$

(It will be shown that this is not a very accurate approximation for a highly coupled system.) Here $\wp(\sigma^n)$ is the n-site probability obtained in the full N-site system. For $n \ll N$, the sum over configurations σ^n is much more tractable than the one over σ^N. Although this expression for the entropy is more accurate for large n, it is also more difficult to obtain $\wp(\sigma^n)$ when n becomes larger.

The formal expression for the n-site probability is

$$\wp(\sigma\prime^n) = \sum_{\sigma^N} \wp(\sigma^N)\,\delta(\sigma^n - \sigma\prime^n).$$

In general terms it is relatively easy to numerically obtain accurate statistical results for the several site probability function, but this becomes more challenging as the order of the probability is increased. What is required is an approximation for the many-body probability in terms of the few body probability.

5.6.3 One dimension

It is most transparent to initially carry out the analysis for a one-dimensional crystal, and then to generalize it to two and three dimensions. We shall mainly consider consecutive sites, with the notation σ_j^n denoting the configurations of the n sites from site j, $\sigma_j^n = \{\sigma_j, \sigma_{j+1}, \ldots, \sigma_{j+n-1}\}$.

We shall write the n-site probability for σ^n at arbitrary sites j^n as

$$\wp^{(n)}(\sigma^n, j^n) = g^{(n)}(\sigma^n, j^n) \prod_{k=1}^{n} \wp^{(1)}(\sigma_k),$$

where the distribution function is unity for far separated sites. For n consecutive sites beginning at j we shall write

$$\wp^{(n)}(\sigma_j^n) = G^{(n)}(\sigma_j^n) \prod_{k=j}^{j+n-1} \wp^{(1)}(\sigma_k).$$

Assuming a homogeneous system, the singlet site probability is $\wp^{(1)}(\sigma)$. The two-site probability is $\wp^{(2)}(\sigma_1, \sigma_2)$. To leading order this can be taken to be the product of singlet probabilities,

$$\wp^{(2)}(\sigma_1, \sigma_2) = \wp^{(1)}(\sigma_1)\wp^{(1)}(\sigma_2)g^{(2)}(\sigma_1, \sigma_2) \approx \wp^{(1)}(\sigma_1)\wp^{(1)}(\sigma_2).$$

(5.171)

This is equivalent to setting the distribution function g or G to unity.

A very common approximation for the three consecutive site probability, $\wp^{(3)}(\sigma_1, \sigma_2, \sigma_3) = \wp^{(1)}(\sigma_1)\wp^{(1)}(\sigma_2)\wp^{(1)}(\sigma_3)G^{(3)}(\sigma_1, \sigma_2, \sigma_3)$, is the Kirkwood superposition approximation,

$$G^{(3)}(\sigma_1, \sigma_2, \sigma_3) \approx G^{(2)}(\sigma_1, \sigma_2)G^{(2)}(\sigma_2, \sigma_3)g^{(2)}(\sigma_1, \sigma_3; 1, 3).$$

(5.172)

The problem with this is that it double counts the correlation between sites 1 and 3, most of which is already counted in the product of the first two factors. It is far better to make a Markov approximation for this,

$$G^{(3)}(\sigma_1, \sigma_2, \sigma_3) \approx G^{(2)}(\sigma_1, \sigma_2)G^{(2)}(\sigma_2, \sigma_3).$$

(5.173)

One can formalize this by defining

$$G^{(3)}(\sigma_1, \sigma_2, \sigma_3) \equiv G^{(2)}(\sigma_1, \sigma_2)G^{(2)}(\sigma_2, \sigma_3)\Delta^{(3)}(\sigma_1, \sigma_2, \sigma_3).$$

(5.174)

The disparity $\Delta^{(3)}$ is what makes this exact. Setting the triplet disparity to unity yields the Markov approximation for the triplet distribution function.

Similarly, for four consecutive sites one defines

$$G^{(4)}(\sigma_1, \sigma_2, \sigma_3, \sigma_4) \equiv \frac{G^{(3)}(\sigma_1, \sigma_2, \sigma_3)G^{(3)}(\sigma_2, \sigma_3, \sigma_4)}{G^{(2)}(\sigma_2, \sigma_3)}\Delta^{(4)}(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$$
$$= G^{(2)}(\sigma_1, \sigma_2)G^{(2)}(\sigma_2, \sigma_3)G^{(2)}(\sigma_3, \sigma_4)$$
$$\times \Delta^{(3)}(\sigma_1, \sigma_2, \sigma_3)\Delta^{(3)}(\sigma_2, \sigma_3, \sigma_4)\Delta^{(4)}(\sigma_1, \sigma_2, \sigma_3, \sigma_4).$$

(5.175)

This again is exact. The $G^{(2)}$ in the denominator of the first equality accounts for double counting. Setting $\Delta^{(4)} = 1$ gives the triplet Markov approximation for the four-particle distribution, and in addition setting $\Delta^{(3)} = 1$ gives the pair Markov approximation for the four- particle distribution.

In general one has the formally exact product form

$$G^{(N)}(\sigma_1, \ldots, \sigma_N) \equiv \frac{G^{(N-1)}(\sigma_1, \ldots, \sigma_{N-1})G^{(N-1)}(\sigma_2, \ldots, \sigma_N)}{G^{(N-2)}(\sigma_2, \ldots, \sigma_{N-1})}\Delta^{(N)}(\sigma_1, \ldots, \sigma_N)$$
$$= G^{(2)}(\sigma_1, \sigma_2)G^{(2)}(\sigma_2, \sigma_3)\cdots G^{(2)}(\sigma_{N-1}, \sigma_N)$$
$$\times \Delta^{(3)}(\sigma_1, \sigma_2, \sigma_3)\cdots\Delta^{(N)}(\sigma_1, \ldots, \sigma_N).$$

(5.176)

Inserting this into the Gibbs–Shannon expression for the sub-system entropy, the logarithm of this product becomes a sum of terms, starting with the singlet, pair, triplet, etc. Summing over the redundant spin variables, the probability multiplying the logarithms is reduced to a singlet, pair, triplet, probability etc,

$$S_{\mathrm{s}}(N) = -k_{\mathrm{B}} \sum_{\sigma^N} \wp^{(N)}(\sigma^N) \ln \left\{ G^{(N)}(\sigma_1, \ldots, \sigma_N) \prod_{j=1}^{N} \wp^{(1)}(\sigma_j) \right\}$$

$$= -k_{\mathrm{B}} \sum_{j=1}^{N} \sum_{\sigma_j} \wp^{(1)}(\sigma_j) \ln \wp^{(1)}(\sigma_j)$$

$$-k_{\mathrm{B}} \sum_{j=1}^{N-1} \sum_{\sigma_j, \sigma_{j+1}} \wp^{(2)}(\sigma_j, \sigma_{j+1}) \ln G^{(2)}(\sigma_j, \sigma_{j+1})$$

$$-k_{\mathrm{B}} \sum_{j=1}^{N-2} \sum_{\sigma_j, \sigma_{j+1}, \sigma_{j+2}} \wp^{(3)}(\sigma_j, \sigma_{j+1}, \sigma_{j+2}) \ln \Delta^{(3)}(\sigma_j, \sigma_{j+1}, \sigma_{j+2})$$

$$- \cdots - k_{\mathrm{B}} \sum_{\sigma^N} \wp^{(N)}(\sigma^N) \ln \Delta^{(N)}(\sigma^N).$$

For a homogeneous system there are N equivalent singlet terms, $N-1$ equivalent pair terms, $N-2$ equivalent triplet terms, etc. In the thermodynamic limit each of these may be replaced by N and the expansion for the sub-system entropy per site is

$$S_{\mathrm{s}}(N)/N = -k_{\mathrm{B}} \sum_{\sigma} \wp^{(1)}(\sigma) \ln \wp^{(1)}(\sigma)$$

$$-k_{\mathrm{B}} \sum_{\sigma_1, \sigma_2} \wp^{(2)}(\sigma_1, \sigma_2) \ln G^{(2)}(\sigma_1, \sigma_2)$$

$$-k_{\mathrm{B}} \sum_{\sigma_1, \sigma_2, \sigma_3} \wp^{(3)}(\sigma_1, \sigma_2, \sigma_3) \ln \Delta^{(3)}(\sigma_1, \sigma_2, \sigma_3)$$

$$- \cdots.$$

If one has obtained the probability $\wp^{(n)}$, then one can evaluate this up to order n, since all the $\wp^{(j)}$ and $\Delta^{(j)}$, $j \leqslant n$ can be obtained by reducing $\wp^{(n)}$. This can be called the nth order Markov superposition approximation to the Gibbs–Shannon form of the sub-system entropy.

Terminating this expansion at some order n is expected to be accurate because it is equivalent to invoking a Markov superposition approximation at that order. This is much more accurate than, for example, the Kirkwood superposition approximation, because it does not double count the correlations, which can be particularly problematic at high couplings. This nth order Markov superposition approximation is also more accurate than simply inserting $\wp^{(n)}$ directly into the Gibbs–Shannon form of the sub-system entropy,

$$S_{\mathrm{s}}^{(n)} \equiv -k_{\mathrm{B}} \sum_{\sigma^n} \wp^{(n)}(\sigma^n) \ln \wp^{(n)}(\sigma^n),$$

and taking $S_{\mathrm{s}}(N)/N \approx S_{\mathrm{s}}^{(n)}/n.$

There is another way to write the *n*th order Markov superposition approximation that is particularly easy to generalize to higher dimensions. In view of the definition of the disparity,

$$\Delta^{(n)}(\sigma_1, \ldots, \sigma_n) \equiv \frac{\wp^{(n)}(\sigma_1, \ldots, \sigma_n)\wp^{(n-2)}(\sigma_2, \ldots, \sigma_{n-1})}{\wp^{(n-1)}(\sigma_1, \ldots, \sigma_{n-1})\wp^{(n-1)}(\sigma_2, \ldots, \sigma_n)},$$

(5.180)

one can write the difference between successive approximations to the entropy as

$$
\begin{aligned}
S_{\mathrm{s}}^{(n)} - S_{\mathrm{s}}^{(n-1)} &= -k_{\mathrm{B}} \sum_{\sigma^n} \wp^{(n)}(\sigma^n) \ln \frac{\wp^{(n)}(\sigma^n)}{\wp^{(n-1)}(\sigma^{n-1})} \\
&= -k_{\mathrm{B}} \sum_{\sigma^n} \wp^{(n)}(\sigma^n) \ln \frac{\Delta^{(n)}(\sigma^n)\wp^{(n-1)}(\sigma^{n-1})}{\wp^{(n-2)}(\sigma^{n-2})} \\
&= S_{\mathrm{s}}^{(n-1)} - S_{\mathrm{s}}^{(n-2)} - k_{\mathrm{B}} \sum_{\sigma^n} \wp^{(n)}(\sigma^n) \ln \Delta^{(n)}(\sigma^n).
\end{aligned}
$$

(5.181)

Obviously, one can repeat this down to *n* = 3 to obtain

$$
\begin{aligned}
S_{\mathrm{s}}^{(n)} - S_{\mathrm{s}}^{(n-1)} &= -k_{\mathrm{B}} \sum_{\sigma^n} \wp^{(n)}(\sigma^n) \ln \Delta^{(n)}(\sigma^n) \\
&\quad - k_{\mathrm{B}} \sum_{\sigma^{n-1}} \wp^{(n-1)}(\sigma^{n-1}) \ln \Delta^{(n-1)}(\sigma^{n-1}) \\
&\quad - \cdots - k_{\mathrm{B}} \sum_{\sigma^3} \wp^{(3)}(\sigma^3) \ln \Delta^{(3)}(\sigma^3) \\
&\quad - k_{\mathrm{B}} \sum_{\sigma^2} \wp^{(2)}(\sigma^2) \ln \wp^{(2)}(\sigma^2) + k_{\mathrm{B}} \sum_{\sigma} \wp^{(1)}(\sigma) \ln \wp^{(1)}(\sigma) \\
&= -k_{\mathrm{B}} \sum_{\sigma^n} \wp^{(n)}(\sigma^n) \ln \Delta^{(n)}(\sigma^n) - \cdots - k_{\mathrm{B}} \sum_{\sigma^3} \wp^{(3)}(\sigma^3) \ln \Delta^{(3)}(\sigma^3) \\
&\quad - k_{\mathrm{B}} \sum_{\sigma^2} \wp^{(2)}(\sigma^2) \ln G^{(2)}(\sigma^2) - k_{\mathrm{B}} \sum_{\sigma} \wp^{(1)}(\sigma) \ln \wp^{(1)}(\sigma).
\end{aligned}
$$

(5.182)

One sees that this is identical to the *n*th order Markov superposition approximation to the Gibbs–Shannon form of the sub-system entropy per site,

$$S_{\mathrm{s}}(N)/N \approx S_{\mathrm{s}}^{(n)} - S_{\mathrm{s}}^{(n-1)} \ (\mathrm{Markov}).$$

(5.183)

The difference between the two terms on the right-hand side reflects the subtraction of the doubly counted correlations, which is a feature of the Markov approach. This approximation is equivalent to setting the higher order disparities to unity, $\Delta^{(m)} = 1$, $m \geqslant n + 1$. This expression is expected to be much more accurate than the simple minded $S_{\mathrm{s}}(N)/N \approx S_{\mathrm{s}}^{(n)}/n$, particularly at higher couplings between the sites.

5.6.4 Two and three dimensions

Consider a two-dimensional lattice. For simplicity take this to be a square lattice, but this is not essential. Let $\wp^{(n,m)}(\sigma^{nm})$ be the probability for the configuration σ^{nm} on an $n \times m$ block of sites. The sub-system entropy of such a block is

$$S_{\mathrm{s}}^{(n,m)} \equiv -k_{\mathrm{B}} \sum_{\sigma^{nm}} \wp^{(n,m)}(\sigma^{nm}) \ln \wp^{(n,m)}(\sigma^{nm}).$$

(5.184)

The n, mth Markov superposition approximation to the Gibbs–Shannon form of the sub-system entropy in two dimensions is

$$S_{\mathrm{s}}(N)/N \approx S_{\mathrm{s}}^{(n,m)} - S_{\mathrm{s}}^{(n-1,m)} - S_{\mathrm{s}}^{(n,m-1)} + S_{\mathrm{s}}^{(n-1,m-1)}.$$

(5.185)

This is generalization of the one-dimensional result. It can be seen that the entropy of the common sites that is subtracted twice is added back by the final term. Again, since this correctly counts the correlated contributions, it can be expected to be much more accurate than the simple minded $S_{\mathrm{s}}(N)/N \approx S_{\mathrm{s}}^{(n,m)}/nm$, particularly at high couplings.

Analogously, the three-dimensional result is

$$
\begin{aligned}
S_{\mathrm{s}}(N)/N \approx\ & S_{\mathrm{s}}^{(n,m,k)} - S_{\mathrm{s}}^{(n-1,m,k)} - S_{\mathrm{s}}^{(n,m-1,k)} - S_{\mathrm{s}}^{(n,m,k-1)} \\
& + S_{\mathrm{s}}^{(n,m-1,k-1)} + S_{\mathrm{s}}^{(n-1,m,k-1)} + S_{\mathrm{s}}^{(n-1,m-1,k)} \\
& - S_{\mathrm{s}}^{(n-1,m-1,k-1)}.
\end{aligned}
$$

(5.186)

Here the Gibbs–Shannon form of the sub-system entropy of a $n \times m \times k$ block of sites is

$$S_{\mathrm{s}}^{(n,m,k)} \equiv -k_{\mathrm{B}} \sum_{\sigma^{nmk}} \wp^{(n,m,k)}(\sigma^{nmk}) \ln \wp^{(n,m,k)}(\sigma^{nmk}).$$

(5.187)

5.6.4.1 Ising model test

These approximations are readily tested for the two-dimensional Ising model, for which exact results are known (Baxter 1982). In this model the Hamiltonian is

$$\mathcal{H}(\sigma^N) = -J \sum_{j,k}^{(\mathrm{nn})} \sigma_j \sigma_k.$$

(5.188)

where nn denotes the nearest neighbors on the lattice, and the spin variable can be $\sigma = \pm 1$. The weight of each microstate for an isolated sub-system is unity, and hence the sub-system microstate entropy vanishes, $S_{\mathrm{s}}(\sigma^N) = 0$.

From Onsager's exact solution, the unconstrained total entropy per site is (Wikipedia 2017)

$$
\begin{aligned}
\frac{S_{\mathrm{total}}(N,T)}{N} &\equiv \frac{1}{N} \ln Z_N(T) \\
&= \frac{1}{2} \ln 2 + \frac{1}{2\pi} \int_0^\pi \mathrm{d}\theta\ \ln\left\{ \cosh^2 2K + k^{-1}\sqrt{1 + k^2 - 2k\,\cos(2\theta)} \right\},
\end{aligned}
$$

(5.189)

where the dimensionless coupling parameter is $K \equiv J/k_{\mathrm{B}}T$, and $k \equiv 1/\sinh^2 2K$. The average (equivalently, most likely) energy per site is (Wikipedia 2017)

$$\frac{\beta}{N}\bar{E}(N,T) = -K \coth 2K \left\{ 1 + \frac{2}{\pi}[2\tanh^2 2K - 1] \int_0^{\pi/2} \frac{\mathrm{d}\theta}{\sqrt{1 - 4k(1+k)^{-2}\sin^2\theta}} \right.$$

(5.1 90)

From these can be calculated the reservoir entropy per site

$$\frac{S_{\mathrm{r}}(N,T)}{N} = \frac{-1}{NT}\bar{E}(N,T),$$

(5.1 91)

and the sub-system entropy per site

$$\frac{S_{\mathrm{s}}(N,T)}{N} = \frac{S_{\mathrm{tot}}(N,T)}{N} + \frac{1}{NT}\bar{E}(N,T).$$

(5.1 92)

Figure 5.6 shows the exact results for the entropy of the two-dimensional Ising model as a function of the coupling parameter. The reservoir entropy goes to zero as the coupling parameter goes to zero because in this regime the sub-system energy vanishes, and hence so does the reservoir entropy. The sub-system entropy decreases as the coupling parameter increases because the spins are increasingly correlated and aligned, which is an ordered arrangement.

Figure 5.6. The entropy per site for the Ising model on a square lattice as a function of the coupling parameter for the canonical equilibrium system. The curves are the exact result for the total (solid), reservoir (dotted), and sub-system (dashed) entropy. The symbols are the Gibbs–Shannon information entropy using the four-spin probability function on a 2 × 2 block, with the filled triangles being the Markov superposition formula, $S_{\mathrm{s}}^{(2,2)} - 2S_{\mathrm{s}}^{(2,1)} + S_{\mathrm{s}}^{(1,1)}$, and the empty circles being $S_{\mathrm{s}}^{(2,2)}/4$. Data from Attard (1999).

The two approximations for the Gibbs–Shannon form of the sub-system entropy are tested, both based on the four-spin probability $\wp^{(2,2)}(\sigma^4)$ on a 2 × 2 block. The probability was obtained by enumerating all configurations on a 5 × 5 lattice with periodic boundary conditions (Attard 1999). It can be seen that the Gibbs–Shannon entropy agrees with the exact sub-system entropy, which confirms that it is just part of the total entropy. The Markov superposition approximation, $S_{\mathrm{s}}^{(2,2)} - 2S_{\mathrm{s}}^{(2,1)} + S_{\mathrm{s}}^{(1,1)}$, performs much better than the naive $S_{\mathrm{s}}^{(2,2)}/4$ at higher couplings. It appears that the simple approach implicitly assumes a disordered state for spins beyond the domain upon which it is calculated. At high couplings when the correlation length extends beyond this domain, this underestimates the real degree of correlation beyond the domain, and hence overestimates the entropy of the system. Evidently, the Markov approach is able to correctly account for the correlations between the spins, even at high couplings. It is surprising how well the Markov superposition approximation performs considering that only the four-spin distribution function was used.

5.6.5 Continuum disordered fluid

In this subsection the focus is on the positional contribution to the sub-system entropy for particles in a continuum. The ideal gas contribution, which is to be added to this excess, was given above as equation (5.116).

The n-particle position probability, $\wp^{(n)}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n)$, is symmetric with respect to interchange of the arguments. In two and three dimensions, this appears to create difficulties for developing a Markov-type approximation for the entropy. Nevertheless, one can proceed at least formally.

Define the thee-particle disparity as

$$g^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) \equiv g^{(2)}(\mathbf{r}_1, \mathbf{r}_2) g^{(2)}(\mathbf{r}_2, \mathbf{r}_3) \Delta^{(3)}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3).$$

(5.193)

Since the left-hand side is symmetric in its arguments, the disparity must be asymmetric in its arguments. One expects the disparity to be close to unity when $\mathbf{r}_2$ lies between the other two positions. One can define the higher order disparities also as in the one-dimensional spin lattice case.

Proceeding as in that case, the excess (i.e. without the ideal gas part) Gibbs–Shannon form of the sub-system entropy is

$$
\begin{aligned}
S_{\mathrm{s}}^{\mathrm{ex}}(N) = & - k_{\mathrm{B}} \int_V \mathrm{d}\mathbf{r}^N \, \wp^{(N)}(\mathbf{r}^N) \ln \left\{ g^{(N)}(\mathbf{r}_1, \ldots, \mathbf{r}_N) \prod_{j=1}^N \wp^{(1)}(\mathbf{r}_j) \right\} \\
= & - k_{\mathrm{B}} \sum_{j=1}^N \int_V \mathrm{d}\mathbf{r}_j \, \wp^{(1)}(\mathbf{r}_j) \ln \wp^{(1)}(\mathbf{r}_j) \\
& - k_{\mathrm{B}} \sum_{j=1}^{N-1} \int_V \mathrm{d}\mathbf{r}_j \, \mathrm{d}\mathbf{r}_{j+1} \, \wp^{(2)}(\mathbf{r}_j, \mathbf{r}_{j+1}) \ln g^{(2)}(\mathbf{r}_j, \mathbf{r}_{j+1}) \\
& - k_{\mathrm{B}} \sum_{j=1}^{N-2} \int_V \mathrm{d}\mathbf{r}_j \, \mathrm{d}\mathbf{r}_{j+1} \, \mathrm{d}\mathbf{r}_{j+2} \, \wp^{(3)}(\mathbf{r}_j, \mathbf{r}_{j+1}, \mathbf{r}_{j+2}) \ln \Delta^{(3)}(\mathbf{r}_j, \mathbf{r}_{j+1}, \mathbf{r}_{j+2}) \\
& - \cdots - k_{\mathrm{B}} \int_V \mathrm{d}\mathbf{r}^N \, \wp^{(N)}(\mathbf{r}^N) \ln \Delta^{(N)}(\mathbf{r}^N).
\end{aligned}
$$

(5.194)

For a homogeneous system each of these contributes essentially N equivalent terms, and the expansion for the excess sub-system entropy per particle is

$$
\begin{aligned}
S_{\mathrm{s}}^{\mathrm{ex}}(N)/N = & -k_{\mathrm{B}} \int_V \mathrm{d}\mathbf{r}\; \wp^{(1)}(\mathbf{r}) \ln\, \wp^{(1)}(\mathbf{r}) \\
& - k_{\mathrm{B}} \int_V \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{s}\; \wp^{(2)}(\mathbf{r},\mathbf{s}) \ln\, g^{(2)}(\mathbf{r},\mathbf{s}) \\
& - k_{\mathrm{B}} \int_V \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{s}\, \mathrm{d}\mathbf{t}\; \wp^{(3)}(\mathbf{r},\mathbf{s},\mathbf{t}) \ln\, \Delta^{(3)}(\mathbf{r},\mathbf{s},\mathbf{t}) - \cdots
\end{aligned}
$$

(5.195)

Recall that $\Delta^{(3)}(\mathbf{r},\mathbf{s},\mathbf{t})$ is symmetric with respect to interchange of $\mathbf{r}$ and $\mathbf{t}$, but that it is asymmetric for any interchange with $\mathbf{s}$. This expansion is formally exact.

Define

$$
S_{\mathrm{s}}^{\mathrm{ex},(n)} = -k_{\mathrm{B}} \int_V \mathrm{d}\mathbf{r}^n\; \wp^{(n)}(\mathbf{r}^n) \ln\, \wp^{(n)}(\mathbf{r}^n).
$$

(5.196)

One has

$$
\begin{aligned}
S_{\mathrm{s}}^{\mathrm{ex},(n)} - S_{\mathrm{s}}^{\mathrm{ex},(n-1)} = & -k_{\mathrm{B}} \int_V \mathrm{d}\mathbf{r}^n\; \wp^{(n)}(\mathbf{r}^n) \ln\, \frac{\wp^{(n)}(\mathbf{r}^n)}{\wp^{(n-1)}(\mathbf{r}^{n-1})} \\
= & -k_{\mathrm{B}} \int_V \mathrm{d}\mathbf{r}^n\; \wp^{(n)}(\mathbf{r}^n) \ln\, \frac{\Delta^{(n)}(\mathbf{r}^n)\wp^{(n-1)}(\mathbf{r}^{n-1})}{\wp^{(n-2)}(\mathbf{r}^{n-2})} \\
= & \; S_{\mathrm{s}}^{\mathrm{ex},(n-1)} - S_{\mathrm{s}}^{\mathrm{ex},(n-2)} - k_{\mathrm{B}} \int_V \mathrm{d}\mathbf{r}^n\; \wp^{(n)}(\mathbf{r}^n) \ln\, \Delta^{(n)}(\mathbf{r}^n),
\end{aligned}
$$

(5.197)

which telescopes to

$$
\begin{aligned}
S_{\mathrm{s}}^{\mathrm{ex},(n)} - S_{\mathrm{s}}^{\mathrm{ex},(n-1)} = & -k_{\mathrm{B}} \int_V \mathrm{d}\mathbf{r}^n\; \wp^{(n)}(\mathbf{r}^n) \ln\, \Delta^{(n)}(\mathbf{r}^n) \\
& - \cdots - k_{\mathrm{B}} \int_V \mathrm{d}\mathbf{r}^3\; \wp^{(3)}(\mathbf{r}^3) \ln\, \Delta^{(3)}(\mathbf{r}^3) \\
& - k_{\mathrm{B}} \int_V \mathrm{d}\mathbf{r}^2\; \wp^{(2)}(\mathbf{r}^2) \ln\, g^{(2)}(\mathbf{r}^2) - k_{\mathrm{B}} \int_V \mathrm{d}\mathbf{r}\; \wp^{(1)}(\mathbf{r}) \ln\, \wp^{(1)}(\mathbf{r}).
\end{aligned}
$$

(5.198)

This is just the Markov expression obtained above. Hence one can take the excess sub-system entropy per particle to be

$$
S_{\mathrm{s}}^{\mathrm{ex}}(N)/N \approx S_{\mathrm{s}}^{\mathrm{ex},(n)} - S_{\mathrm{s}}^{\mathrm{ex},(n-1)}.
$$

(5.199)

Whatever the general utility of the Markov expansion for the particle distribution functions in two and three dimensions, this is, presumably, the best approximation to the sub-system entropy that one can make.

5.7 Time correlation and van Hove functions

5.7.1 Time correlation function

Let $A(\mathbf{\Gamma})$ and $B(\mathbf{\Gamma})$ be functions of phase space. The time correlation function for them is usually defined in terms of the correlation between the fluctuations from their average value,

$$C_{AB}(\tau) = \langle [A(\tau) - \langle A \rangle] [B(0) - \langle B \rangle] \rangle_{N,V,T}.$$

(5.200)

In the long-time limit the two variables must be uncorrelated and so one has

$$C_{AB}(\tau) \to \langle [A(\tau) - \langle A \rangle] \rangle_{N,V,T} \langle [B(0) - \langle B \rangle] \rangle_{N,V,T}, \quad \tau \to \infty$$
$$= 0.$$

(5.201)

In terms of the stochastic, dissipative trajectory $\mathbf{\Gamma}(t \mid \mathbf{\Gamma}_0)$ this is

$$C_{AB}(\tau) = \frac{1}{t} \int_0^t dt\prime\, [A(\mathbf{\Gamma}(t\prime+\tau \mid \mathbf{\Gamma}_0)) - \langle A \rangle][B(\mathbf{\Gamma}(t\prime\mid \mathbf{\Gamma}_0)) - \langle B \rangle].$$

(5.202)

For large t this should be independent of the starting position $\mathbf{\Gamma}_0$. One often sees in the literature the stochastic, dissipative trajectory that appears here approximated by an adiabatic or even by a deterministic thermostatted trajectory.

Let $\varepsilon = \pm 1$ be the time parity such that

$$A(\mathbf{\Gamma}^\dagger) = \varepsilon_A A(\mathbf{\Gamma}) \ \text{and} \ B(\mathbf{\Gamma}^\dagger) = \varepsilon_B B(\mathbf{\Gamma}).$$

(5.203)

Recall that the conjugate phase point has the same positions but reversed momenta. Note that $\langle A \rangle$ is nonzero only if A has even parity. Hence one can always write $\langle A \rangle = \varepsilon_A \langle A \rangle$ and $\langle B \rangle = \varepsilon_B \langle B \rangle$.

As shown in section 5.4.2, the stochastic, dissipative equations of motion possess microscopic reversibility. Equation (5.67) reads,

$$\mathbf{\Gamma}_2 = \mathbf{\Gamma}(t \mid \mathbf{\Gamma}_1) \Leftrightarrow \mathbf{\Gamma}_2^\dagger = \mathbf{\Gamma}(-t \mid \mathbf{\Gamma}_1^\dagger),$$

(5.204)

which is to be understood in a probabilistic sense. With these, and starting the trajectory from the equally likely conjugate phase space point $\mathbf{\Gamma}_0^\dagger$, the time correlation function has the symmetry

$$C_{AB}(\tau) = \frac{1}{t} \int_0^t dt\prime\, \left[A(\mathbf{\Gamma}(t\prime+\tau \mid \mathbf{\Gamma}_0^\dagger)) - \langle A \rangle \right] \left[B(\mathbf{\Gamma}(t\prime\mid \mathbf{\Gamma}_0^\dagger)) - \langle B \rangle \right]$$
$$= \frac{1}{t} \int_0^t dt\prime\, \left[A(\mathbf{\Gamma}(-t\prime-\tau \mid \mathbf{\Gamma}_0)^\dagger) - \langle A \rangle \right] \left[B(\mathbf{\Gamma}(-t\prime\mid \mathbf{\Gamma}_0)^\dagger) - \langle B \rangle \right]$$
$$= \frac{\varepsilon_A \varepsilon_B}{t} \int_0^t dt\prime\, [A(\mathbf{\Gamma}(-t\prime-\tau \mid \mathbf{\Gamma}_0)) - \langle A \rangle][B(\mathbf{\Gamma}(-t\prime\mid \mathbf{\Gamma}_0)) - \langle B \rangle]$$
$$= \varepsilon_A \varepsilon_B C_{AB}(-\tau).$$

(5.205)

The final equality follows because of time homogeneity: only the relative times at which A and B are evaluated matters.

One can conclude from this that

$$C_{AB}(0) = 0, \quad \varepsilon_A \neq \varepsilon_B.$$

This says that there can be no instantaneous correlation between variables of opposite time parity.

5.7.2 van Hove function

The one-particle number density phase space function is defined as

$$\hat{\rho}^{(1)}(\mathbf{r}, t) = \sum_{j=1}^{N} \delta(\mathbf{r} - \mathbf{q}_j(t)),$$

where $\mathbf{q}_j$ is the position of molecule j. For a closed system, this integrates to the total number, $\int d\mathbf{r}\, \hat{\rho}(\mathbf{r}, t) = N$.

The caret distinguishes this phase space function from the average particle density of the type discussed in section 5.5.7. The singlet density is

$$\left\langle \hat{\rho}^{(1)}(\mathbf{r}, t) \right\rangle = \rho^{(1)}(\mathbf{r}).$$

An equilibrium system is assumed here.

The density–density correlation function is a form of time correlation function. Taking $A \equiv \hat{\rho}^{(1)}(\mathbf{r}\prime + \mathbf{r}, t)$ and $B \equiv \hat{\rho}^{(1)}(\mathbf{r}\prime, 0)$, this can be defined as

$$G^{(2)}(\mathbf{r}, \mathbf{r}\prime, t) \equiv N^{-1} \left\langle \hat{\rho}^{(1)}(\mathbf{r}\prime + \mathbf{r}, t)\, \hat{\rho}^{(1)}(\mathbf{r}\prime, 0) \right\rangle.$$

This gives the probability of a molecule being at $\mathbf{r}\prime$ and a molecule being located a distance $\mathbf{r}$ from $\mathbf{r}\prime$ at a time t later. These may be the same molecule.

This density–density correlation function should not be confused with the pair distribution function $g^{(2)}$ introduced in section 5.5.7, or the consecutive site correlation function of equation (5.170). Asymptotically, $G^{(2)}(\mathbf{r}, \mathbf{r}\prime, t) \sim \rho^{(1)}(\mathbf{r}\prime + \mathbf{r})\rho^{(1)}(\mathbf{r})/N$, as $\mathbf{r}\prime \to \infty$ or $t \to \infty$. In contrast, $g^{(2)}(\mathbf{r}, \mathbf{r}\prime) \sim 1$, as $|\mathbf{r} - \mathbf{r}\prime| \to \infty$.

For a homogenous system only the distance matters. Integrating the density–density correlation function over the volume gives the van Hove function (Hansen and McDonald 1990),

$$G(\mathbf{r}, t) = \int d\mathbf{r}\prime\, G(\mathbf{r}, \mathbf{r}\prime, t)$$

$$= \int d\mathbf{r}\prime\, \frac{1}{N} \left\langle \sum_{i=1}^{N} \sum_{j=1}^{N} \delta(\mathbf{r}\prime + \mathbf{r} - \mathbf{q}_i(t))\, \delta(\mathbf{r}\prime - \mathbf{q}_j(0)) \right\rangle$$

$$= \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \langle \delta(\mathbf{r} + \mathbf{q}_j(0) - \mathbf{q}_i(t)) \rangle.$$

This is equivalent to $G(\mathbf{r}, t) = VG(\mathbf{r}, \mathbf{0}, t) = \rho^{-1}\left\langle \hat{\rho}^{(1)}(\mathbf{r}, t)\hat{\rho}^{(1)}(\mathbf{0}, 0)\right\rangle$. The van Hove (5.2 10) function can be measured by x-ray or neutron scattering techniques.

The self-part of this gives the motion of an individual molecule over time,

$$G_{\mathrm{s}}(\mathbf{r}, t) = \frac{1}{N} \sum_{j=1}^{N} \langle \delta(\mathbf{r} + \mathbf{q}_j(0) - \mathbf{q}_j(t))\rangle.$$

(5.2 11)

In essence, this is the probability that a molecule moves $\mathbf{r}$ in time t. The distinct part gives the correlation of different molecules

$$G_{\mathrm{d}}(\mathbf{r}, t) = \frac{1}{N} \sum_{i=1}^{N}\sum_{j=1}^{N} {}^{(i\neq j)}\langle \delta(\mathbf{r} + \mathbf{q}_i(0) - \mathbf{q}_j(t))\rangle.$$

(5.2 12)

At large distances and at large times the molecules are uncorrelated, so that one has $\lim_{\mathbf{r}\to\infty} G_{\mathrm{d}}(\mathbf{r}, t) = \lim_{t\to\infty} G_{\mathrm{d}}(\mathbf{r}, t) = \rho$, assuming a homogeneous system.

Summary

- Classical phase space forms the microstates for a classical system. Hamilton's equations of motion govern the adiabatic evolution of an isolated sub-system in phase space. For an isolated sub-system, classical phase space has uniform weight density and therefore zero internal entropy.
- The canonical equilibrium probability density in phase space is the Maxwell–Boltzmann distribution. In general equilibrium probability distributions in phase space are proportional to the exponential of the reservoir entropy.
- Equilibrium averages can be given as either a time average over the stochastic, dissipative trajectory in phase space, or else as an integral over phase space weighted by the equilibrium probability distribution.
- The Gibbs–Shannon information entropy equals the sub-system entropy. Markov expansions for lattice and continuum systems in various dimensions facilitate its evaluation and are more accurate than direct approaches.
- The time correlation function for two phase functions can be obtained as a time average over the stochastic, dissipative trajectory.

References

Attard P 1999 Markov superposition expansion for the entropy and correlation functions in two and three dimensions *Statistical Physics on the Eve of the Twenty-First Century* ed Batchelor M T (ed) and Wille L T (ed) (Singapore: World Scientific)

Baxter R J 1982 *Exactly Solved Models in Statistical Mechanics* (London: Academic)

Boltzmann L 1871 Über das wärmegleichgewicht zwischen mehratomigen gasmolekule *Wien. Ber.* **63** 397 679 712

Boltzmann L 1877 On the relation of a general mechanical theorem to the second law of thermodynamics *Irreversible Processes: Kinetic Theory* ed Brush S G (ed) (Oxford: Pergamon) p 191

Galea T M and Attard P 2002 Constraint method for deriving non-equilibrium molecular dynamics equations of motion *Phys. Rev.* E **66** 041207

Hansen J-P and McDonald I R 1990 *Theory of Simple Liquids* 2nd edn (London: Academic)

Münster A 1969 *Statistical Thermodynamics* **vol 1** (Berlin: Springer)

Pathria R K 1972 *Statistical Mechanics* (Oxford: Pergamon)

Planck M 1949 *Scientific Autobiography, and Other Papers in Gaynor* (New York: Philosophical Library) p 33 (English translation by F Gaynor)

Wikipedia 2017 *Square-Lattice Ising Model* (accessed 8 April 2017) https://en.wikipedia.org/wiki/Square_lattice_Ising_model

Entropy Beyond the Second Law (Second Edition)

Thermodynamics and statistical mechanics for equilibrium, non-equilibrium, classical, and quantum systems

Phil Attard

Chapter 6

Non-equilibrium statistical mechanics and path entropy

'Time that is moved by little fidget wheels / Is not my time, the flood that does not flow'

K Slessor (1939)

'It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience'

A Einstein (1933)

'Medicine is a philosophy and this is not compatible with the renouncement of criticism of the leading authors'

al-Razi (Weinberg 2015 page 111)

The main aim of this chapter is to develop non-equilibrium classical statistical mechanics by deriving the general form for the probability density in phase space for a thermodynamic or a mechanical non-equilibrium system.

The need for such a first principles derivation is no better illustrated than by an expression often found in books on statistical mechanics that is meant to give the phase space probability density for a non-equilibrium system with Hamiltonian that varies with time,

$$\wp(\mathbf{\Gamma}, t) \propto e^{-\beta \mathcal{H}(\mathbf{\Gamma}, t)}.$$

(6.1)

Although this expression has the great virtue of being simple, it has the significant drawback of being wrong. The chapter begins by showing why this expression goes awry.

6.1 Reservoir entropy

There are two generic types of non-equilibriums systems, mechanical and thermodynamic. A mechanical non-equilibrium system has an explicitly time-dependent Hamiltonian, $\mathcal{H}(\mathbf{\Gamma}, t)$. Typically, the time variation is due to a time-varying external potential that is applied to the sub-system. Usually the sub-system is in thermal contact with a heat reservoir.

A thermodynamic non-equilibrium system is typically a sub-system across which there is a thermodynamic gradient applied by two spatially separated reservoirs with different values of certain field variables. In this case there is a steady flux of the conjugate exchangeable material from one reservoir to the other through the sub-system. Steady heat flow that was treated in chapter 3 is an example of thermodynamic non-equilibrium system.

Both of these generic non-equilibrium systems are sensitive to the direction of time, either because the external potential is explicitly time dependent, or else because of the steady induced flux that increases the entropy of the total system with time. This time sensitivity means that the probability distribution for the molecular configurations must be sensitive to the molecular velocities, since reversing time is equivalent to reversing all the

velocities. Recall the definition of the conjugation operation, namely that each point in the phase space of the particles' positions and momenta, $\mathbf{\Gamma} \equiv \{\mathbf{q}, \mathbf{p}\}$, has a conjugate point with all the momenta reversed, $\mathbf{\Gamma}^\dagger \equiv \{\mathbf{q}, -\mathbf{p}\}$. The fact that by definition a non-equilibrium system must be sensitive to time means that the phase space point and its conjugate must have different probabilities,

$$\wp(\mathbf{\Gamma}, t) \neq \wp(\mathbf{\Gamma}^\dagger, t).$$

(6.2)

This is a formally exact requirement that must hold for both mechanical and thermodynamic non-equilibrium systems.

However, the kinetic energy in the time-dependent Hamiltonian is a quadratic function of velocities. Hence it is unchanged by their reversal,

$$\mathcal{H}(\mathbf{\Gamma}, t) = \mathcal{H}(\mathbf{\Gamma}^\dagger, t).$$

(6.3)

This means that the Maxwell–Boltzmann probability distribution that is the simplistic equation (6.1) cannot be the probability distribution for any non-equilibrium system.

6.1.1 Trajectory entropy

In chapter 5, it was shown that the weight density of a point in the phase space of an isolated system was uniform, which means that its entropy vanishes. (The constant contributions, Planck's constant h^{3N} for the volume of a quantum microstate, and $N!$ for distinct microstates, can be defined as the phase point entropy, or else they can be shown explicitly, as will be done here.) This result was derived from the adiabatic equations of motion for an isolated system. Since the same Hamilton's equations hold for the isolated sub-system even when the Hamiltonian is time-dependent, one can assume that the same result holds in the non-equilibrium case. That is, the internal entropy of a point in the phase space of the isolated sub-system vanishes.

In view of this, for the case of a sub-system that can exchange with a reservoir, the entropy of a sub-system phase space point $\mathbf{\Gamma}$ is just the entropy of the reservoir that is associated with that point at that time, $S_\mathrm{r}(\mathbf{\Gamma}, t)$. Hence the non-equilibrium phase space probability density is

$$\wp(\mathbf{\Gamma}, t) = \frac{\mathrm{e}^{S_\mathrm{r}(\mathbf{\Gamma}, t)/k_\mathrm{B}}}{h^{3N} N! Z(t)}.$$

(6.4)

This result is formally exact but practically useless until an explicit expression for the reservoir entropy can be derived.

In order to be definite, and to give the clearest and most transparent derivation, we shall henceforth assume the canonical mechanical non-equilibrium system until it is stated otherwise. The sub-system Hamiltonian is $\mathcal{H}(\mathbf{\Gamma}, t)$, and the sub-system is assumed to be able to exchange energy with a heat reservoir of temperature T.

Consider a specific trajectory in the sub-system, $\mathbf{\Gamma}(t)$. Let us take this to be a forward trajectory from $t = 0$, which is convenient but not essential. This contains both adiabatic, superscript 0, and reservoir, $\mathbf{R}$, influences, as is sketched in figure 6.1. The rate of change of the trajectory is

$$\dot{\mathbf{\Gamma}}(t) = \dot{\mathbf{\Gamma}}^0(t) + \dot{\mathbf{R}}(t).$$

(6.5)
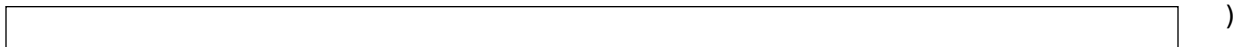
Figure 6.1. Sketch of a trajectory in the sub-system phase space. The arrows represent the force at intervals, with the solid tangential arrows being the total force, the dashed arrows being the adiabatic force, and the dotted arrows being the reservoir force.

This is meant to be an actual trajectory, and the reservoir forces are the real forces that are necessary to give the specified departure from the adiabatic trajectory. For this reason we do not have to distinguish between forward and backward time derivatives. This is in contrast to the cases discussed with reference to equation (4.22) and to figure 5.5, where the statistical reservoir contributions to the trajectory are irreversible and create a discontinuity in the time derivative.

The rate of change of energy of the sub-system on the trajectory is

$$
\begin{aligned}
\frac{\mathrm{d}\mathcal{H}(\boldsymbol{\Gamma}(t),t)}{\mathrm{d}t} &= \frac{\partial\mathcal{H}(\boldsymbol{\Gamma},t)}{\partial t} + \dot{\boldsymbol{\Gamma}}(t) \cdot \nabla\mathcal{H}(\boldsymbol{\Gamma},t) \\
&= \frac{\partial\mathcal{H}(\boldsymbol{\Gamma},t)}{\partial t} + \dot{\mathbf{R}}(t) \cdot \nabla\mathcal{H}(\boldsymbol{\Gamma},t).
\end{aligned}
$$

(6.6)

The adiabatic velocity does not contribute to the change in energy. The final term is the reservoir-induced rate of change of energy. By energy conservation, this is equal and opposite to the rate of change of the reservoir energy. Hence the rate of change of the reservoir entropy on this particular trajectory is

$$
\dot{S}_{\mathrm{r}}(\boldsymbol{\Gamma}(t),t) = \frac{-1}{T}\dot{\mathbf{R}}(t) \cdot \nabla\mathcal{H}(\boldsymbol{\Gamma}(t),t).
$$

(6.7)

Finally, the trajectory entropy, which is the change in reservoir entropy over the particular trajectory is

$$
\begin{aligned}
S_{\mathrm{r}}([\boldsymbol{\Gamma}(t)],t) &= \frac{-1}{T}\int_0^t \mathrm{d}t\prime\, \dot{\mathbf{R}}(t\prime) \cdot \nabla\mathcal{H}(\boldsymbol{\Gamma}(t\prime),t\prime) \\
&= \frac{-1}{T}\int_0^t \mathrm{d}t\prime\, \left[\dot{\mathcal{H}}(\boldsymbol{\Gamma}(t\prime),t\prime) - \dot{\mathcal{H}}^0(\boldsymbol{\Gamma}(t\prime),t\prime)\right] \\
&= \frac{-1}{T}\left[\mathcal{H}(\boldsymbol{\Gamma}(t),t) - \mathcal{H}(\boldsymbol{\Gamma}(0),0)\right] + \frac{1}{T}\int_0^t \mathrm{d}t\prime\, \frac{\partial\mathcal{H}(\boldsymbol{\Gamma}(t\prime),t\prime)}{\partial t\prime}.
\end{aligned}
$$

(6.8)

It is convenient but not essential to insist that this is a forward trajectory and that $t > 0$.

The first term on the right-hand side of the final equality is the instantaneous or static reservoir entropy at the end of the trajectory less than that at the beginning. This static part of the entropy is the entropy based solely on the instantaneous structure of the sub-

system, which is the entropy that would occur if the system were an equilibrium system. In the present mechanical case the static entropy is defined as

$$S_{\mathrm{r,st}}(\mathbf{\Gamma},t) = \frac{-1}{T}\mathcal{H}(\mathbf{\Gamma},t).$$

(6.9)

The second term on the right-hand side may be called the dynamic part of the entropy. It is the correction to the static entropy that arises from the non-equilibrium nature of the system. It subtracts the sum total of the change of the static part of the reservoir entropy that is due to adiabatic processes,

$$\begin{aligned}
S_{\mathrm{r,dyn}}([\mathbf{\Gamma}(t)],t) &= \frac{1}{T}\int_0^t \mathrm{d}t\prime\, \frac{\partial\mathcal{H}(\mathbf{\Gamma}(t\prime),t\prime)}{\partial t\prime}\\
&= \frac{1}{T}\int_0^t \mathrm{d}t\prime\, \dot{\mathcal{H}}^0(\mathbf{\Gamma}(t\prime),t\prime)\\
&= -\int_0^t \mathrm{d}t\prime\, \dot{S}_{\mathrm{r,st}}^0(\mathbf{\Gamma}(t\prime),t\prime).
\end{aligned}$$

(6.10)

The integral of the first equality may be recognized as the work done by the external potential on the system over the particular trajectory. With these the trajectory entropy is

$$\begin{aligned}
S_{\mathrm{r}}([\mathbf{\Gamma}(t)],t) &= S_{\mathrm{r,st}}(\mathbf{\Gamma},t) + S_{\mathrm{r,dyn}}([\mathbf{\Gamma}(t)],t)\\
&= S_{\mathrm{r,st}}(\mathbf{\Gamma},t) - \int_0^t \mathrm{d}t\prime\, \dot{S}_{\mathrm{r,st}}^0(\mathbf{\Gamma}(t\prime),t\prime).
\end{aligned}$$

(6.11)

This neglects the static entropy at the start of the trajectory, $S_{\mathrm{r,st}}(\mathbf{\Gamma}(0),0)$, since for a long enough trajectory this is a constant that is uncorrelated with the end point.

Subtracting the second term, the dynamic part of the entropy, ensures that only reservoir-induced changes in energy contribute to the reservoir trajectory entropy. One virtue of writing the trajectory entropy this way is that the non-equilibrium part is explicit and can easily be treated as a perturbation about the equilibrium part, if desired.

A second virtue of writing the trajectory entropy this way is is that it holds formally unchanged for a thermodynamic non-equilibrium system. For example, in the case of steady heat that was treated in section 3.4, the static part of the entropy is just

$$S_{\mathrm{r,st}}(\mathbf{\Gamma}) = \frac{-E_0(\mathbf{\Gamma})}{T_0} - \frac{E_1(\mathbf{\Gamma})}{T_1},$$

(6.12)

and its adiabatic rate of change is

$$\dot{S}_{\mathrm{r,st}}^0(\mathbf{\Gamma}) = \frac{-1}{T_1}\dot{E}_1^0(\mathbf{\Gamma}) = \frac{-1}{T_1}\dot{\mathbf{\Gamma}}^0\cdot\nabla E_1(\mathbf{\Gamma}).$$

(6.13)

The larger point is that for any non-equilibrium system, it is trivial to write explicitly the expression for the static part of the reservoir entropy. It is defined as the instantaneous reservoir entropy that changes directly from exchanges between the sub-system and the reservoir.

From this point on, it is no longer assumed that the analysis is restricted to a mechanical non-equilibrium system.

6.1.2 Point entropy

What is required to formulate non-equilibrium statistical mechanics is the reservoir entropy associated with a phase point $\mathbf{\Gamma}$, whereas the above expression is for the reservoir entropy of a particular trajectory $\mathbf{\Gamma}(t)$. The obvious way to proceed is to neglect fluctuations and to focus on the most likely backward trajectory from the current point $\mathbf{\Gamma}$, namely $\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma},t)$, $t\prime \leqslant t$. (It is also possible to invoke instead the most likely forward trajectory from the current point.) With this the reservoir entropy for $\mathbf{\Gamma}$ for the non-equilibrium system is

$$S_{\mathrm{r}}(\mathbf{\Gamma},t) = S_{\mathrm{r,st}}(\mathbf{\Gamma},t) + S_{\mathrm{r,dyn}}(\mathbf{\Gamma},t),$$

(6.14)

where the dynamic part of the entropy on the most likely backward trajectory is

$$S_{\mathrm{r,dyn}}(\mathbf{\Gamma},t) \equiv S_{\mathrm{r,dyn}}([\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma},t)],t)$$
$$= -\int_0^t \mathrm{d}t\prime\, \dot{S}_{\mathrm{r,st}}^0(\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma},t),t\prime).$$

(6.15)

Note that although the most likely trajectory has a discontinuous first time derivative, the adiabatic time derivative on this backward most likely trajectory is continuous, and so there is no distinction between the forward and backward adiabatic time derivatives.

This result formally gives the explicit expression for the phase space non-equilibrium probability density, equation (6.4). Details for the most likely trajectory in the non-equilibrium system are derived in section 6.2.2.

6.1.2.1 Asymptote

The integrand of the dynamic entropy has asymptote

$$\dot{S}_{\mathrm{r,st}}^0(\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma},t),t\prime) \to \overline{\dot{S}}_{\mathrm{r,st}}^0(t\prime), \quad |\,t\prime{-}t\,| \to \infty.$$

(6.16)

This says that with overwhelming probability the system reverts back to the most likely rate of adiabatic entropy production no matter how far the current point $\mathbf{\Gamma}$ is from a likely point in phase space.

With this one can write

$$S_{\mathrm{r,dyn}}(\mathbf{\Gamma},t) = \bar{S}_{\mathrm{r,dyn}}(t) - \int_0^t \mathrm{d}t\prime\, \left[\dot{S}_{\mathrm{r,st}}^0(\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma},t),t\prime) - \overline{\dot{S}}_{\mathrm{r,st}}^0(t\prime)\right],$$

(6.17)

where the constant is $\bar{S}_{\mathrm{r,dyn}}(t) \equiv -\int_0^t \mathrm{d}t\prime\, \overline{\dot{S}}_{\mathrm{r,st}}^0(t\prime)$. The integrand of the dynamic entropy is now short-ranged and so it is insensitive to the value chosen for the lower limit, which facilitates its numerical evaluation.

6.1.2.2 Forward trajectory

As was mentioned above, the dynamic part of the entropy can equally well be written as an integral over the forward trajectory as the backward one,

$$S_{\mathrm{r,dyn}}(\mathbf{\Gamma}, t) = - \int_{-\infty}^{t} dt\prime \; \dot{S}_{\mathrm{r,st}}^{0}(\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma}, t), t\prime)$$

$$= - \int_{t}^{\infty} dt\prime \; \dot{S}_{\mathrm{r,st}}^{0}(\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma}, t), t\prime).$$

These two expression may differ by an immaterial term that is constant in phase space.

6.1.3 Changes in entropy

6.1.3.1 Reservoir-induced change in reservoir entropy

In section 6.2.2 we shall derive the stochastic, dissipative equations of motion for the non-equilibrium system. In view of the equilibrium result one can anticipate that they are of the form

$$\mathbf{\Gamma}_2 = \mathbf{\Gamma}_1 + \Delta_t \dot{\mathbf{\Gamma}}^{0} + \mathbf{R}_p(\Delta_t, \mathbf{\Gamma}, t).$$

As usual, the superscript 0 denotes the adiabatic velocity and the reservoir contribution effects only the momenta over a single infinitesimal time step.

One can rewrite the result for the reservoir entropy directly in terms of its change over the most likely trajectory. With the initial most likely point being $\mathbf{\Gamma}_0 \equiv \bar{\mathbf{\Gamma}}(0\,|\,\mathbf{\Gamma}, t)$, the change in reservoir entropy is

$$
\begin{aligned}
S_{\mathrm{r}}(\mathbf{\Gamma}, t) - S_{\mathrm{r}}(\mathbf{\Gamma}_0, 0) &= S_{\mathrm{r,st}}(\mathbf{\Gamma}, t) - S_{\mathrm{r,st}}(\mathbf{\Gamma}_0, 0) - \int_{0}^{t} dt\prime \; \dot{S}_{\mathrm{r,st}}^{0}(\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma}, t), t\prime) \\
&= \int_{0}^{t} dt\prime \; \left\{ \dot{S}_{\mathrm{r,st}}^{-}(\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma}, t), t\prime) - \dot{S}_{\mathrm{r,st}}^{0}(\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma}, t), t\prime) \right\} \\
&= \int_{0}^{t} dt\prime \; \dot{\bar{\mathbf{R}}}_{p}^{-}(\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma}, t), t\prime) \cdot \nabla_p S_{\mathrm{r,st}}(\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma}, t), t\prime).
\end{aligned}
$$

On both sides of the first equality, the previously neglected constant static entropy at the start of the trajectory has been reinserted. Because the first time derivative is discontinuous on the most likely trajectory, it is necessary to specify that it is the backward derivative that appears here. One can see that the integrand is entirely the reservoir-induced change in reservoir entropy. That the reservoir force $\mathbf{R}_p$ couples to $\nabla_p S_{\mathrm{r,st}}$ defines the static reservoir entropy as the instantaneous reservoir entropy that changes directly from exchanges between the sub-system and the reservoir.

One can also give an explicit expression for the change in reservoir entropy on the most likely backward trajectory, $\Delta_t < 0$. Since the integral for the dynamic entropy is precisely over this most likely backward trajectory, by the fundamental theorem of calculus one has

$$
\begin{aligned}
&S_{\mathrm{r}}(\bar{\mathbf{\Gamma}}(t + \Delta_t \,|\, \mathbf{\Gamma}, t), t) - S_{\mathrm{r}}(\mathbf{\Gamma}, t) \\
&= \Delta_t \dot{S}_{\mathrm{r,st}}^{0}(\mathbf{\Gamma}, t) + \bar{\mathbf{R}}_p(\mathbf{\Gamma}, t, \Delta_t) \cdot \nabla_p S_{\mathrm{r,st}}(\mathbf{\Gamma}, t) - \Delta_t \dot{S}_{\mathrm{r,st}}^{0}(\mathbf{\Gamma}, t) \\
&= \bar{\mathbf{R}}_p(\mathbf{\Gamma}, t, \Delta_t) \cdot \nabla_p S_{\mathrm{r,st}}(\mathbf{\Gamma}, t), \quad \Delta_t < 0.
\end{aligned}
$$

Notice that the adiabatic contributions cancel. Again given the definition of $\mathbf{R}_p$ as the reservoir forces, and the definition of the static entropy as the reservoir entropy instantaneously effected by exchange with the sub-system, this result is physically sensible.

Formally, the rate of change of the reservoir entropy on a trajectory is

$$
\begin{aligned}
\frac{\mathrm{d}S_\mathrm{r}(\mathbf{\Gamma},t)}{\mathrm{d}t} &= \frac{\partial S_\mathrm{r}(\mathbf{\Gamma},t)}{\partial t} + \dot{\mathbf{\Gamma}}^0 \cdot \nabla S_\mathrm{r}(\mathbf{\Gamma},t) + \frac{1}{\Delta_t}\mathbf{R}_p \cdot \nabla S_\mathrm{r}(\mathbf{\Gamma},t) \\
&= \dot{S}_\mathrm{r}^0(\mathbf{\Gamma},t) + \frac{1}{\Delta_t}\mathbf{R}_p \cdot \nabla_p S_\mathrm{r}(\mathbf{\Gamma},t).
\end{aligned}
$$

(6.22)

The first term on the right-hand side of the final equality is the adiabatic rate of change of the reservoir entropy. Evaluating the total rate of change on the most likely backward trajectory and using the above result this is

$$
\begin{aligned}
\dot{S}_\mathrm{r}^0(\mathbf{\Gamma},t) &= \frac{1}{\Delta_t}\bar{\mathbf{R}}_p(\mathbf{\Gamma},t,\Delta_t) \cdot \nabla_p S_\mathrm{r,st}(\mathbf{\Gamma},t) - \frac{1}{\Delta_t}\bar{\mathbf{R}}_p(\mathbf{\Gamma},t,\Delta_t) \cdot \nabla_p S_\mathrm{r}(\mathbf{\Gamma},t) \\
&= \frac{-1}{\Delta_t}\bar{\mathbf{R}}_p(\mathbf{\Gamma},t,\Delta_t) \cdot \nabla_p S_\mathrm{r,dyn}(\mathbf{\Gamma},t), \quad \Delta_t < 0.
\end{aligned}
$$

(6.23)

The dynamic part of the reservoir entropy depends upon the most likely backward trajectory. In the next section, section 6.2, the stochastic dissipative equations that give this are derived. In section 6.3 the dynamic part of the entropy is reformulated in a more computationally efficient form, which includes utilizing adiabatic trajectories, and the connection of this to the Green–Kubo relations is elucidated.

6.1.3.2 Rate of change of total entropy

The partition function normalizes the probability distribution, equation (6.4),

$$
Z(t) = \frac{1}{N!h^{3N}} \int \mathrm{d}\mathbf{\Gamma}\ \mathrm{e}^{S_\mathrm{r}(\mathbf{\Gamma},t)/k_\mathrm{B}}.
$$

(6.24)

Its logarithm gives the total unconstrained entropy of the total system,

$$
S_\mathrm{tot}(t) = k_\mathrm{B} \ln Z(t).
$$

(6.25)

The rate of change of the total entropy is

$$
\begin{aligned}
\dot{S}_\mathrm{tot}(t) &= \frac{k_\mathrm{B}\dot{Z}(t)}{Z(t)} \\
&= \frac{k_\mathrm{B}}{Z(t)} \frac{1}{N!h^{3N}} \int \mathrm{d}\mathbf{\Gamma}\ \mathrm{e}^{S_\mathrm{r}(\mathbf{\Gamma},t)/k_\mathrm{B}} \frac{\partial S_\mathrm{r}(\mathbf{\Gamma},t)}{k_\mathrm{B}\partial t} \\
&= \left\langle \frac{\partial S_\mathrm{r}(\mathbf{\Gamma},t)}{\partial t} \right\rangle.
\end{aligned}
$$

(6.26)

Let $\boldsymbol{\Gamma}_2 = \boldsymbol{\Gamma}^0(t_2 \mid \boldsymbol{\Gamma}_1, t_1) = \boldsymbol{\Gamma}_1 + t_{21}\dot{\boldsymbol{\Gamma}}_1^0$ be the adiabatic evolution of $\boldsymbol{\Gamma}_1$ over an infinitesimal time step. From the incompressibility of phase space under Hamilton's equations of motion one has

$$
\begin{aligned}
S_{\text{tot}}(t_2) &= k_{\text{B}}\,\ln\,\frac{1}{N!h^{3N}}\int d\boldsymbol{\Gamma}_2\,\, e^{S_{\text{r}}(\boldsymbol{\Gamma}_2,t_2)/k_{\text{B}}} \\
&= k_{\text{B}}\,\ln\,\frac{1}{N!h^{3N}}\int d\boldsymbol{\Gamma}_1\,\, e^{\left\{S_{\text{r}}(\boldsymbol{\Gamma}_1,t_1)+t_{21}\dot{S}_{\text{r}}^0(\boldsymbol{\Gamma}_1,t_1)\right\}/k_{\text{B}}} \\
&= k_{\text{B}}\,\ln\,Z(t_1)\int d\boldsymbol{\Gamma}_1\,\, \wp(\boldsymbol{\Gamma}_1,t_1)\left\{1+t_{21}\dot{S}_{\text{r}}^0(\boldsymbol{\Gamma}_1,t_1)/k_{\text{B}}\right\} \\
&= S_{\text{tot}}(t_1) + t_{21}\left\langle \dot{S}_{\text{r}}^0(\boldsymbol{\Gamma},t_1)\right\rangle.
\end{aligned}
$$

(6.27)

Hence the rate of change of the total entropy is also given by

$$
\dot{S}_{\text{tot}}(t) = \left\langle \dot{S}_{\text{r}}^0(\boldsymbol{\Gamma},t)\right\rangle.
$$

(6.28)

Subtracting these two expressions for $\dot{S}_{\text{tot}}(t)$ yields

$$
\left\langle \dot{\boldsymbol{\Gamma}}^0 \cdot \nabla S_{\text{r}}(\boldsymbol{\Gamma},t)\right\rangle = 0.
$$

(6.29)

Because these are the averages of an extensive variable, one can expect fluctuations to be relatively negligible. In this case these should hold locally, on points in phase space that are likely to occur. Hence to a good approximation,

$$
\dot{S}_{\text{r}}^0(\boldsymbol{\Gamma},t) \approx \frac{\partial S_{\text{r}}(\boldsymbol{\Gamma},t)}{\partial t} \approx \dot{S}_{\text{tot}}(t),
$$

(6.30)

and

$$
\dot{\boldsymbol{\Gamma}}^0 \cdot \nabla S_{\text{r}}(\boldsymbol{\Gamma},t) \approx 0.
$$

(6.31)

6.2 Stochastic, dissipative equations of motion

6.2.1 Foundations for the transition probability in the time-varying case

Following Attard (2012, section 8.3.1), the set theoretic formulation of weight, entropy, and probability is now given in the case of time-varying systems. This extends the equilibrium formalities given in sections 1.1.2, 3.1.3, and 3.1.4. Although the transition analysis of section 3.1.4 was applied to the non-equilibrium system of steady heat flow, in a sense steady-state systems are a special non-equilibrium system in which the sub-system does not change macroscopically with time. The analysis given here removes that restriction.

Suppose that the weights depend on time. These are microstate, $\omega(i,t)$, macrostate, $\omega(\alpha,t) = \sum_{i\in\alpha}\omega(i,t)$, and total $W(t) = \sum_i \omega(i,t) = \sum_\alpha \omega(\alpha,t)$. This is sketched in figure 6.2 The entropies are the logarithm of these: microstate, $S(i,t) = k_{\text{B}}\,\ln\,\omega(i,t)$, macrostate, $S(\alpha,t) = k_{\text{B}}\,\ln\,\omega(\alpha,t)$, and total $S(t) = k_{\text{B}}\,\ln\,W(t)$. Again as usual, the

probabilities are proportional to the weight, and hence to the exponential of the entropy, $\wp(i,t) = \omega(i,t)/W(t) = \exp[S(i,t)/k_{\mathrm{B}}]/W(t)$ and $\wp(\alpha,t) = \omega(\alpha,t)/W(t) = \exp[S(\alpha,t)/k_{\mathrm{B}}]/W(t)$.

Figure 6.2. Sketch of the evolution of a collective of macrostates in a time-varying system. The weight of a macrostate is proportional to the size of its cell.
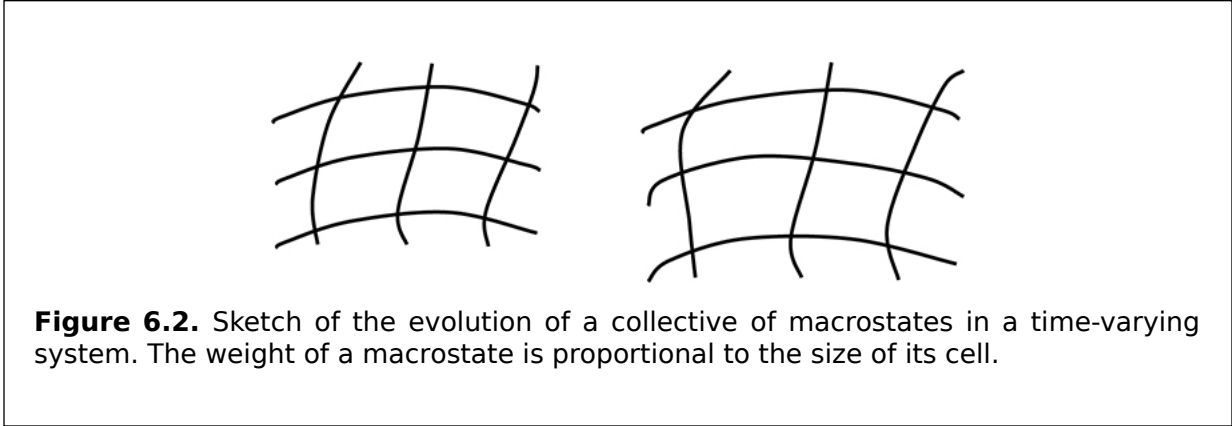
The macrostate transition, $\{\alpha,t\} \to \{\gamma,t\prime\}$, has weight $\omega(\gamma,t\prime;\alpha,t)$. There is similarly a weight for microstate transitions, $\omega(j,t\prime;k,t)$. From statistical symmetry the unconditional weight is unchanged by swapping the order of the arguments, $\omega(\gamma,t\prime;\alpha,t) = \omega(\alpha,t;\gamma,t\prime)$. This is the weight attached to the system being in the macrostate α at time t *and* in the macrostate γ at time $t\prime$.

For an equilibrium system two conservation laws for weight were established. The first was for the weight of a system being jointly in two macrostates, equation (3.9), $\sum_\gamma \omega(\alpha,\gamma) = \omega(\alpha)$. The second was for the system making a transition between two macrostates, equation (3.17), $\sum_\gamma \omega(\alpha,\gamma\,|\,\tau) = \omega(\alpha)$, and $\sum_\alpha \omega(\alpha,\gamma\,|\,\tau) = \omega(\gamma)$. These may also be called reduction conditions (of the first type), although they are slightly different to the reduction condition (of the second type) given in equation (3.38).

Fundamentally the basis of both of these conservation laws is the fact that the macrostates of a given collective are disjoint and form a complete set. One therefore expects similar conservation conditions to hold in the present time-varying case. One has to modify the conservation condition on the transition weight to account for the fact that since the total weight changes with time, $W(t)$, it cannot be conserved in a transition. For this reason perhaps it is better to call this a reduction condition of the first type, or else just a sum rule for the transition weight.

This variation in the total weight can be accounted for by including a time-dependent scale factor in the sum rule,

$$\sum_\gamma \omega(\gamma,t\prime;\alpha,t) = f(t\prime,t)\omega(\alpha,t).$$

(6.32)

With this the total weight of the transition is

$$\sum_{\alpha,\gamma} \omega(\gamma,t\prime;\alpha,t) = f(t\prime,t)W(t).$$

(6.33)

For this to be symmetric in both times the scale factor must be of the form

$$f(t\prime,t) = W(t\prime)^a W(t)^{a-1}.$$

This must give the equilibrium result in the time-independent case, namely $f(t\prime, t)$ which means that $a = 1/2$, and that $f(t\prime, t) = \sqrt{W(t\prime)/W(t)}$. That is, the sum rule for the transition weight in a time-varying system is

$$\sum_{\gamma} \omega(\gamma, t\prime; \alpha, t) = \sqrt{\frac{W(t\prime)}{W(t)}} \omega(\alpha, t),$$

and the total weight of the transition is

$$\sum_{\alpha\gamma} \omega(\gamma, t\prime; \alpha, t) = \sqrt{W(t\prime)W(t)}.$$

This is just the geometric mean of the total weights at the two termini of the transition, which has a pleasing symmetry.

In the time-independent case of equation (3.17), the weight was conserved as it was distributed amongst the target states in a transition. In the present time-dependent case, the weight of the initial state α is scaled and distributed amongst all the target states γ. This scaling is necessary whenever the weights of the states and the total weight change with time.

It is clear that a similar scaling has to occur for the sum over initial states,

$$\sum_{\alpha} \omega(\gamma, t\prime; \alpha, t) = \sqrt{\frac{W(t)}{W(t\prime)}} \omega(\gamma, t\prime).$$

This says that the weight of the target macrostate γ must have come from somewhere, and been scaled. The laws of probability require these two sum rules in the time-varying case for its internal consistency.

The unconditional transition probability has the total transition weight as the normalizing factor,

$$\wp(\gamma, t\prime; \alpha, t) = \frac{\omega(\gamma, t\prime; \alpha, t)}{\sqrt{W(t\prime)W(t)}}.$$

As usual, the conditional transition probability is

$$\wp(\gamma, t\prime | \alpha, t) = \frac{\wp(\gamma, t\prime; \alpha, t)}{\wp(\alpha, t)}$$

$$= \frac{\omega(\gamma, t\prime; \alpha, t)}{\omega(\alpha, t)} \sqrt{\frac{W(t)}{W(t\prime)}}$$

$$= \frac{\omega(\gamma, t\prime; \alpha, t)}{\sum_{\gamma} \omega(\gamma, t\prime; \alpha, t)}.$$

The first equality is just Bayes' theorem, $\wp(\gamma, t\prime; \alpha, t) = \wp(\gamma, t\prime | \alpha, t)\wp(\alpha, t)$.

As in the equilibrium case, the second entropy is defined to be essentially the logarithm of the transition weight, $S^{(2)}(\gamma, t\prime; \alpha, t) = k_{\mathrm{B}} \ln \omega(\gamma, t\prime; \alpha, t)$. In terms of it the transition probability is

$$\wp(\gamma, t\prime; \alpha, t) = \frac{e^{S^{(2)}(\gamma, t\prime; \alpha, t)/k_{\mathrm{B}}}}{\sqrt{W(t\prime)W(t)}}.$$

(6.40)

The conditional transition probability follows as

$$\wp(\gamma, t\prime|\, \alpha, t) = \sqrt{\frac{W(t)}{W(t\prime)}}\, e^{[S^{(2)}(\gamma, t\prime; \alpha, t) - S(\alpha, t)]/k_{\mathrm{B}}}.$$

(6.41)

This is evidently normalized to unity when summed over γ. The conditional probability refers to a forward transition if $t\prime > t$, which is to say that tells where the system will go to in the future given the current state. If $t\prime < t$ it refers to a backward transition, which tells where the system came from in the past to get to the current state.

The important equilibrium reduction condition of the second type, equation (3.38), also has to be modified slightly for the time-dependent case. As previously, define the conditionally most likely target state $\bar{\gamma} \equiv \bar{\gamma}(t\prime|\, \alpha, t)$ as the state that maximizes the second entropy,

$$\frac{\partial S^{(2)}(\gamma, t\prime; \alpha, t)}{\partial \gamma}\Big|_{\gamma = \bar{\gamma}} = 0.$$

(6.42)

The reduction condition of the second type says that the maximal value of the conditional second entropy equals the entropy of the initial state α. That is,

$$S(\alpha, t) = k_{\mathrm{B}} \ln \omega(\alpha, t)$$
$$= k_{\mathrm{B}} \ln \left[\sqrt{\frac{W(t)}{W(t\prime)}} \sum_{\gamma} \omega(\gamma, t\prime; \alpha, t)\right]$$
$$\approx k_{\mathrm{B}} \ln \left[\sqrt{\frac{W(t)}{W(t\prime)}} \omega(\bar{\gamma}, t\prime; \alpha, t)\right]$$
$$= S^{(2)}(\bar{\gamma}, t\prime; \alpha, t) + \frac{1}{2}[S(t) - S(t\prime)].$$

(6.43)

Since one expects the probability distribution to be sharply peaked in any macroscopic physical system, one can approximate the logarithm of a sum over states by the logarithm of the largest term in the sum, as in the third equality. This is equivalent to assuming that the fluctuations about the most likely state are relatively negligible. Here $S(t) = k_{\mathrm{B}} \ln W(t)$ is the total unconstrained entropy. This second type of reduction condition can be re-written as

$$S^{(2)}(\bar{\gamma}, t\prime; \alpha, t) = S(\alpha, t) + \frac{1}{2}[S(t\prime) - S(t)].$$

It is obvious that the reduction condition holds for both forward and backward transitions, and so one also has

$$S^{(2)}(\gamma, t\prime; \bar{\alpha}, t) = S(\gamma, t\prime) + \frac{1}{2}[S(t) - S(t\prime)],$$

(6.4 4)

(6.4 5)

where $\bar{\alpha} \equiv \bar{\alpha}(t \mid \gamma, t\prime)$. Of course the order of the arguments in the second entropy is irrelevant since the times at which each state occurs are stated explicitly. In the time-independent case, equation (3.38), the maximal value of the second entropy equals the first entropy alone. In the present time-varying non-equilibrium system, the maximal value of the second entropy equals the first entropy plus half the difference in the total entropy of the two states.

6.2.2 Second entropy and the transition probability

In previous work (Attard 2012, section 8.2.3 and section 8.3.2; Attard 2014), I derived the stochastic, dissipative equations of motion for a non-equilibrium system from a fluctuation form for the second entropy. The analysis that is now given improves upon this earlier work in three respects. First, the adiabatic contributions were originally treated as fluctuating quantities within the mixed parity second entropy formulation, whereas here the pure parity slave formulation used for Brownian motion, equation (4.21), and for a phase space trajectory in the canonical equilibrium system, equation (5.50), is invoked. Second, the original analysis was based on fluctuations about the most likely point in phase space, $\bar{\Gamma}(t)$, aspects of which may be questioned. Fluctuations play a much smaller role in the present analysis, and a more direct physical interpretation of various quantities is given. And third, the present derivation is about one quarter the length of the original, which makes for greater simplicity and transparency, and which leaves less room for error. As the quote from Einstein given at the head of the chapter has been paraphrased: a theory should be as simple as possible, but no simpler.

We shall take the equations of motion in phase space over an infinitesimal time step to be of the form

$$\Gamma_q(t + \Delta_t) = \Gamma_q(t) + \Delta_t \dot{\Gamma}_q^0(\Gamma, t),$$
$$\Gamma_p(t + \Delta_t) = \Gamma_p(t) + \Delta_t \dot{\Gamma}_p^0(\Gamma, t) + \mathbf{R}_p(\Gamma, t, \Delta_t).$$

(6.4 6)

As usual the superscript 0 represents the adiabatic velocity,

$$\dot{\Gamma}_q^0(\Gamma, t) = \frac{\partial \mathcal{H}(\Gamma, t)}{\partial \Gamma_p}, \quad \text{and} \quad \dot{\Gamma}_p^0(\Gamma, t) = \frac{-\partial \mathcal{H}(\Gamma, t)}{\partial \Gamma_q}.$$

(6.4 7)

The force due to the reservoir $\mathbf{R}_p$ only effects the momenta directly. In this formulation the position coordinates are treated as slaves to the momenta, as was done in the case of the Langevin equation for Brownian motion, equation (4.21), and in the case of phase space for the canonical equilibrium system, equation (5.50). The reservoir force can be divided into a deterministic, dissipative part, $\bar{\mathbf{R}}_p$, and a stochastic part of zero mean $\tilde{\mathbf{R}}_p$,

$$\mathbf{R}_p = \bar{\mathbf{R}}_p + \tilde{\mathbf{R}}_p.$$

(6.4 8)

The explicit form for the former will shortly be determined from the second entropy form of the transition probability.

We assume that the probability distribution for the fluctuations in the present non-equilibrium system has the same time symmetry as for an equilibrium system, namely that the progression and regression of a fluctuation are equally likely. This means that the stochastic reservoir force is Gaussian distributed, with variance proportional to the magnitude of the time step,

$$\wp(\tilde{\mathbf{R}}_p) = \frac{\mathrm{e}^{-\Lambda^{-1}:\tilde{\mathbf{R}}_p\tilde{\mathbf{R}}_p/2k_{\mathrm{B}}|\Delta_t|}}{\sqrt{\mathrm{Det}\{2\pi k_{\mathrm{B}} \mid \Delta_t \mid \Lambda\}}}.$$

(6.49)

Evidently the random force has zero mean, $\langle \tilde{\mathbf{R}} \rangle = 0$, and its variance is

$$\left\langle \tilde{\mathbf{R}}(t)\,\tilde{\mathbf{R}}(t) \right\rangle = \mid \Delta_t \mid k_{\mathrm{B}}\Lambda.$$

(6.50)

Random forces at different time steps are uncorrelated. It would be possible to have a time-dependent variance, $\Lambda(t)$, but this is not pursued here. Typically, and most simply, one may take the fluctuation matrix to be diagonal, $\Lambda = \lambda \mathrm{I}_{\mathrm{pp}}$. In some cases (e.g. boundary driven flow) there may be value in making the parameter position dependent.

In summary, the equations of motion for the transition $\boldsymbol{\Gamma}_1 \overset{\Delta_t}{\rightarrow} \boldsymbol{\Gamma}_2$ are

$$\boldsymbol{\Gamma}_2 = \boldsymbol{\Gamma}_1 + \Delta_t \dot{\boldsymbol{\Gamma}}^0 + \bar{\mathbf{R}}_p + \tilde{\mathbf{R}}_p.$$

(6.51)

The total reservoir force, $\mathbf{R}(t)$, for the present transition, $\boldsymbol{\Gamma}_1 \overset{\Delta_t}{\rightarrow} \boldsymbol{\Gamma}_2$, is equal and opposite to that for the opposite transition, $\boldsymbol{\Gamma}_2 \overset{-\Delta_t}{\longrightarrow} \boldsymbol{\Gamma}_1$. The same is true for the adiabatic force. The two individual components of the total reservoir force, the most likely and the stochastic, are not individually equal and opposite their respective counterpart.

The transition probability is just the exponential of the second entropy, equation (6.40),

$$\wp(\boldsymbol{\Gamma}_2, t_2; \boldsymbol{\Gamma}_1, t_1) = \frac{\mathrm{e}^{S^{(2)}(\boldsymbol{\Gamma}_2, t_2; \boldsymbol{\Gamma}_1, t_1)/k_{\mathrm{B}}}}{\sqrt{Z(t_2)Z(t_1)}}$$
$$= \wp(\boldsymbol{\Gamma}_2, t_2 \mid \boldsymbol{\Gamma}_1, t_1)\,\wp(\boldsymbol{\Gamma}_1, t_1).$$

(6.52)

The second equality writes this as the conditional probability times the probability of the initial state, which is Bayes' theorem, equation (6.39). The conditional transition probability is just the probability of the random force, $\wp(\boldsymbol{\Gamma}_2, t_2 \mid \boldsymbol{\Gamma}_1, t_1) = \wp(\tilde{\mathbf{R}}_p)$, since $\boldsymbol{\Gamma}_2$ is randomly distributed about $\bar{\boldsymbol{\Gamma}}(t_2 \mid \boldsymbol{\Gamma}_1, t_1)$. Hence the unconditional probability of the transition may also be written

$$\wp(\boldsymbol{\Gamma}_2, t_2; \boldsymbol{\Gamma}_1, t_1) = \wp(\boldsymbol{\Gamma}_2, t_2 \mid \boldsymbol{\Gamma}_1, t_1)\,\wp(\boldsymbol{\Gamma}_1, t_1)$$
$$= \frac{\mathrm{e}^{-\Lambda^{-1}:\tilde{\mathbf{R}}_p\tilde{\mathbf{R}}_p/2k_{\mathrm{B}}|\Delta_t|}}{\sqrt{\mathrm{Det}\{2\pi k_{\mathrm{B}} \mid \Delta_t \mid \Lambda\}}}\,\frac{\mathrm{e}^{S_{\mathrm{r}}(\boldsymbol{\Gamma}_1, t_1)/k_{\mathrm{B}}}}{Z(t_1)}.$$

(6.53)

Writing the exponent in this form is equivalent to the reduction condition of the second type, equation (6.44).

Now we determine the most likely reservoir force. Unlike in chapter 3, we do not further use the reduction condition of the second type, equation (6.44), because this is for discrete states and it does not account for the compressibility of phase space. Instead we invoke the first type of reduction condition, equation (6.37), which involves an integral over the initial phase space transition point. This is equivalent to deriving the Fokker–Planck equation for the evolution of the probability density, as is discussed in section 6.2.4.

From the reduction condition of the first type, equation (6.37), one has

$$
\begin{aligned}
\wp(\mathbf{\Gamma}_2, t_2) \;=\;& \int d\mathbf{\Gamma}_1 \; \wp(\mathbf{\Gamma}_2, t_2; \mathbf{\Gamma}_1, t_1) \\[2mm]
=\;& \int d\bar{\mathbf{\Gamma}}_2 \; \Big| \frac{d\mathbf{\Gamma}_1}{d\bar{\mathbf{\Gamma}}_2} \Big| \frac{e^{-\Lambda^{-1} : \tilde{\mathbf{R}}_p \tilde{\mathbf{R}}_p / 2 k_{\mathrm B} |\Delta_t|}}{\sqrt{\mathrm{Det}\{2\pi k_{\mathrm B} \mid \Delta_t \mid \Lambda\}}} \frac{e^{S_{\mathrm r}(\mathbf{\Gamma}_1, t_1)/k_{\mathrm B}}}{Z(t_1)} \\[2mm]
=\;& \Big| \frac{d\mathbf{\Gamma}_1}{d\bar{\mathbf{\Gamma}}_2} \Big| \frac{e^{S_{\mathrm r}(\mathbf{\Gamma}_2, t_1)/k_{\mathrm B}}}{Z(t_1)} \\[2mm]
& \times \int d\bar{\mathbf{\Gamma}}_2 \; \frac{e^{-\Lambda^{-1} : \tilde{\mathbf{R}}_p \tilde{\mathbf{R}}_p / 2 k_{\mathrm B} |\Delta_t|}}{\sqrt{\mathrm{Det}\{2\pi k_{\mathrm B} \mid \Delta_t \mid \Lambda\}}} \Big\{ 1 + \frac{1}{k_{\mathrm B}} [\mathbf{\Gamma}_1 - \mathbf{\Gamma}_2] \cdot \nabla S_{\mathrm r} \\[2mm]
& + \frac{1}{2 k_{\mathrm B}} [\mathbf{\Gamma}_1 - \mathbf{\Gamma}_2]^2 : \nabla \nabla S_{\mathrm r} + \frac{1}{2 k_{\mathrm B}^2} ([\mathbf{\Gamma}_1 - \mathbf{\Gamma}_2] \cdot \nabla S_{\mathrm r})^2 \Big\} \\[2mm]
=\;& \frac{e^{S_{\mathrm r}(\mathbf{\Gamma}_2, t_2)/k_{\mathrm B}}}{Z(t_2)} \Big\{ 1 - t_{21} \nabla \cdot \dot{\mathbf{\Gamma}}^{\mathrm{det}} \Big\} \Big\{ 1 - t_{21} \frac{\partial S_{\mathrm r}(\mathbf{\Gamma}, t)}{k_{\mathrm B} \partial t} + t_{21} \frac{\dot{Z}(t)}{Z(t)} \Big\} \\[2mm]
& \times \Big\{ 1 - \frac{t_{21}}{k_{\mathrm B}} \dot{\mathbf{\Gamma}}^{\mathrm{det}} \cdot \nabla S_{\mathrm r} + \frac{1}{2 k_{\mathrm B}} k_{\mathrm B} \mid t_{21} \mid \Lambda : \nabla_p \nabla_p S_{\mathrm r} \\[2mm]
& + \frac{1}{2 k_{\mathrm B}^2} k_{\mathrm B} \mid t_{21} \mid \Lambda : [\nabla_p S_{\mathrm r}][\nabla_p S_{\mathrm r}] \Big\}.
\end{aligned}
$$

(6.54)

This uses $\tilde{\mathbf{R}}_p = \mathbf{\Gamma}_2 - \bar{\mathbf{\Gamma}}_2$, $\bar{\mathbf{\Gamma}}_2 \equiv \bar{\mathbf{\Gamma}}(t_2 \mid \mathbf{\Gamma}_1, t_1) = \mathbf{\Gamma}_1 + t_{21} \dot{\mathbf{\Gamma}}^{\mathrm{det}}$, $|d\mathbf{\Gamma}_1 / d\bar{\mathbf{\Gamma}}_2| = \{1 - t_{21} \nabla \cdot \dot{\mathbf{\Gamma}}^{\mathrm{det}}\}$, and $S_{\mathrm r}(\mathbf{\Gamma}_1, t_1) = S_{\mathrm r}(\mathbf{\Gamma}_2, t_1) + [\mathbf{\Gamma}_1 - \mathbf{\Gamma}_2] \cdot \nabla S_{\mathrm r} + [\mathbf{\Gamma}_1 - \mathbf{\Gamma}_2]^2 : \nabla \nabla S_{\mathrm r}/2$. Only terms to linear order in t_{21} and in Λ have been kept. This obviously places an upper limit on their magnitude in practical computations involving the stochastic, dissipative equations of motion.

Since the first factor on the right-hand side of the final equality is $\wp(\mathbf{\Gamma}_2, t_2)$, the product of the three remaining factors must be unity to linear order in the time step. Hence the sum of the terms proportional to the time step must vanish,

$$
\begin{aligned}
0 =\;& - t_{21} \dot{S}_{\mathrm r}^0(\mathbf{\Gamma}, t) + t_{21} \dot{S}_{\mathrm{tot}}(t) - t_{21} k_{\mathrm B} \nabla \cdot \dot{\mathbf{\Gamma}}^{\mathrm{det}} \\[2mm]
& - \bar{\mathbf{R}}_p \cdot \nabla S_{\mathrm r} + \frac{k_{\mathrm B} \mid t_{21} \mid}{2} \Lambda : \nabla_p \nabla_p S_{\mathrm r} + \frac{\mid t_{21} \mid}{2} \Lambda : [\nabla_p S_{\mathrm r}][\nabla_p S_{\mathrm r}],
\end{aligned}
$$

(6.55)

where $S_r \equiv S_r(\boldsymbol{\Gamma}, t)$. Recall that the total unconstrained entropy is $S_{\text{tot}}(t) = k_B \ln Z(t)$, and that its rate of change is $\dot{S}_{\text{tot}}(t) = k_B \dot{Z}(t)/Z(t)$. This is equivalent to the Fokker–Planck equation, as is discussed in section 6.2.4.

The deterministic velocity is the sum of the adiabatic and the most likely reservoir-induced velocity,

$$\dot{\boldsymbol{\Gamma}}^{\text{det}}(\hat{t}_{21}, \boldsymbol{\Gamma}, t) = \dot{\boldsymbol{\Gamma}}^0(\boldsymbol{\Gamma}, t) + \frac{1}{t_{21}} \bar{\mathbf{R}}_p(t_{21}, \boldsymbol{\Gamma}, t).$$

(6.56)

Here and below, $\hat{t}_{21} \equiv \text{sign } t_{21}$. The adiabatic compressibility vanishes, $\nabla \cdot \dot{\boldsymbol{\Gamma}}^0 = 0$. Since the above equation contains both reversible, $\propto t_{21}$, and irreversible $\propto |t_{21}|$ terms, it is evident that the most likely reservoir force should be similarly decomposed,

$$\bar{\mathbf{R}}_p(t_{21}, \boldsymbol{\Gamma}, t) = |t_{21}| \, \mathbf{a}(\boldsymbol{\Gamma}, t) + t_{21} \mathbf{b}(\boldsymbol{\Gamma}, t).$$

(6.57)

This only has momentum components. According to the pure parity analysis equation (3.86) or equation (4.19) or equation (4.38), only the irreversible term survives, so that $\mathbf{b}(\boldsymbol{\Gamma}, t) = 0$. Strictly speaking, however, position coordinates are also present, which makes this a mixed parity case, and so the reversible term may be nonzero (Attard 2012, section 8.3.4). Using this and equating the irreversible terms to zero in the above yields

$$0 = -k_B \nabla_p \cdot \mathbf{a}(\boldsymbol{\Gamma}, t) - \mathbf{a}(\boldsymbol{\Gamma}, t) \cdot \nabla_p S_r(\boldsymbol{\Gamma}, t)$$
$$+ \frac{k_B}{2} \Lambda : \nabla_p \nabla_p S_r(\boldsymbol{\Gamma}, t) + \frac{1}{2} \Lambda : [\nabla_p S_r(\boldsymbol{\Gamma}, t)] \, [\nabla_p S_r(\boldsymbol{\Gamma}, t)].$$

(6.58)

By inspection, this has the exact solution

$$\mathbf{a}(\boldsymbol{\Gamma}, t) = \frac{1}{2} \Lambda \nabla_p S_r(\boldsymbol{\Gamma}, t)$$
$$= \frac{1}{2} \Lambda \nabla_p S_{r,\text{st}}(\boldsymbol{\Gamma}, t) + \frac{1}{2} \Lambda \nabla_p S_{r,\text{dyn}}(\boldsymbol{\Gamma}, t).$$

(6.59)

The reversible terms must satisfy

$$\dot{S}_r^0(\boldsymbol{\Gamma}, t) - \dot{S}_{\text{tot}}(t) = -k_B \nabla_p \cdot \mathbf{b}(\boldsymbol{\Gamma}, t) - \mathbf{b}(\boldsymbol{\Gamma}, t) \cdot \nabla_p S_r(\boldsymbol{\Gamma}, t).$$

(6.60)

By equation (6.30) (see also equation (6.115)), the left-hand side is approximately zero, and so one expects $\mathbf{b}(\boldsymbol{\Gamma}, t)$ to be small or negligible.

The exact solution for $\mathbf{a}$ is difficult to work with, primarily because of the challenging nature of $S_{r,\text{dyn}}$. In contrast, $S_{r,\text{st}}$ is always relatively trivial to manipulate and to calculate. In the time honored tradition of physics, one simply discards such difficult terms, which gives the zeroth order approximation,

$$\mathbf{a}^{(0)}(\boldsymbol{\Gamma}, t) = \frac{1}{2} \Lambda \nabla_p S_{r,\text{st}}(\boldsymbol{\Gamma}, t), \quad \mathbf{b}^{(0)}(\boldsymbol{\Gamma}, t) = 0.$$

(6.61)

The first of these is equivalent to $\nabla_p S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t) = 0$. Of course in so far as the non-equilibrium contributions are expected to be a small perturbation on the instantaneous static structure of the system, $\mid S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t) \mid \ll \mid S_{\mathrm{r,st}}(\boldsymbol{\Gamma}, t) \mid$, one can expect this zeroth order approximation to be relatively accurate.

To do better than this, define the most likely configuration as the one that maximizes the reservoir entropy,

$$\nabla S_{\mathrm{r}}(\boldsymbol{\Gamma}, t)\big|_{\boldsymbol{\Gamma} = \bar{\boldsymbol{\Gamma}}(t)} = 0.$$

(6.62)

Equivalently (see figure 6.3),

$$\nabla S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t) = -\nabla S_{\mathrm{r,dyn}}(\bar{\boldsymbol{\Gamma}}(t), t).$$

(6.63)

The abbreviated notation here means that the gradient is evaluated at the most likely configuration. We shall not address issues such as the multiplicity of solutions to this, other than to say that $\bar{\boldsymbol{\Gamma}}(t)$ is to be interpreted as the zero of the reservoir entropy gradient that is closest to $\boldsymbol{\Gamma}$ at time t. With this, instead of discarding completely the dynamic part of the reservoir entropy, one can approximate it by its value at the most likely point,

$$\nabla_p S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t) \approx \nabla_p S_{\mathrm{r,dyn}}(\bar{\boldsymbol{\Gamma}}(t), t) = -\nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t).$$

(6.64)

This approximation can be expected to be accurate in so far as departures from the most likely configuration are small. The utility of the most likely configuration is that it allows the dynamic entropy to be replaced by the static entropy, which is relatively easy to evaluate. It is not essential for what follows, but it may be a reasonable to approximate this gradient at the most likely configuration by the most likely gradient, $\nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t) \approx \overline{\mathbf{S}\prime_{\mathrm{r,st}}}(t)$.

Figure 6.3. Sketch of the reservoir entropy (solid curve), and its static (dashed curve) and dynamic (dotted curve) parts in the vicinity of the most likely point in phase space at that time.
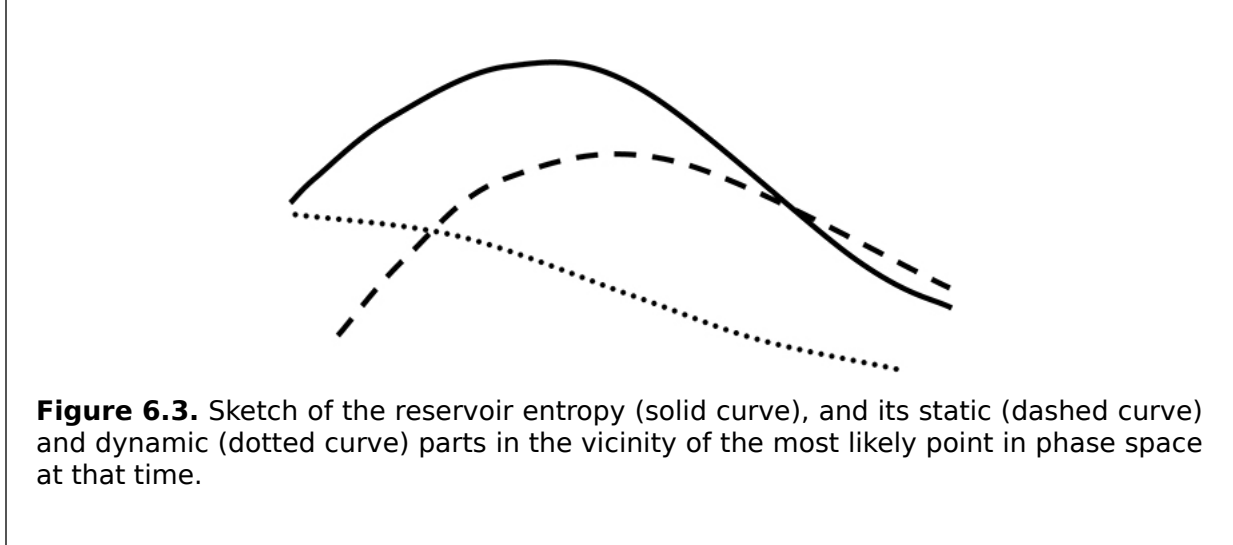
With this the more refined expression for the irreversible contribution to the reservoir force is

$$\mathbf{a}(\boldsymbol{\Gamma}, t) = \frac{1}{2}\Lambda\nabla_p S_{\mathrm{r,st}}(\boldsymbol{\Gamma}, t) - \frac{1}{2}\Lambda\nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t).$$

The most likely configuration $\bar{\mathbf{\Gamma}}(t)$ evolves in time. This evolution can be expected to be given by the dissipative equations of motion, $\bar{\mathbf{\Gamma}}(t\prime|\,\bar{\mathbf{\Gamma}}(t),t) = \bar{\mathbf{\Gamma}}(t\prime)$. Since this is assumed to be a single trajectory, this evolution must be reversible and given by the coefficient $\mathbf{b}(\mathbf{\Gamma},t)$. Hence the more refined expression for the reversible contribution to the reservoir force is to set it equal to the zeroth approximation $\mathbf{a}^{(0)}$ evaluated at the most likely configuration,

$$\mathbf{b}(\mathbf{\Gamma},t) = \mathbf{a}^{(0)}(\bar{\mathbf{\Gamma}}(t),t) = \frac{1}{2}\Lambda\nabla_p S_{\mathrm{r,st}}(\bar{\mathbf{\Gamma}}(t),t).$$

In the present approximation this is equivalent to $\mathbf{b}(\mathbf{\Gamma},t) \approx -\Lambda\nabla_p S_{\mathrm{r,dyn}}(\mathbf{\Gamma},t)/2$, which point is discussed in connection with equation (6.72). Since the dynamic part of the entropy is a small perturbation, this is consistent with the above observation that $\mathbf{b}(\mathbf{\Gamma},t) \approx 0$.

With these, the most likely reservoir force is

$$\bar{\mathbf{R}}_p = \frac{|\,\Delta_t\,|}{2}\Lambda\left[\nabla_p S_{\mathrm{r,st}}(\mathbf{\Gamma},t) - \nabla_p S_{\mathrm{r,st}}(\bar{\mathbf{\Gamma}}(t),t)\right] + \frac{\Delta_t}{2}\Lambda\nabla_p S_{\mathrm{r,st}}(\bar{\mathbf{\Gamma}}(t),t)$$

$$= \begin{cases} \dfrac{|\,\Delta_t\,|}{2}\Lambda\nabla_p S_{\mathrm{r,st}}(\mathbf{\Gamma},t), & \Delta_t > 0, \\[2mm] \dfrac{|\,\Delta_t\,|}{2}\Lambda\nabla_p S_{\mathrm{r,st}}(\mathbf{\Gamma},t) -\,|\,\Delta_t\,|\,\Lambda\nabla_p S_{\mathrm{r,st}}(\bar{\mathbf{\Gamma}}(t),t), & \Delta_t < 0. \end{cases}$$

Notice that on a forward trajectory, the most likely configuration is not required. Further, one can see by inspection that on the most likely configuration, $\mathbf{\Gamma} = \bar{\mathbf{\Gamma}}(t)$, the most likely reservoir force is purely reversible, $\bar{\mathbf{R}}_p(\Delta_t, \bar{\mathbf{\Gamma}}(t),t) = \Delta_t\Lambda\nabla_p S_{\mathrm{r,st}}(\bar{\mathbf{\Gamma}}(t),t)/2$.

The extra term, $(\Delta_t -\,|\,\Delta_t\,|)\nabla_p S_{\mathrm{r,st}}(\bar{\mathbf{\Gamma}}(t),t)/2$, compared to the zeroth order approximation, (see equation (6.74)), is a constant in phase space. Hence it can play no role in determining the relative probability of points in phase space. See equation (6.171) for a discussion of this in the context of a detailed calculation of driven Brownian motion.

One can look at the consequences of the full result and check a number of consistency requirements. On the most likely backward trajectory, the change in reservoir entropy is given exactly by the reservoir force times the gradient in the static part of the reservoir entropy,

$$\Delta S_{\mathrm{r}} = \bar{\mathbf{R}}_p(-\,|\,\Delta_t\,|, \mathbf{\Gamma},t) \cdot \nabla_p S_{\mathrm{r,st}}(\mathbf{\Gamma},t)$$

$$= \left(\frac{|\,\Delta_t\,|}{2}\Lambda\nabla_p S_{\mathrm{r,st}}(\mathbf{\Gamma},t) -\,|\,\Delta_t\,|\,\Lambda\nabla_p S_{\mathrm{r,st}}(\bar{\mathbf{\Gamma}}(t),t)\right) \cdot \nabla_p S_{\mathrm{r,st}}(\mathbf{\Gamma},t).$$

But formally one also has the Taylor expansion

$$\Delta S_{\mathrm{r}} = -\,|\,\Delta_t\,|\,\dot{S}_{\mathrm{r}}^{0}(\mathbf{\Gamma},t) + \bar{\mathbf{R}}_p(-\,|\,\Delta_t\,|, \mathbf{\Gamma},t) \cdot \nabla_p S_{\mathrm{r}}(\mathbf{\Gamma},t).$$

Subtracting these and rearranging yields

$$\dot{S}_{\mathrm{r}}^{0}(\boldsymbol{\Gamma}, t) = \frac{1}{\mid \Delta_t \mid} \bar{\mathbf{R}}_p(- \mid \Delta_t \mid, \boldsymbol{\Gamma}, t) \cdot \nabla_p S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t)$$

$$= \left(\frac{1}{2} \Lambda \nabla_p S_{\mathrm{r,st}}(\boldsymbol{\Gamma}, t) - \Lambda \nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t) \right) \cdot \nabla_p S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t)$$

$$\approx \frac{-1}{2} \Lambda : [\nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t)] \, [\nabla_p S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t)]$$

$$\approx \frac{1}{2} \Lambda : [\nabla_p S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t)] \, [\nabla_p S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t)]$$

$$\approx \frac{1}{2} \Lambda : [\nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t)] \, [\nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t)].$$

(6.70)

The third equality approximates the gradient of the static part of the reservoir entropy at an arbitrary configuration by its value at the most likely configuration, $\nabla_p S_{\mathrm{r,st}}(\boldsymbol{\Gamma}, t) \approx \nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t)$, which is reasonable since one expects fluctuations to be small. The next two approximations invoke the above result for the dynamic part of the entropy, $\nabla_p S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t) \approx -\nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t)$.

Now equation (6.30), shows that the adiabatic rate of entropy production, the left-hand side, can be approximated by the time-dependent constant in phase space, namely the rate of total entropy production, $\dot{S}_{\mathrm{r}}^{0}(\boldsymbol{\Gamma}, t) \approx \dot{S}_{\mathrm{tot}}(t)$. (See also equation (6.115).) In view of the final equality, also a time-dependent constant in phase space, since Λ is positive definite, one concludes that the rate of total entropy production is positive, as it must be. One sees therefore a certain consistency in the above expression for the most likely reservoir force. Equation (6.60) gave the exact equation that the reversible coefficient had to satisfy. Inserting the present expression for $\mathbf{b}(\boldsymbol{\Gamma}, t)$ into this yields

$$\dot{S}_{\mathrm{r}}^{0}(\boldsymbol{\Gamma}, t) - \dot{S}_{\mathrm{tot}}(t) = - k_{\mathrm{B}} \nabla_p \cdot \mathbf{b}(\boldsymbol{\Gamma}, t) - \mathbf{b}(\boldsymbol{\Gamma}, t) \cdot \nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma}, t)$$

$$= \frac{-k_{\mathrm{B}}}{2} \nabla_p \cdot \left[\Lambda \nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t) \right]$$

$$- \frac{1}{2} \Lambda \nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t) \cdot [\nabla_p S_{\mathrm{r,st}}(\boldsymbol{\Gamma}, t) + \nabla_p S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t)]$$

$$\approx \frac{-1}{2} \Lambda \nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t) \cdot \left[\nabla_p S_{\mathrm{r,st}}(\boldsymbol{\Gamma}, t) - \nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t) \right]$$

$$\approx 0.$$

(6.71)

The final two equalities invoke the most likely gradients, $\nabla_p S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t) \approx \nabla_p S_{\mathrm{r,dyn}}(\bar{\boldsymbol{\Gamma}}(t), t) = -\nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t)$, and $\nabla_p S_{\mathrm{r,st}}(\boldsymbol{\Gamma}, t) \approx \nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t)$. One sees again that this is consistent with equation (6.30), which is also based upon the concept that fluctuations are relatively negligible in the thermodynamic limit.

The above results were based on the approximation that the reversible coefficient was given by $\mathbf{b}(\boldsymbol{\Gamma}, t) = \Lambda \nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t)/2 \approx -\Lambda \nabla_p S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t)/2$. One might speculate that the final approximation is in fact the exact result. Combining this with the exact result for the irreversible coefficient, $\mathbf{a}(\boldsymbol{\Gamma}, t) = \Lambda \nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma}, t)/2$ this would mean the reservoir force is exactly given by

$$\bar{\mathbf{R}}_p = \frac{|\Delta_t|}{2}\Lambda\nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}, t) - \frac{\Delta_t}{2}\Lambda\nabla_p S_{\mathrm{r,dyn}}(\mathbf{\Gamma}, t)$$

$$= \begin{cases} \dfrac{|\Delta_t|}{2}\Lambda\nabla_p S_{\mathrm{r,st}}(\mathbf{\Gamma}, t), & \Delta_t > 0, \\[2ex] \dfrac{|\Delta_t|}{2}\Lambda\nabla_p S_{\mathrm{r,st}}(\mathbf{\Gamma}, t) + |\Delta_t|\,\Lambda\nabla_p S_{\mathrm{r,dyn}}(\mathbf{\Gamma}, t), & \Delta_t < 0. \end{cases}$$

(6.72)

This has a certain beauty to it, which may signify truth. Continuing with the hypothesis, the exact condition equation (6.60) becomes

$$\dot{S}_{\mathrm{r}}^0(\mathbf{\Gamma}, t) - \dot{S}_{\mathrm{tot}}(t)$$
$$= \frac{k_{\mathrm{B}}}{2}\Lambda : \nabla_p\nabla_p S_{\mathrm{r,dyn}}(\mathbf{\Gamma}, t) + \frac{1}{2}\Lambda : [\nabla_p S_{\mathrm{r,dyn}}(\mathbf{\Gamma}, t)]\,[\nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}, t)].$$

(6.73)

The left-hand side of this equation is approximately zero, equation (6.30). (See also equation (6.115).)

In any case, using either the approximate result or the putative exact result, the dynamic entropy gradient is a second order contribution that can be neglected. As such, the most likely reservoir force to leading order is

$$\bar{\mathbf{R}}_p \approx \frac{|\Delta_t|}{2}\Lambda\nabla_p S_{\mathrm{r,st}}(\mathbf{\Gamma}, t).$$

(6.74)

To this leading order the dissipative force is independent of the sign of the time step, $\bar{\mathbf{R}}_p^{21} = \bar{\mathbf{R}}_p^{12}$, where the superscript represents the direction of the transition. This makes it fully irreversible, just as in the equilibrium case. In fact it has the same functional form as in the equilibrium case, with the gradient of the static entropy providing the driving force toward the optimum non-equilibrium state. This zeroth order approximation is tested for driven Brownian motion in equation (6.171), and it is found to correctly give the relative entropy and hence relative probability of points in phase space.

The irreversible contributions break the symmetry of time reversibility, which is essential for consistency with the Second Law of Thermodynamics. The approximation to the gradient of the dynamic part of the entropy provides a correction to the zeroth order expression so that the reservoir force has both reversible and irreversible components. The fact that Λ gives the magnitude of the dissipation and the variance of the stochastic force shows that the fluctuation–dissipation theorem holds also for a non-equilibrium system.

Ultimately the approximations invoked in deriving these stochastic, dissipative equations of motion for the non-equilibrium system, should be judged by the results of their application to specific systems. It can be mentioned that they have been tested by computer simulation for both mechanical (Attard and Gray-Weale 2008, Attard 2009a, 2009b) and thermodynamic (Attard 2006b, 2009a) non-equilibrium systems. They are further tested analytically below with the derivation of the behavior of a driven Brownian particle, section 6.4.

6.2.3 Generalized equipartition theorem

It is of interest to discuss the result for the most likely reservoir force a little more generally in the context of the generalized equipartition theorem for a non-equilibrium system (cf section 4.4). Since the probability density and its gradient must vanish at the extremes of phase space, one must have that

$$\int d\mathbf{\Gamma} \; \nabla \nabla \wp(\mathbf{\Gamma}, t) = 0.$$

(6.7
5)

This implies that

$$\langle \nabla \nabla S_{\mathrm{r}}(\mathbf{\Gamma}, t) \rangle = -k_{\mathrm{B}}^{-1} \langle [\nabla S_{\mathrm{r}}(\mathbf{\Gamma}, t)] \, [\nabla S_{\mathrm{r}}(\mathbf{\Gamma}, t)] \rangle.$$

(6.7
6)

This is a generalized equipartition theorem for a non-equilibrium system.

Taking the trace, or the double scalar product with a matrix, makes these the average of an extensive quantity. One can assume that in these cases they hold locally on the likely points in phase space. That is

$$\nabla^2 S_{\mathrm{r}}(\mathbf{\Gamma}, t) \approx -k_{\mathrm{B}}^{-1} [\nabla S_{\mathrm{r}}(\mathbf{\Gamma}, t)] \cdot [\nabla S_{\mathrm{r}}(\mathbf{\Gamma}, t)],$$

(6.7
7)

and

$$\Lambda : \nabla_p \nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}, t) \approx -k_{\mathrm{B}}^{-1} \Lambda : [\nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}, t)] \, [\nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}, t)].$$

(6.7
8)

In the thermodynamic limit these are likely exact.

Applying the second of these to equation (6.58), one would deduce that

$$0 \approx -k_{\mathrm{B}} \nabla_p \cdot \mathbf{a}(\mathbf{\Gamma}, t) - \mathbf{a}(\mathbf{\Gamma}, t) \cdot \nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}, t).$$

(6.7
9)

This implies that $\mathbf{a}(\mathbf{\Gamma}, t)$ must be small, if not zero. This is consistent with the result invoked above, $\mathbf{a}(\mathbf{\Gamma}, t) = \Lambda[\nabla_p S_{\mathrm{r,st}}(\mathbf{\Gamma}, t) - \nabla_p S_{\mathrm{r,st}}(\bar{\mathbf{\Gamma}}(t), t)]/2$, equation (6.65), which is indeed zero on the most likely points in phase space.

6.2.4 The Fokker–Planck equation

Inserting the second of these into equation (6.4), one deduces that

$$
\begin{aligned}
\frac{\partial^{\pm} \wp(\mathbf{\Gamma}, t)}{\partial t} &\equiv \left\{ \frac{\partial S_{\mathrm{r}}(\mathbf{\Gamma}, t)}{k_{\mathrm{B}} \partial t} - \frac{\dot{Z}(t)}{Z(t)} \right\} \wp(\mathbf{\Gamma}, t) \\
&= \left\{ -\nabla \cdot \dot{\mathbf{\Gamma}}^{\mathrm{det}} - \dot{\mathbf{\Gamma}}^{\mathrm{det}} \cdot \nabla S_{\mathrm{r}}/k_{\mathrm{B}} \right. \\
&\quad \left. + \frac{\hat{\Delta}_t}{2} \Lambda : \nabla_p \nabla_p S_{\mathrm{r}} + \frac{\hat{\Delta}_t}{2k_{\mathrm{B}}} \Lambda : [\nabla_p S_{\mathrm{r}}][\nabla_p S_{\mathrm{r}}] \right\} \wp(\mathbf{\Gamma}, t) \\
&= -[\nabla \cdot \dot{\mathbf{\Gamma}}^{\mathrm{det}}] \wp(\mathbf{\Gamma}, t) - \dot{\mathbf{\Gamma}}^{\mathrm{det}} \cdot \nabla \wp(\mathbf{\Gamma}, t) + \frac{\hat{\Delta}_t k_{\mathrm{B}}}{2} \Lambda : \nabla_p \nabla_p \wp(\mathbf{\Gamma}, t).
\end{aligned}
$$

(6.8
0)

The second equality is equation (6.55) rearranged. This is the Fokker–Planck equation. Since the stochastic, dissipative equations of motion for a non-equilibrium system have essentially the same functional form as those for an equilibrium system, it should not be surprising that this is identical in form to the equilibrium result, equation (5.68).

The superscript ± on the left-hand side is formally required because there are irreversible terms (those proportional to $\hat{\Delta}_t \equiv \mathrm{sign} \, \Delta_t = \pm 1$, both explicit and implicit in $\dot{\mathbf{\Gamma}}^{\mathrm{det}}$) on the right-hand side. However, these irreversible terms must cancel with each

other because they are just the irreversible terms in equation (6.58). This makes the superscript ± on the left-hand side redundant. (It is redundant for the exact equations; the approximate ansatz will be discussed below.)

With the reversible part of the deterministic velocity that remains, $\dot{\boldsymbol{\Gamma}}^{\mathrm{det,rev}} = \dot{\boldsymbol{\Gamma}}^0 + \mathbf{b}$, the Fokker–Planck equation becomes

$$\frac{\partial \wp(\boldsymbol{\Gamma}, t)}{\partial t} \equiv \left\{ \frac{\partial S_{\mathrm{r}}(\boldsymbol{\Gamma}, t)}{k_{\mathrm{B}} \partial t} - \frac{\dot{Z}(t)}{Z(t)} \right\} \wp(\boldsymbol{\Gamma}, t)$$
$$= -[\nabla \cdot \mathbf{b}] \wp(\boldsymbol{\Gamma}, t) - \left[\dot{\boldsymbol{\Gamma}}^0 + \mathbf{b} \right] \cdot \nabla \wp(\boldsymbol{\Gamma}, t).$$

(6.81)

The formally exact expression for the reversible reservoir contribution, equation (6.60), makes this an identity.

Now let us analyze the Fokker–Planck equation for the the approximation to the most likely reservoir force, equation (6.67). In this case the irreversible part of the equation is

$$\frac{1}{2} \left\{ \frac{\partial^+ \wp(\boldsymbol{\Gamma}, t)}{\partial t} - \frac{\partial^- \wp(\boldsymbol{\Gamma}, t)}{\partial t} \right\}$$

$$= -[\nabla \cdot \mathbf{a}(\boldsymbol{\Gamma}, t)] \wp(\boldsymbol{\Gamma}, t) - \mathbf{a}(\boldsymbol{\Gamma}, t) \cdot \nabla \wp(\boldsymbol{\Gamma}, t) + \frac{k_{\mathrm{B}}}{2} \Lambda : \nabla_p \nabla_p \wp(\boldsymbol{\Gamma}, t)$$

$$= \left\{ \frac{-1}{2} \Lambda : \nabla_p \nabla_p S_{\mathrm{r,st}}(\boldsymbol{\Gamma}, t) \right.$$

$$- \frac{1}{2k_{\mathrm{B}}} \Lambda : [\nabla_p S_{\mathrm{r,st}}(\boldsymbol{\Gamma}, t) - \nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t)] \, [\nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma}, t)]$$

$$+ \frac{1}{2} \Lambda : \nabla_p \nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma}, t) + \frac{1}{2k_{\mathrm{B}}} \Lambda : [\nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma}, t)] \, [\nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma}, t)] \right\} \wp(\boldsymbol{\Gamma}, t)$$

$$= \left\{ \frac{1}{2} \Lambda : \nabla_p \nabla_p S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t) \right.$$

$$+ \frac{k_{\mathrm{B}}^{-1}}{2} \Lambda : [\nabla_p S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t) - \nabla_p S_{\mathrm{r,dyn}}(\bar{\boldsymbol{\Gamma}}(t), t)][\nabla_p S_{\mathrm{r}}(\boldsymbol{\Gamma}, t)] \right\} \wp(\boldsymbol{\Gamma}, t).$$

(6.82)

This is evidently nonzero, although one might argue that it is small if not negligible. In this approximation the reversible part of the deterministic velocity is

$$\dot{\boldsymbol{\Gamma}}^{\mathrm{det,rev}} = \dot{\boldsymbol{\Gamma}}^0 + \frac{1}{2} \Lambda \nabla_p S_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t), t).$$

(6.83)

Since the reservoir part is a constant in phase space, one has that $\nabla \cdot \dot{\boldsymbol{\Gamma}}^{\mathrm{det,rev}} = 0$, and the reversible part of the Fokker–Planck equation in this case reduces to

$$\frac{\partial \wp(\mathbf{\Gamma}, t)}{\partial t} = -\dot{\mathbf{\Gamma}}^{\mathrm{det,rev}} \cdot \nabla \wp(\mathbf{\Gamma}, t)$$

$$= -\left[\dot{\mathbf{\Gamma}}^0 + \frac{1}{2} \Lambda \nabla_p S_{\mathrm{r,st}}(\bar{\mathbf{\Gamma}}(t), t)\right] \cdot \nabla \wp(\mathbf{\Gamma}, t).$$

(6.8
4)

To the extent that equation (6.71) is satisfied, this is the expected result.

The hypothetical exact result for the most likely reservoir force, equation (6.72), gives for the irreversible part of the Fokker–Planck equation as

$$\frac{1}{2}\left\{\frac{\partial^+ \wp(\mathbf{\Gamma}, t)}{\partial t} - \frac{\partial^- \wp(\mathbf{\Gamma}, t)}{\partial t}\right\}$$

$$= -[\nabla \cdot \mathbf{a}(\mathbf{\Gamma}, t)]\wp(\mathbf{\Gamma}, t) - \mathbf{a}(\mathbf{\Gamma}, t) \cdot \nabla \wp(\mathbf{\Gamma}, t) + \frac{k_{\mathrm{B}}}{2} \Lambda : \nabla_p \nabla_p \wp(\mathbf{\Gamma}, t)$$

$$= \left\{\frac{-1}{2} \Lambda : [\nabla_p \nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}, t)] [\nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}, t)]\right.$$

$$-\frac{1}{2k_{\mathrm{B}}} \Lambda : [\nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}, t)][\nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}, t)] + \frac{1}{2} \Lambda : \nabla_p \nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}, t)$$

$$+ \frac{1}{2k_{\mathrm{B}}} \Lambda : [\nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}, t)] \quad [\nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}, t)]\right\}\wp(\mathbf{\Gamma}, t)$$

$$= 0.$$

(6.8
5)

This vanishes, which is not surprising because $\mathbf{a}(\mathbf{\Gamma}, t)$ is formally exact in this case. The reversible part is

$$\frac{\partial \wp(\mathbf{\Gamma}, t)}{\partial t} = -[\nabla \cdot \mathbf{b}(\mathbf{\Gamma}, t)]\wp(\mathbf{\Gamma}, t) - [\dot{\mathbf{\Gamma}}^0 + \mathbf{b}(\mathbf{\Gamma}, t)] \cdot \nabla \wp(\mathbf{\Gamma}, t)$$

$$= \left\{\frac{1}{2} \Lambda : \nabla_p \nabla_p S_{\mathrm{r,dyn}}(\mathbf{\Gamma}, t)\right.$$

$$-\left[\dot{\mathbf{\Gamma}}^0 - \frac{1}{2} \Lambda \nabla_p S_{\mathrm{r,dyn}}(\mathbf{\Gamma}, t)\right] \cdot \nabla_p S_{\mathrm{r}}(\mathbf{\Gamma}, t)\right\}\wp(\mathbf{\Gamma}, t).$$

(6.8
6)

I have no independent evidence for the veracity or otherwise of this.

6.3 Odd projection of the dynamic entropy

This section is concerned in the first place with using only the odd time parity projection of the dynamic part of the reservoir entropy, and in the second place with replacing the backwards most likely trajectories that appear in the integrand with adiabatic trajectories.

There are several reasons for rewriting the dynamic part of the reservoir entropy in these two ways. First, it identifies and focusses on the dominant contributions to the reservoir entropy. Second, it allows certain analytic results for the rate of change of the entropy to be obtained. Third, it gives an explicit connection between the novel dynamic part of the reservoir entropy and the well-known Green–Kubo expressions for the transport coefficients of a steady-state system, which provides additional confirmation of the validity of the former. And fourth it offers a different, possibly more efficient procedure for the

numerical computation of the reservoir entropy associated with a point in the sub-system phase space.

6.3.1 Odd projection

It is now argued that the odd projection of the dynamic part of the reservoir entropy is dominant. Two definitions of the odd projection are given. The first has the virtue of being physically intuitive and it can be applied immediately to steady-state thermodynamic systems. The second is more abstract and more general, and it can be applied both to steady-state thermodynamic systems and to mechanical non-equilibrium systems.

6.3.1.1 Simplest definition

As was mentioned in section 6.1, the crucial distinction between an equilibrium and a non-equilibrium system is that the probability distribution for the latter depends upon the sign of the molecular velocities, $\wp(\mathbf{\Gamma}, t) \neq \wp(\mathbf{\Gamma}^\dagger, t)$. For a sub-system phase space point $\mathbf{\Gamma} = \{\mathbf{q}^N, \mathbf{p}^N\}$, the conjugate phase space point is the one with the velocities reversed, $\mathbf{\Gamma}^\dagger = \{\mathbf{q}^N, (-\mathbf{p})^N\}$. Since the probability density cannot have even parity, neither can its exponent, the reservoir entropy, $S_\mathrm{r}(\mathbf{\Gamma}, t) \neq S_\mathrm{r}(\mathbf{\Gamma}^\dagger, t)$. Since the static part of the reservoir entropy is a purely equilibrium quantity, it necessarily has even parity, $S_\mathrm{r,st}(\mathbf{\Gamma}, t) = S_\mathrm{r,st}(\mathbf{\Gamma}^\dagger, t)$. This means that the dynamic part of the reservoir entropy cannot be even, $S_\mathrm{r,dyn}(\mathbf{\Gamma}, t) \neq S_\mathrm{r,dyn}(\mathbf{\Gamma}^\dagger, t)$.

Further, because the non-equilibrium aspects of the system are a perturbation on the equilibrium aspects, one can neglect the even projection of $S_\mathrm{r,dyn}(\mathbf{\Gamma}, t)$ in comparison with $S_\mathrm{r,st}(\mathbf{\Gamma}, t)$ and so write

$$S_\mathrm{r}(\mathbf{\Gamma}, t) \approx S_\mathrm{r,st}(\mathbf{\Gamma}, t) + S_\mathrm{r,dyn}^\mathrm{odd}(\mathbf{\Gamma}, t) + \bar{S}_\mathrm{r,dyn}(t).$$

(6.87)

First the odd projection is discussed, and then the time-dependent constant that is retained here.

The simplest definition of the odd projection of the dynamic part of the reservoir entropy is

$$
\begin{aligned}
S_\mathrm{r,dyn}^\mathrm{odd}(\mathbf{\Gamma}, t) &\equiv \frac{1}{2}\left[S_\mathrm{r,dyn}(\mathbf{\Gamma}, t) - S_\mathrm{r,dyn}(\mathbf{\Gamma}^\dagger, t)\right] \\
&= \frac{-1}{2} \int_0^t \mathrm{d}t\prime\, \left[\dot{S}_\mathrm{r,st}^0(\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma}, t), t\prime) - \dot{S}_\mathrm{r,st}^0(\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma}^\dagger, t), t\prime)\right].
\end{aligned}
$$

(6.88)

A more abstract form of the odd projection will be given shortly.

In addition to the odd projection, there also has been retained in the above the asymptotic contribution of the dynamic part of the reservoir entropy, which is constant in phase space and which therefore has even parity. On the most likely trajectory the asymptote of the integrand of the dynamic part of the reservoir entropy is

$$\dot{S}_\mathrm{r,st}^0(\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma}, t), t\prime) \to \overline{\dot{S}_\mathrm{r,st}^0}(t\prime), \quad |\,t\prime - t\,| \to \infty.$$

(6.89)

This asymptote arises from the fact that with overwhelming probability the system came from its most likely value in the past (and will return there in the future), independent of the current phase space point of the sub-system. As a consequence of this

independence, the asymptote holds unchanged for the conjugate point,

$$\dot{S}^0_{\mathrm{r,st}}(\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma}^\dagger, t), t\prime) \to \overline{\dot{S}^0_{\mathrm{r,st}}}(t\prime), \mid t\prime{-}t \mid\, \to \infty.$$

One can add and subtract this asymptote before taking the odd projection,

$$S_{\mathrm{r,dyn}}(\mathbf{\Gamma}, t) = -\int_0^t \mathrm{d}t\prime \left[\dot{S}^0_{\mathrm{r,st}}(\bar{\mathbf{\Gamma}}(t\prime|\,\mathbf{\Gamma}, t), t\prime) - \overline{\dot{S}^0_{\mathrm{r,st}}}(t\prime) \right] + \bar{S}_{\mathrm{r,dyn}}(t),$$

(6.90)

where

$$\bar{S}_{\mathrm{r,dyn}}(t) = -\int_0^t \mathrm{d}t\prime \, \overline{\dot{S}^0_{\mathrm{r,st}}}(t\prime).$$

(6.91)

(This result was given above as equation (6.17).) Since this is constant in phase space, it has even parity. Hence the asymptote does not contribute to the odd projection.

Although it could be neglected, it is arguably best to retain the even parity asymptotic contribution to the dynamic part of the entropy. This is a constant in phase space, so it has no effect on the relative probability of phase space points. It could simply be neglected, which effectively cancels it with the same contribution to the partition function. However, it does represent part of the total entropy produced as a function of time in the total system, which is a useful physical quantity. For this reason this even parity, constant term has been retained explicitly.

6.3.1.2 Abstract definition

The argument for keeping the odd time projection, and only the odd projection, of the dynamic part of the reservoir entropy was that the even time projection was dominated by the static part, and that the non-equilibrium probability could not have even time parity if it is to satisfy the Second Law of Thermodynamics. Now a second definition of conjugation is given that has more general application than the physically appealing momentum reversal definition used above.

For each phase space point $\mathbf{\Gamma}$ denote a conjugate point as $\mathbf{\Gamma}^\#$. This new conjugation operation is meant to include velocity reversal, but it can also include other operations besides. (I shall be more specific in the treatment of mechanical non-equilibrium systems below.) The conjugation operation is defined to have two properties. First it leaves the static part of the reservoir entropy unchanged,

$$S_{\mathrm{r,st}}(\mathbf{\Gamma}^\#, t) = S_{\mathrm{r,st}}(\mathbf{\Gamma}, t).$$

(6.92)

This is exactly as in the case of velocity reversal alone, $S_{\mathrm{r,st}}(\mathbf{\Gamma}^\dagger, t) = S_{\mathrm{r,st}}(\mathbf{\Gamma}, t)$.

Second, it reverses the sign of the adiabatic rate of static entropy production,

$$\dot{S}^0_{\mathrm{r,st}}(\mathbf{\Gamma}^\#, t) = -\dot{S}^0_{\mathrm{r,st}}(\mathbf{\Gamma}, t).$$

(6.93)

For the case of a steady-state system, which is not explicitly dependent on time, $\dot{S}^0_{\mathrm{r,st}}(\mathbf{\Gamma}) = \dot{\mathbf{\Gamma}}^0 \cdot \nabla S_{\mathrm{r,st}}(\mathbf{\Gamma})$, this condition is also the same as for velocity reversal, $\dot{S}^0_{\mathrm{r,st}}(\mathbf{\Gamma}^\dagger) = -\dot{S}^0_{\mathrm{r,st}}(\mathbf{\Gamma})$. However, for a system explicitly dependent on time, the extra term in the adiabatic time derivative, $\partial S_{\mathrm{r,st}}(\mathbf{\Gamma}, t)/\partial t$, has even parity, and so

$\dot{S}^0_{\mathrm{r,st}}(\boldsymbol{\Gamma}^\dagger,t) \neq -\dot{S}^0_{\mathrm{r,st}}(\boldsymbol{\Gamma},t)$. Hence this second condition defines the more general conjugation operation $\#$ such that the adiabatic rate of static entropy production must reverse sign also for explicitly time-dependent systems such as mechanical non-equilibrium systems.

Define the odd projection of the dynamic part of the reservoir entropy with respect to this conjugation operation as

$$
\begin{aligned}
S^{\mathrm{odd}}_{\mathrm{r,dyn}}(\boldsymbol{\Gamma},t) &\equiv \frac{1}{2}\left[S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma},t) - S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}^\#,t)\right] \\
&= \frac{-1}{2}\int_0^t \mathrm{d}t\prime \left[\dot{S}^0_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t\prime|\,\boldsymbol{\Gamma},t),t\prime) - \dot{S}^0_{\mathrm{r,st}}(\bar{\boldsymbol{\Gamma}}(t\prime|\,\boldsymbol{\Gamma}^\#,t),t\prime)\right].
\end{aligned}
$$

(6.94)

As for the usual conjugation operation of velocity reversal, since the asymptote in independent of the starting point, one can add and subtract the asymptote before taking the odd projection.

The even projection is defined analogously and one has for the reservoir entropy

$$
\begin{aligned}
S_{\mathrm{r}}(\boldsymbol{\Gamma},t) &= S_{\mathrm{r,st}}(\boldsymbol{\Gamma},t) + S^{\mathrm{odd}}_{\mathrm{r,dyn}}(\boldsymbol{\Gamma},t) + S^{\mathrm{even}}_{\mathrm{r,dyn}}(\boldsymbol{\Gamma},t) \\
&\approx S_{\mathrm{r,st}}(\boldsymbol{\Gamma},t) + S^{\mathrm{odd}}_{\mathrm{r,dyn}}(\boldsymbol{\Gamma},t) + \bar{S}_{\mathrm{r,dyn}}(t).
\end{aligned}
$$

(6.95)

where, as above, the asymptotic contribution is $\bar{S}_{\mathrm{r,dyn}}(t) = -\int_0^t \mathrm{d}t\prime\, \overline{\dot{S}^0_{\mathrm{r,st}}}(t\prime)$. The justification for the second equality is as in the first definition of conjugation, namely that the static part of the reservoir entropy has even parity and it dominates the even projection of the dynamic part of the reservoir entropy. And since the Second Law of Thermodynamics mandates that a non-equilibrium system must have time asymmetry, it is necessary to retain the odd projection of the dynamic part of the reservoir entropy since this is the only part of the reservoir entropy with such a time asymmetry. Since the Second Law of Thermodynamics refers to the change in entropy, it makes sense to take the definition of the conjugation operation to be that it reverses exactly the rate of entropy production, rather than the more obvious definition that it reverses the velocities, since there are some systems (i.e. mechanical non-equilibrium systems, which have a time-dependent external potential) where the two are not equivalent.

Retaining the even parity asymptotic contribution that is constant in phase space is not necessary for determining the relative probability of phase space points, but it does have a transparent physical interpretation and it contributes to the rate of dissipation of the total system, which make it useful to retain this constant term explicitly rather than incorporating it into the partition function.

6.3.2 Adiabatic odd projection, steady-state system

Consider in this subsection a steady-state, thermodynamic, non-equilibrium system. Examples include steady heat flow, shear flow, and diffusion. Since it is steady state, the static part of the reservoir entropy is independent of time, $S_{\mathrm{r,st}}(\boldsymbol{\Gamma},t) = S_{\mathrm{r,st}}(\boldsymbol{\Gamma})$, and velocity reversal alone is sufficient to negate the adiabatic rate of static entropy production, $\dot{S}^0_{\mathrm{r,st}}(\boldsymbol{\Gamma}^\dagger) = \dot{\boldsymbol{\Gamma}}^0(\boldsymbol{\Gamma}^\dagger) \cdot \nabla^\dagger S_{\mathrm{r,st}}(\boldsymbol{\Gamma}^\dagger) = -\dot{S}^0_{\mathrm{r,st}}(\boldsymbol{\Gamma})$. Hence velocity reversal serves as the conjugation operation for this steady-state system.

The behavior of the adiabatic rate of entropy production on various trajectories for a steady-state thermodynamic system is sketched in figure 6.4. On the most likely trajectory

the asymptotes are

$$\dot{S}^0_{\mathrm{r,st}}(\bar{\Gamma}(t\prime|\,\Gamma,t),t\prime) \to \overline{\dot{S}^0_{\mathrm{r,st}}}(t\prime), \quad |\,t\prime-t\,| \to \infty.$$

(6.96)

This asymptote arises from the fact that with overwhelming probability the system came from its most likely value in the past (and will return there in the future), independent of the current phase space point of the sub-system. For the present steady-state system, the asymptote is independent of $t\prime$ and t.

Figure 6.4. Sketch of the adiabatic rate of entropy production, $\dot{S}^0_{\mathrm{r,st}}$, for a steady-state thermodynamic non-equilibrium system, in the future (right-hand side) and in the past (left-hand side), starting from a likely phase space point Γ (upper), and from its unlikely conjugate $\Gamma^\dagger$ (lower). The solid curves are on most likely trajectories, $\bar{\Gamma}(t\prime|\,\Gamma,0)$ and $\bar{\Gamma}(t\prime|\,\Gamma^\dagger,0)$, and the short dashed curves are on adiabatic trajectories, $\Gamma^0(t\prime|\,\Gamma,0)$ and $\Gamma^0(t\prime|\,\Gamma^\dagger,0)$. After Attard (2009c).

In contrast, the asymptotic behavior on the adiabatic trajectory is

$$\dot{S}^0_{\mathrm{r,st}}(\Gamma^0(t\prime|\,\Gamma,t),t\prime) \sim \mathrm{sign}(t\prime-t)\,\overline{\dot{S}^0_{\mathrm{st}}}(t\prime).$$

(6.97)

This behavior can be seen in the case of steady heat flow in figure 3.2, where the tangent of the curve is proportional to the adiabatic rate of entropy production. The asymptotic behavior sets in for $|\,t\prime-t\,| \gtrsim \tau_{\mathrm{relax}}$, where τ_{relax} is a relaxation time that is long enough for the system to reach its asymptote, but not so long that the structure has changed significantly, $|\,\tau_{\mathrm{relax}}\dot{S}^0_{\mathrm{r,st}}\,| \ll |\,S_{\mathrm{r,st}}\,|$. (One does not need to impose this condition for the dissipative trajectory because the interactions with the reservoir maintain the structure of the sub-system.) For an isolated system, the structure represents a fluctuation, and $\dot{S}^0_{\mathrm{r,st}}$ represents its regression, which must be an odd function of time, at least for a

steady-state system. For $t\prime > 0$, the adiabatic asymptote and the actual asymptote approximately coincide, which is just Onsager's regression hypothesis (Onsager 1931).

In view of the trajectories shown in figure 6.4 and the above discussion, the odd projection of the dynamic part of the reservoir entropy may be transformed from an integral over the most likely trajectory to an integral over the adiabatic trajectories. Successive transformations yield

$$
\begin{aligned}
S_{\mathrm{r,dyn}}^{\mathrm{odd}}(\boldsymbol{\Gamma}, t) &= \frac{-1}{2} \int_0^t \mathrm{d}t\prime \left[\dot{S}_{\mathrm{r,st}}^0(\bar{\boldsymbol{\Gamma}}(t\prime|\, \boldsymbol{\Gamma}, t), t\prime) - \dot{S}_{\mathrm{r,st}}^0(\bar{\boldsymbol{\Gamma}}(t\prime|\, \boldsymbol{\Gamma}^\dagger, t), t\prime) \right] \\
&\approx \frac{-1}{2} \int_t^{2t} \mathrm{d}t\prime \left[\dot{S}_{\mathrm{r,st}}^0(\bar{\boldsymbol{\Gamma}}(t\prime|\, \boldsymbol{\Gamma}, t), t\prime) - \dot{S}_{\mathrm{r,st}}^0(\bar{\boldsymbol{\Gamma}}(t\prime|\, \boldsymbol{\Gamma}^\dagger, t), t\prime) \right] \\
&\approx \frac{-1}{2} \int_t^{2t} \mathrm{d}t\prime \left[\dot{S}_{\mathrm{r,st}}^0(\boldsymbol{\Gamma}^0(t\prime|\, \boldsymbol{\Gamma}, t), t\prime) - \dot{S}_{\mathrm{r,st}}^0(\boldsymbol{\Gamma}^0(t\prime|\, \boldsymbol{\Gamma}^\dagger, t), t\prime) \right] \\
&= \frac{-1}{2} \int_0^t \mathrm{d}t\prime \left[\dot{S}_{\mathrm{r,st}}^0(\boldsymbol{\Gamma}^0(t\prime|\, \boldsymbol{\Gamma}^\dagger, t)^\dagger, t\prime) - \dot{S}_{\mathrm{r,st}}^0(\boldsymbol{\Gamma}^0(t\prime|\, \boldsymbol{\Gamma}, t)^\dagger, t\prime) \right] \\
&= \frac{-1}{2} \int_0^t \mathrm{d}t\prime \left[\dot{S}_{\mathrm{r,st}}^0(\boldsymbol{\Gamma}^0(t\prime|\, \boldsymbol{\Gamma}, t), t\prime) - \dot{S}_{\mathrm{r,st}}^0(\boldsymbol{\Gamma}^0(t\prime|\, \boldsymbol{\Gamma}^\dagger, t), t\prime) \right] \\
&\equiv S_{\mathrm{r,dyn}}^{\mathrm{odd};0}(\boldsymbol{\Gamma}, t).
\end{aligned}
$$

(6.98)

The result says that for a steady-state thermodynamic system, the odd projection of the dynamic part of the reservoir entropy can be well-approximated by adiabatic trajectories. The first equality is the area between the solid curves in the left half of the figure. The second equality is the area between the solid curves in the right half of the figure. This follows because the dissipation on the most likely trajectory is to a good approximation even in time. The third equality is the area between the short dashed curves in the right half of the figure. This follows from Onsager's regression hypothesis. The fourth equality is the area between the short dashed curves in the left half of the figure. This follows from the microscopic reversibility of Hamilton's equations of motion,

$$
\boldsymbol{\Gamma}^0(t\prime|\, \boldsymbol{\Gamma}^\dagger, t) = \boldsymbol{\Gamma}^0(2t - t\prime|\, \boldsymbol{\Gamma}, t)^\dagger.
$$

(6.99)

This holds for a steady-state system in which the internal and any external potentials do not depend explicitly on time. The penultimate equality also holds for a steady-state system, since in this case the adiabatic rate of change of the static part of the reservoir entropy is $\dot{S}_{\mathrm{r,st}}^0(\boldsymbol{\Gamma}) = \dot{\boldsymbol{\Gamma}}^0 \cdot \nabla S_{\mathrm{r,st}}(\boldsymbol{\Gamma})$, and this has odd parity, $\dot{S}_{\mathrm{r,st}}^0(\boldsymbol{\Gamma}^\dagger) = -\dot{S}_{\mathrm{r,st}}^0(\boldsymbol{\Gamma})$.

It is obvious from the figure that the integrand asymptotes to zero. This means that the lower limit of the integral can be replaced by $t - \tau$ for some convenient interval $\tau > 0$. Although the integrand is an exact differential, there is no point in analytically evaluating the integral because the actual value at the lower limit would be required, $S_{\mathrm{r,dyn}}^{\mathrm{odd}}(\boldsymbol{\Gamma}, t) = [S_{\mathrm{r,st}}(\boldsymbol{\Gamma}^0(t - \tau \,|\, \boldsymbol{\Gamma}, t), t - \tau) - S_{\mathrm{r,st}}(\boldsymbol{\Gamma}^0(t - \tau \,|\, \boldsymbol{\Gamma}^\dagger, t), t - \tau)]/2$. (Although $\dot{S}_{\mathrm{r,st}}^0$ has the same asymptote starting at $\boldsymbol{\Gamma}$ and at $\boldsymbol{\Gamma}^\dagger$, there is a finite difference between the respective asymptotes of $S_{\mathrm{r,st}}$ that corresponds to the area between the two curves in the left half of the figure.) It takes no more computational effort to perform the quadrature numerically than it does to calculate the adiabatic trajectories backward to their lower limit.

In summary, this section argues that in some circumstances the odd projection of the dynamic part of the reservoir entropy is either dominant or is all that is required. Further, it says that for a steady-state thermodynamic system, the odd projection of the dynamic part of the reservoir entropy may be evaluated on the past adiabatic trajectories. With this result, one does not need to evaluate the most likely backwards trajectory. (On the other hand, calculating the dissipative trajectory, or even the stochastic dissipative trajectory may be advantageous because it is thermostatted and therefore stable, in contrast to the adiabatic trajectory.) This adiabatic expression for the dynamic part of the reservoir entropy has been tested with computer simulations of non-equilibrium systems and found to be accurate (Attard 2006b). This adiabatic approximation for the trajectories of the dynamic part of the entropy, and the general formulation of the non-equilibrium probability, will be tested analytically by direct comparison with the Green–Kubo relations in section 6.3.6.

6.3.3 Adiabatic odd projection, mechanical system

In this subsection is considered a mechanical non-equilibrium system that has a time-dependent external potential. A specific example will be considered from which the general procedure can be extrapolated.

Consider a solvent in which there is a solute acted upon by a time-varying potential, and also a thermal reservoir with which energy can be exchanged. Suppose that the sub-system Hamiltonian is of the form $\mathcal{H}(\mathbf{\Gamma}, t) = \mathcal{H}_0(\mathbf{\Gamma}) + U(\mathbf{x}, t)$. The time-independent Hamiltonian $\mathcal{H}_0(\mathbf{\Gamma})$ describes the interactions of the solvent particles amongst themselves and with the solute particle. The solute particle is acted upon by a time-dependent external potential that takes the form of a moving parabolic trap,

$$
U(\mathbf{x}, t) = \frac{\kappa}{2}[\mathbf{x} - \mathbf{b}(t)] \cdot [\mathbf{x} - \mathbf{b}(t)].
$$

(6.100)

Here $\mathbf{x}$ is the position of the solute and $\mathbf{b}(t)$ is location of the trap. The sub-system phase space point $\mathbf{\Gamma}$ includes the solute as well as the solvent particles, $\mathbf{\Gamma} = \{\mathbf{\Gamma}_q, \mathbf{x}; \mathbf{\Gamma}_p, M\mathbf{x}\}$, where M is the solute mass. Velocity conjugation applies to the solvent and to the solute,

$$
\mathbf{\Gamma}^\dagger = \{\mathbf{\Gamma}_q, \mathbf{x}; -\mathbf{\Gamma}_p, -M\mathbf{x}\}.
$$

(6.101)

The sub-system can exchange energy with a thermal reservoir of temperature T, and therefore the static part of the reservoir entropy is

$$
S_{\mathrm{r,st}}(\mathbf{\Gamma}, t) = \frac{-1}{T}\mathcal{H}(\mathbf{\Gamma}, t).
$$

(6.102)

Evidently this has even parity with respect to velocity reversal, $S_{\mathrm{r,st}}(\mathbf{\Gamma}^\dagger, t) = S_{\mathrm{r,st}}(\mathbf{\Gamma}, t)$. The adiabatic rate of change of the static entropy is

$$
\dot{S}^0_{\mathrm{r,st}}(\mathbf{\Gamma}, t) = \frac{-1}{T}\left[\frac{\partial \mathcal{H}(\mathbf{\Gamma}, t)}{\partial t} + \dot{\mathbf{\Gamma}}^0 \cdot \nabla \mathcal{H}(\mathbf{\Gamma}, t)\right] = \frac{-1}{T}\frac{\partial U(\mathbf{x}, t)}{\partial t}.
$$

(6.103)

This has even parity with respect to velocity reversal, $\dot{S}^0_{\mathrm{r,st}}(\mathbf{\Gamma}^\dagger, t) = \dot{S}^0_{\mathrm{r,st}}(\mathbf{\Gamma}, t)$, since the right-hand side is independent of the solvent and solute velocity.

In addition to velocity reversal symmetry, the Hamiltonian is symmetric with respect to reflection of the solute and all of the solvent particles about the trap minimum,

$$\mathbf{\Gamma}^{\ddagger} = \{2\mathbf{b}(t) - \mathbf{\Gamma}_q, 2\mathbf{b}(t) - \mathbf{x}; \mathbf{\Gamma}_p, M\mathbf{x}\}.$$

(6.104)

Hence the static part of the reservoir entropy is unchanged by this reflection operation,

$$S_{\mathrm{r,st}}(\mathbf{\Gamma}^{\ddagger}, t) = S_{\mathrm{r,st}}(\mathbf{\Gamma}, t).$$

(6.105)

For this symmetry to hold, all the solvent particles have to be reflected at the same time as the solute particle. This assumes that the solvent and the solute are not chiral, and also that the sub-system is homogeneous apart from the moving trap potential.

This reflection symmetry as well as the velocity reversal symmetry is sketched in figure 6.5. The four configurations shown have the same energy and hence static part of the reservoir entropy,

$$S_{\mathrm{r,st}}(\mathbf{\Gamma}, t) = S_{\mathrm{r,st}}(\mathbf{\Gamma}^{\dagger}, t) = S_{\mathrm{r,st}}(\mathbf{\Gamma}^{\ddagger}, t) = S_{\mathrm{r,st}}(\mathbf{\Gamma}^{\dagger\ddagger}, t).$$

(6.106)

(Since this depends on the total energy of the solute and solvent, $\mathcal{H}(\mathbf{\Gamma}, t)$, one can see why the solvent particles also have to be reflected in the trap minimum, since this is the only way that their interaction potential with the solute is unchanged.)

Figure 6.5. Four configurations of a Brownian particle (circle) in a solvent (not shown) in a moving potential trap (curve). The arrows indicate the velocity of the particle and of the trap. The phase space point $\mathbf{\Gamma}$ is a likely point. The dagger signifies velocity reversal, and the double dagger signifies reflection in the trap minimum. These operations apply to the solute and the solvent particles. All four configurations have the same Hamiltonian energy.

As mentioned, the adiabatic rate of change of the static part of the reservoir entropy is

$$\dot{S}^{0}_{\mathrm{r,st}}(\mathbf{\Gamma}, t) = \frac{-1}{T}\frac{\partial U(\mathbf{x}, t)}{\partial t} = \frac{\kappa}{T}\dot{\mathbf{b}}(t) \cdot [\mathbf{x} - \mathbf{b}(t)].$$

(6.107)

By inspection one sees that that this is independent of solvent and solute velocity, $\dot{S}^{0}_{\mathrm{r,st}}(\mathbf{\Gamma}, t) = \dot{S}^{0}_{\mathrm{r,st}}(\mathbf{\Gamma}^{\dagger}, t)$, but that it changes sign upon position reflection,

$$\dot{S}_{\mathrm{r,st}}^{0}(\boldsymbol{\Gamma}, t) = -\dot{S}_{\mathrm{r,st}}^{0}(\boldsymbol{\Gamma}^{\ddagger}, t), \text{and also } \dot{S}_{\mathrm{r,st}}^{0}(\boldsymbol{\Gamma}, t) = -\dot{S}_{\mathrm{r,st}}^{0}(\boldsymbol{\Gamma}^{\dagger\ddagger}, t).$$

In view of the definition of the generalized conjugation operation, equation (6.93), one can choose either $\# \equiv \ddagger$, or else $\# \equiv \dagger\ddagger$.

In previous work it has been argued that the combined symmetry operation $\dagger\ddagger$ should be used to define the odd projection (Attard 2009c), and this has been successfully used in computer simulations of a driven Brownian particle (Attard 2009b). A possible reason for needing to combine both is that non-equilibrium systems are necessarily sensitive to both velocity reversal and the rate of entropy production. The combined symmetry operation is discussed for an explicit calculation of driven Brownian motion in equation (6.172). Following that earlier work, the odd projection of the dynamic part of the reservoir entropy is here also defined as

$$
\begin{aligned}
S_{\mathrm{r,dyn}}^{\mathrm{odd}}(\boldsymbol{\Gamma}, t) &= \frac{-1}{2} \int_{0}^{t} \mathrm{d}t\prime \left[\dot{S}_{\mathrm{r,st}}^{0}(\bar{\boldsymbol{\Gamma}}(t\prime| \boldsymbol{\Gamma}, t), t\prime) - \dot{S}_{\mathrm{r,st}}^{0}(\bar{\boldsymbol{\Gamma}}(t\prime| \boldsymbol{\Gamma}^{\dagger\ddagger}, t), t\prime) \right] \\
&\approx \frac{-1}{2} \int_{0}^{t} \mathrm{d}t\prime \left[\dot{S}_{\mathrm{r,st}}^{0}(\boldsymbol{\Gamma}^{0}(t\prime| \boldsymbol{\Gamma}, t), t\prime) - \dot{S}_{\mathrm{r,st}}^{0}(\boldsymbol{\Gamma}^{0}(t\prime| \boldsymbol{\Gamma}^{\dagger\ddagger}, t), t\prime) \right] \\
&\equiv S_{\mathrm{r,dyn}}^{\mathrm{odd};0}(\boldsymbol{\Gamma}, t).
\end{aligned}
$$

In the second equality the most likely trajectories have been replaced by adiabatic trajectories. This result is formally identical to equation (6.98) for a steady-state thermodynamic system (but with the present definition of conjugation).

Figure 6.6 sketches the backwards evolution from a likely point $\boldsymbol{\Gamma}_1$ and from its unlikely conjugate $\boldsymbol{\Gamma}_1^{\dagger\ddagger}$. As in the preceding derivation, it is assumed that the most likely trajectory is approximately the same as the adiabatic trajectory in each case. This is a reasonable assumption in view of the fact that the solute is surrounded by solvent that is included in the adiabatic evolution. In this case over not too long-time periods, the reservoir is redundant, since the solvent itself acts like a heat sink. Hence the dissipative trajectories of the system can be replaced by adiabatic trajectories of the sub-system, provided that the time period concerned is short enough that any perturbation by the reservoir is insignificant.

Figure 6.6. A Brownian particle in a solvent in a moving potential trap. Time increases from left to right, with the current time t_1 being on the far right. The first line shows a backward trajectory from the current point $\boldsymbol{\Gamma}_1$, $\bar{\boldsymbol{\Gamma}}(t\prime| \boldsymbol{\Gamma}_1, t_1) \approx \boldsymbol{\Gamma}^{0}(t\prime| \boldsymbol{\Gamma}_1, t_1)$,
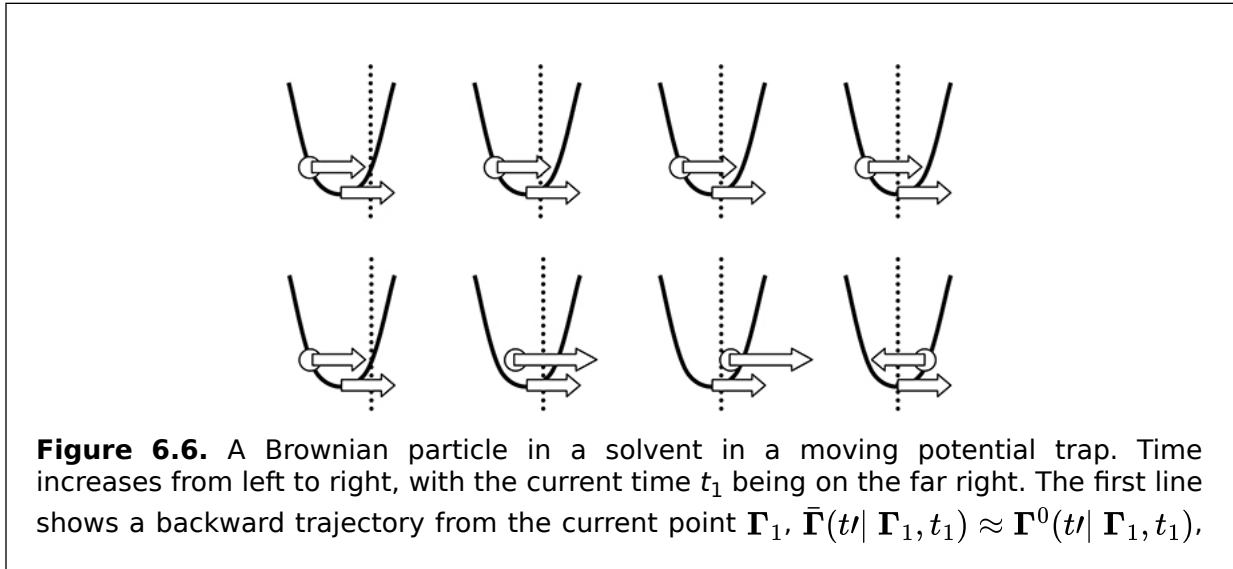
and the second line shows a trajectory from the unlikely conjugate point, $\bar{\mathbf{\Gamma}}(t\prime|\ \mathbf{\Gamma}_1^{\dagger\ddagger}, t_1) \approx \mathbf{\Gamma}^0(t\prime|\ \mathbf{\Gamma}_1^{\dagger\ddagger}, t_1)$. The dotted line is fixed in space as a guide to the eye.

The two different starting configurations at t_1 in figure 6.6 both most likely end up at the same destination at $t_2 < t_1$. This means that the integrand for $S_{\mathrm{r,dyn}}^{\mathrm{odd};0}(\mathbf{\Gamma}, t)$ goes to zero, and that the lower limit on the integral can be replaced by $t - \tau$ for some convenient time interval.

It is quite common in the treatment of a driven Brownian particle to make the particle the entire sub-system and to include all of the solvent as the reservoir; such a model is explored in section 6.4. In this approach solvent effects are subsumed entirely into the diffusion constant (or the related memory function for rapidly varying applied potentials). The adiabatic approximation would be quite inappropriate in such a picture, since one can only justify replacing the most likely dissipative trajectories by adiabatic trajectories if the latter include the solvent to provide a mechanism and sink for dissipation.

6.3.4 Adiabatic time derivative of the entropy

In this section the adiabatic time derivative of the odd projection of the dynamic part of the reservoir entropy is sought. It is actually much simpler to evaluate the adiabatic derivative first, and then take the odd projection. However, the operations of trajectory derivative and odd projection do not commute. In using the result now derived, one implicitly assumes that the non-commuting part is negligible compared to the commuting part.

The odd projection of the adiabatic derivative of the dynamic part of the reservoir entropy may be obtained by approximating the dissipative trajectories by adiabatic ones. This is more valid for a mechanical non-equilibrium system than for a steady-state thermodynamic system. One has

$$
\begin{aligned}
\frac{\mathrm{d}^0 S_{\mathrm{r,dyn}}(\mathbf{\Gamma}, t)}{\mathrm{d}t} \ &\approx\ \frac{\mathrm{d}^0 S_{\mathrm{r,dyn}}^0(\mathbf{\Gamma}, t)}{\mathrm{d}t} \\
&= \frac{S_{\mathrm{r,dyn}}^0(\mathbf{\Gamma} + \Delta_t \dot{\mathbf{\Gamma}}^0, t + \Delta_t) - S_{\mathrm{r,dyn}}^0(\mathbf{\Gamma}, t)}{\Delta_t} \\
&= \frac{-1}{\Delta_t}\left\{ \Delta_t \dot{S}_{\mathrm{r,st}}^0(\mathbf{\Gamma}^0(t\mid \mathbf{\Gamma}, t), t) - \int_0^t \mathrm{d}t\prime \Big[\dot{S}_{\mathrm{r,st}}^0(\mathbf{\Gamma}^0(t\prime|\ \mathbf{\Gamma}, t), t\prime) \right. \\
&\quad \left. - \dot{S}_{\mathrm{r,st}}^0(\mathbf{\Gamma}^0(t\prime|\ \{\mathbf{\Gamma} + \Delta_t \dot{\mathbf{\Gamma}}^0\}, t + \Delta_t), t\prime) \Big] \right\} \\
&= -\dot{S}_{\mathrm{r,st}}^0(\mathbf{\Gamma}, t).
\end{aligned}
$$

(6.1
10)

The surviving term is the integrand evaluated at the upper limit of the integral (Leibnitz rule). The remaining terms cancel because a point on an adiabatic trajectory at time $t\prime$ is unchanged by specifying any point and time that lies on it, $\mathbf{\Gamma}^0(t\prime|\ \mathbf{\Gamma} + \Delta_t \dot{\mathbf{\Gamma}}^0, t + \Delta_t) = \mathbf{\Gamma}^0(t\prime|\ \mathbf{\Gamma}, t)$. With this the odd projection of the adiabatic derivative of the dynamic part of the reservoir entropy is

$$\dot{S}_{\mathrm{r,dyn}}^{0;\mathrm{odd}}(\boldsymbol{\Gamma}, t) \equiv \frac{1}{2} \left\{ \frac{\mathrm{d}^0 S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t)}{\mathrm{d}t} - \frac{\mathrm{d}^0 S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}^{\#}, t)}{\mathrm{d}t} \right\}$$

$$\approx -\dot{S}_{\mathrm{r,st}}^{0}(\boldsymbol{\Gamma}, t).$$

(6.1
11)

since by definition, $\dot{S}_{\mathrm{r,st}}^{0}(\boldsymbol{\Gamma}^{\#}, t) = -\dot{S}_{\mathrm{r,st}}^{0}(\boldsymbol{\Gamma}, t)$. The accuracy or utility of this result is unclear.

This result should not be confused with the adiabatic derivative of the odd projection of the dynamic part of the reservoir entropy. The latter is not so easy to obtain explicitly because the conjugate starting points of the adiabatic trajectories do not lie on the same trajectory,

$$\boldsymbol{\Gamma}^0(t\prime|\ \{\boldsymbol{\Gamma} + \Delta_t \dot{\boldsymbol{\Gamma}}^0\}^{\dagger}, t + \Delta_t) = \boldsymbol{\Gamma}^0(t\prime|\ \boldsymbol{\Gamma}^{\dagger} + \Delta_t \{\dot{\boldsymbol{\Gamma}}^0\}^{\dagger} - \Delta_t \dot{\boldsymbol{\Gamma}}^0(\boldsymbol{\Gamma}^{\dagger}), t)$$

$$= \boldsymbol{\Gamma}^0(t\prime|\ \boldsymbol{\Gamma}^{\dagger} + \Delta_t \{\dot{\boldsymbol{\Gamma}}^0\}^{\dagger} + \Delta_t \{\dot{\boldsymbol{\Gamma}}^0\}^{\dagger}), t)$$

$$= \boldsymbol{\Gamma}^0(t\prime|\ \boldsymbol{\Gamma}^{\dagger} + 2\Delta_t \{\dot{\boldsymbol{\Gamma}}^0\}^{\dagger}, t).$$

(6.1
12)

The even projection of the dynamic part of the reservoir entropy can be approximated by the asymptote of the integrand. Hence the adiabatic derivative of the even projection of the dynamic part of the reservoir entropy is approximately

$$\frac{\mathrm{d}^0 S_{\mathrm{r,dyn}}^{\mathrm{even}}(\boldsymbol{\Gamma}, t)}{\mathrm{d}t} \approx \frac{\mathrm{d}^0 \bar{S}_{\mathrm{r,dyn}}(t)}{\mathrm{d}t}$$

$$= \frac{\partial \bar{S}_{\mathrm{r,dyn}}(t)}{\partial t}$$

$$= -\overline{\dot{S}_{\mathrm{r,st}}^{0}}(t).$$

(6.1
13)

With this the adiabatic total time derivative of the dynamic part of the reservoir entropy may be taken to be

$$\dot{S}_{\mathrm{r,dyn}}^{0}(\boldsymbol{\Gamma}, t) \equiv \frac{\mathrm{d}^0 S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t)}{\mathrm{d}t} \approx -\dot{S}_{\mathrm{r,st}}^{0}(\boldsymbol{\Gamma}, t) - \overline{\dot{S}_{\mathrm{r,st}}^{0}}(t).$$

(6.1
14)

As mentioned above, the form of the first term on the right-hand side assumes that the commuting part of the adiabatic derivative of the odd projection dominates the non-commuting part.

It follows that the adiabatic total time derivative of the reservoir entropy is

$$\dot{S}_{\mathrm{r}}^{0}(\boldsymbol{\Gamma}, t) \equiv \frac{\mathrm{d}^0 S_{\mathrm{r}}(\boldsymbol{\Gamma}, t)}{\mathrm{d}t}$$

$$\approx -\overline{\dot{S}_{\mathrm{r,st}}^{0}}(t).$$

This is evidently a constant in phase space. This holds for both steady-state thermodynamic systems and for mechanical non-equilibrium system (subject to the validity of the various approximations invoked above).

This is consistent with equation (6.30), which assumed that the exact result for the average adiabatic time derivative of the reservoir entropy held locally. That result is

$$\dot{S}_{\rm r}^0(\mathbf{\Gamma}, t) = \frac{k_{\rm B}\dot{Z}(t)}{Z(t)} = \dot{S}_{\rm tot}(t).$$

One can therefore conclude that the rate of change of the total entropy is equal and opposite to the most likely adiabatic rate of change of the static part of the reservoir entropy,

$$\dot{S}_{\rm tot}(t) = -\overline{\dot{S}_{\rm r,st}^0}(t).$$

6.3.5 Steady-state system

This general result applied to a general steady-state system is

$$\dot{S}_{\rm tot}(x_{\rm r}, \nabla x_{\rm r}, t) = \dot{S}_{\rm r}(x_{\rm r}, \nabla x_{\rm r}, t) = -\overline{\dot{S}_{\rm r,st}^0}.$$

Here $x_{\rm r}$ is the reservoir field variable or variables. Because this is a steady-state system, the structure and hence the sub-system entropy does not change with time, $\dot{S}_{\rm s}(t) = 0$, and so only the reservoir entropy contributes to the rate of change of the total entropy. For a steady-state system, $\overline{\dot{S}_{\rm r,st}^0}$ is independent of time. The physical interpretation of this result is that in a steady state, the internal change in the structure of the sub-system is canceled by exchange with the reservoir, which means that the internal change in entropy is equal and opposite to the change in reservoir entropy.

Since the material flux is constant, so is the rate of change of reservoir entropy, and so integrating the latter gives the reservoir entropy as a linear function of time, $S_{\rm r}(x_{\rm r}, \nabla x_{\rm r}, t) = S_{\rm r}(x_{\rm r}, \nabla x_{\rm r}, 0) - t\overline{\dot{S}_{\rm r,st}^0}$. The partition function is therefore of the form

$$Z(x_{\rm r}, \nabla x_{\rm r}, t) = Z(x_{\rm r}, \nabla x_{\rm r})\mathrm{e}^{-t\overline{\dot{S}_{\rm r,st}^0}/k_{\rm B}},$$

which gives the total unconstrained entropy as

$$S_{\rm tot}(x_{\rm r}, \nabla x_{\rm r}, t) = S_{\rm tot}(x_{\rm r}, \nabla x_{\rm r}) - t\overline{\dot{S}_{\rm r,st}^0}.$$

A steady-state system is homogeneous in time,

$$\bar{\mathbf{\Gamma}}(t \mid \mathbf{\Gamma}_0, t_0) = \bar{\mathbf{\Gamma}}(t - t_0 \mid \mathbf{\Gamma}_0).$$

Hence the dynamic part of the reservoir entropy may be written

$$S_{\text{r,dyn}}(\boldsymbol{\Gamma} \mid \nabla x_{\text{r}}, t) = -\int_0^t dt\prime \; \dot{S}_{\text{r,st}}^0(\bar{\boldsymbol{\Gamma}}(t\prime \mid \boldsymbol{\Gamma}, t))$$

$$= -\int_{-t}^0 dt\prime\prime \; \dot{S}_{\text{r,st}}^0(\bar{\boldsymbol{\Gamma}}(t\prime\prime \mid \boldsymbol{\Gamma})), \quad t\prime\prime = t\prime - t.$$

The partial time derivative of this is

$$\frac{\partial S_{\text{r,dyn}}(\boldsymbol{\Gamma} \mid \nabla x_{\text{r}}, t)}{\partial t} = -\dot{S}_{\text{r,st}}^0(\bar{\boldsymbol{\Gamma}}(-t \mid \boldsymbol{\Gamma}, 0))$$

$$= -\overline{\dot{S}_{\text{r,st}}^0}.$$

This is evidently a constant in phase space. The second equality holds because the dynamic part of the entropy is formulated such that t is assumed large enough for the system to have settled into a steady state and the integrand to have reached its asymptotic limit.

For a steady-state system the macroscopic structure of the sub-system is fixed and the static part of the reservoir entropy is not explicitly dependent on time, $S_{\text{r,st}}(\boldsymbol{\Gamma})$. Hence in this case the partial time derivative of the reservoir entropy is just that of the dynamic part alone,

$$\frac{\partial S_{\text{r}}(\boldsymbol{\Gamma} \mid x_{\text{r}}, \nabla x_{\text{r}}, t)}{\partial t} = \frac{\partial S_{\text{r,dyn}}(\boldsymbol{\Gamma} \mid \nabla x_{\text{r}}, t)}{\partial t} = -\overline{\dot{S}_{\text{r,st}}^0}.$$

In view of equation (6.115), $\dot{S}_{\text{r}}^0(\boldsymbol{\Gamma}, t) = -\overline{\dot{S}_{\text{r,st}}^0}(t)$, this implies that

$$\dot{\boldsymbol{\Gamma}}^0 \cdot \nabla S_{\text{r}}(\boldsymbol{\Gamma}, t) = 0.$$

This agrees with equation (6.31).

With this result for the partial time derivative of the dynamic entropy for a steady-state system, one can write

$$S_{\text{r,dyn}}(\boldsymbol{\Gamma} \mid x_{\text{r}}, \nabla x_{\text{r}}, t) = S_{\text{r,dyn}}(\boldsymbol{\Gamma} \mid \nabla x_{\text{r}}) - t\overline{\dot{S}_{\text{r,st}}^0}$$

with

$$S_{\text{r,dyn}}(\boldsymbol{\Gamma} \mid \nabla x_{\text{r}}) = -\int_{-\infty}^0 dt\prime\prime \left\{ \dot{S}_{\text{r,st}}^0(\bar{\boldsymbol{\Gamma}}(t\prime\prime \mid \boldsymbol{\Gamma}, 0)) - \overline{\dot{S}_{\text{r,st}}^0} \right\}.$$

The probability density is

$$\wp(\mathbf{\Gamma} \mid x_{\mathrm{r}}, \nabla x_{\mathrm{r}}) = \frac{\mathrm{e}^{S_{\mathrm{r,st}}(\mathbf{\Gamma}|x_{\mathrm{r}},\nabla x_{\mathrm{r}})/k_{\mathrm{B}}}\mathrm{e}^{S_{\mathrm{r,dyn}}(\mathbf{\Gamma}|\nabla x_{\mathrm{r}},t)/k_{\mathrm{B}}}}{h^{3N}N!Z(x_{\mathrm{r}}, \nabla x_{\mathrm{r}}, t)}$$

$$= \frac{\mathrm{e}^{S_{\mathrm{r,st}}(\mathbf{\Gamma}|x_{\mathrm{r}},\nabla x_{\mathrm{r}})/k_{\mathrm{B}}}\mathrm{e}^{S_{\mathrm{r,dyn}}(\mathbf{\Gamma}|\nabla x_{\mathrm{r}})/k_{\mathrm{B}}}}{h^{3N}N!Z(x_{\mathrm{r}}, \nabla x_{\mathrm{r}})}.$$

(6.1
28)

The passage from the first to the second equality confirms the first three equations in this subsection, equation (6.118) *et seq.*

The constant part of the partition function is just

$$Z(x_{\mathrm{r}}, \nabla x_{\mathrm{r}}) = \frac{1}{h^{3N}N!}\int \mathrm{d}\mathbf{\Gamma} \ \mathrm{e}^{S_{\mathrm{r,st}}(\mathbf{\Gamma}|x_{\mathrm{r}},\nabla x_{\mathrm{r}})/k_{\mathrm{B}}}\mathrm{e}^{S_{\mathrm{r,dyn}}(\mathbf{\Gamma}|\nabla x_{\mathrm{r}})/k_{\mathrm{B}}}$$

$$= \mathrm{e}^{S_{\mathrm{s}}(\bar{X}_{\mathrm{s}},\overline{\dot{X}_{\mathrm{s}}^{0}}|x_{\mathrm{r}},\nabla x_{\mathrm{r}})/k_{\mathrm{B}}}\mathrm{e}^{S_{\mathrm{r,st}}(\bar{X}_{\mathrm{s}}|x_{\mathrm{r}},\nabla x_{\mathrm{r}})/k_{\mathrm{B}}},$$

(6.1
29)

The second equality is exact when fluctuations are negligible, which they are in the thermodynamic limit. In terms of the total unconstrained entropy this is

$$S_{\mathrm{tot}}(x_{\mathrm{r}}, \nabla x_{\mathrm{r}}) = S_{\mathrm{s}}(\bar{X}_{\mathrm{s}}, \overline{\dot{X}_{\mathrm{s}}^{0}} \mid x_{\mathrm{r}}, \nabla x_{\mathrm{r}}) + S_{\mathrm{r,st}}(\bar{X}_{\mathrm{s}} \mid x_{\mathrm{r}}, \nabla x_{\mathrm{r}}).$$

(6.1
30)

The sub-system entropy in the most likely macrostate, $\bar{S}_{\mathrm{s}}$, contains the effects of the dynamic order due to the induced energy flux, which is embodied in the dynamic part of the entropy.

6.3.5.1 Steady heat flow

For steady heat flow, the static part of the reservoir entropy is

$$S_{\mathrm{r,st}}(\mathbf{\Gamma}) = \frac{-E_{0}(\mathbf{\Gamma})}{T_{0}} - \frac{E_{1}(\mathbf{\Gamma})}{T_{1}}.$$

(6.1
31)

The most likely heat flux is

$$\bar{J}_{\mathrm{e}}^{0} = \frac{1}{V}\overline{\dot{E}_{1}^{0}}$$

$$= \frac{\hat{\tau}}{2V}\Lambda\frac{1}{T_{1}}.$$

(6.1
32)

Since the present case is that of an applied temperature gradient rather than a spontaneous fluctuation, one has $\hat{\tau} = +1$ here.

The adiabatic rate of change of the static part of the reservoir entropy is

$$\overline{\dot{S}_{\mathrm{r,st}}^{0}} = \frac{-1}{T_{1}}\overline{\dot{E}_{1}^{0}}.$$

(6.1
33)

Since the flux is constant, the reservoir entropy increases linearly with time, and it may be written as $S_{\mathrm{r}}(t \mid T_{0}, T_{1}) = S_{\mathrm{r}}(0 \mid T_{0}, T_{1}) + t\overline{\dot{E}_{1}^{0}}/T_{1}$, with the (immaterial) value at $t = 0$ taken to be

$$S_{\mathrm{r}}(0 \mid T_0, T_1) = \bar{S}_{\mathrm{r,st}}(T_0, T_1) = \frac{-\bar{E}_0(\boldsymbol{\Gamma})}{T_0} - \frac{\bar{E}_1(\boldsymbol{\Gamma})}{T_1}.$$

(6.134)

The partition function is therefore of the form

$$Z(T_0, T_1, t) = Z(T_0, T_1)\mathrm{e}^{-t\overline{\dot{S}^0_{\mathrm{r,st}}}/k_{\mathrm{B}}},$$

(6.135)

which gives the total unconstrained entropy as

$$S_{\mathrm{tot}}(T_0, T_1, t) = S_{\mathrm{tot}}(T_0, T_1) - t\overline{\dot{S}^0_{\mathrm{r,st}}}.$$

(6.136)

The partial time derivative of the dynamic part of the reservoir entropy is

$$\frac{\partial S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}, t \mid T_1)}{\partial t} = -\overline{\dot{S}^0_{\mathrm{r,st}}} = \frac{1}{T_1}\overline{\dot{E}^0_1}.$$

(6.137)

The probability density is

$$\wp(\boldsymbol{\Gamma} \mid T_0, T_1) = \frac{\mathrm{e}^{S_{\mathrm{r,st}}(\boldsymbol{\Gamma}\mid T_0,T_1)/k_{\mathrm{B}}}\mathrm{e}^{S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}\mid T_1,t)/k_{\mathrm{B}}}}{h^{3N}N!Z(T_0, T_1, t)}$$
$$= \frac{\mathrm{e}^{S_{\mathrm{r,st}}(\boldsymbol{\Gamma}\mid T_0,T_1)/k_{\mathrm{B}}}\mathrm{e}^{S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}\mid T_1)/k_{\mathrm{B}}}}{h^{3N}N!Z(T_0, T_1)}.$$

(6.138)

The time-independent part of the partition function is just

$$Z(T_0, T_1) = \frac{1}{h^{3N}N!}\int \mathrm{d}\boldsymbol{\Gamma}\; \mathrm{e}^{S_{\mathrm{r,st}}(\boldsymbol{\Gamma}\mid T_0,T_1)/k_{\mathrm{B}}}\mathrm{e}^{S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}\mid T_1)/k_{\mathrm{B}}}$$
$$= \mathrm{e}^{S_{\mathrm{s}}(\bar{E}_0,\bar{E}_1,\overline{\dot{E}^0_1}\mid T_0,T_1)/k_{\mathrm{B}}}\mathrm{e}^{S_{\mathrm{r,st}}(\bar{E}_0,\bar{E}_1\mid T_0,T_1)/k_{\mathrm{B}}}.$$

(6.139)

The second equality is exact when fluctuations are negligible, which they are in the thermodynamic limit. In terms of the total unconstrained entropy this is

$$S_{\mathrm{tot}}(T_0, T_1) = S_{\mathrm{s}}(\bar{E}_0, \bar{E}_1, \overline{\dot{E}^0_1} \mid T_0, T_1) + S_{\mathrm{r,st}}(\bar{E}_0, \bar{E}_1 \mid T_0, T_1)$$
$$= S_{\mathrm{s}}(\bar{E}_0, \bar{E}_1, \overline{\dot{E}^0_1} \mid T_0, T_1) - \frac{\bar{E}_0}{T_0} - \frac{\bar{E}_1}{T_1}.$$

(6.140)

The sub-system entropy in the most likely macrostate, $S_{\mathrm{s}}(\bar{E}_0, \bar{E}_1, \overline{\dot{E}^0_1} \mid T_0, T_1)$, contains the effects of the dynamic order due to the induced energy flux, which is embodied in the dynamic part of the entropy.

6.3.6 Green–Kubo relations

The validity and utility of the expression for the non-equilibrium probability will now be illustrated with a simple derivation of the Green–Kubo relations (Onsager 1931, Green 1954, Kubo 1966). These relate the hydrodynamic transport coefficients to the equilibrium time correlation functions of the fluxes.

For the particular case of heat flow, the static part of the reservoir entropy is (see chapter 3)

$$S_{\mathrm{r,st}}(\mathbf{\Gamma}) = \frac{-E_0(\mathbf{\Gamma})}{T_0} - \frac{E_1(\mathbf{\Gamma})}{T_1},$$

(6.141)

where the nth energy moment in the z-direction is $E_n(\mathbf{\Gamma}) \equiv \int d\mathbf{r}\, \epsilon(\mathbf{r}; \mathbf{\Gamma}) z^n$, with $\epsilon(\mathbf{r}; \mathbf{\Gamma})$ being the energy density at $\mathbf{r}$. Also the zeroth temperature is the mid-temperature of the two reservoirs, $T_0^{-1} \equiv [T_+^{-1} + T_-^{-1}]/2 = T^{-1} + \mathcal{O}(\nabla T)^2$, and the first temperature is essentially the temperature gradient imposed by them, $T_1^{-1} \equiv [T_+^{-1} - T_-^{-1}]/L_z = -T^{-2}\nabla T + \mathcal{O}(\nabla T)^2$.

The instantaneous heat flux, a phase function of the isolated sub-system, is essentially the adiabatic rate of change of the first energy moment

$$J_{\mathrm{E}}(\mathbf{\Gamma}) \equiv \dot{E}_1^0(\mathbf{\Gamma})/V,$$

(6.142)

where V is the volume of the sub-system. Due to energy conservation of the isolated system, $\dot{E}_0^0(\mathbf{\Gamma}) = 0$ and $\dot{S}_{\mathrm{r,st}}^0(\mathbf{\Gamma}) = -\dot{E}_1^0(\mathbf{\Gamma})/T_1$.

Fourier's law gives the heat flow in the presence of an applied thermal gradient, and it is (Kubo *et al* 1978, Zwanzig 2001, Bellac *et al* 2004, Pottier 2010, Attard 2012)

$$\bar{J}_{\mathrm{E}} = -\lambda \nabla T,$$

(6.143)

where λ is the thermal conductivity. The left-hand side is the most likely heat flux, which equals the average heat flux. This law of course holds to linear order in the temperature gradient.

The average heat flux given by the present non-equilibrium theory is

$$\langle J_{\mathrm{E}}\rangle_{\mathrm{non-equil}} = \frac{1}{V}\int \mathrm{d}\boldsymbol{\Gamma}\ \wp(\boldsymbol{\Gamma}\mid T_0,T_1)\dot{E}_1^0(\boldsymbol{\Gamma})$$

$$= \frac{1}{V}\frac{\displaystyle\int \mathrm{d}\boldsymbol{\Gamma}\ \mathrm{e}^{[S_{\mathrm{r,st}}(\boldsymbol{\Gamma})+S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma})]/k_{\mathrm{B}}}\dot{E}_1^0(\boldsymbol{\Gamma})}{\displaystyle\int \mathrm{d}\boldsymbol{\Gamma}\ \mathrm{e}^{[S_{\mathrm{r,st}}(\boldsymbol{\Gamma})+S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma})]/k_{\mathrm{B}}}}$$

$$= \frac{1}{Vk_{\mathrm{B}}}\frac{\displaystyle\int \mathrm{d}\boldsymbol{\Gamma}\ \mathrm{e}^{-E_0(\boldsymbol{\Gamma})/k_{\mathrm{B}}T_0}\dot{E}_1^0(\boldsymbol{\Gamma})S_{\mathrm{r,dyn}}^{\mathrm{odd};0}(\boldsymbol{\Gamma})}{\displaystyle\int \mathrm{d}\boldsymbol{\Gamma}\ \mathrm{e}^{-E_0(\boldsymbol{\Gamma})/k_{\mathrm{B}}T_0}} + \mathcal{O}(\nabla T)^2$$

$$= \frac{1}{2Vk_{\mathrm{B}}T_1}\int_{-\tau}^{0}\mathrm{d}t\prime\left\langle \dot{E}_1^0(\boldsymbol{\Gamma})\left[\dot{E}_1^0(\boldsymbol{\Gamma}^0(t\prime|\ \boldsymbol{\Gamma},0) - \dot{E}_1^0(\boldsymbol{\Gamma}^0(t\prime|\ \boldsymbol{\Gamma}^\dagger,0)\right]\right\rangle_{\mathrm{equil}}$$

$$= \frac{-\nabla T}{2Vk_{\mathrm{B}}T_0^2}\int_{-\tau}^{\tau}\mathrm{d}t\prime\left\langle \dot{E}_1^0(\boldsymbol{\Gamma})\dot{E}_1^0(\boldsymbol{\Gamma}^0(t\prime|\ \boldsymbol{\Gamma},0)\right\rangle_{\mathrm{equil}}.$$

(6.1
44)

In the third equality the exponentials have been expanded in powers of the temperature gradient and second order terms have been neglected. As well, terms that are the product of an even parity function, $S_{\mathrm{r,st}}(\boldsymbol{\Gamma})$ or $S_{\mathrm{r,dyn}}^{\mathrm{even}}(\boldsymbol{\Gamma})$, and an odd parity function, $\dot{S}_{\mathrm{r,st}}^0(\boldsymbol{\Gamma})$ or $\dot{E}_1^0(\boldsymbol{\Gamma})$, vanish upon integration over phase space. In addition, the most likely trajectory has been replaced by the adiabatic trajectory, equation (6.98). The equilibrium average arises because $Z(T_0)^{-1}\mathrm{e}^{-E_0(\boldsymbol{\Gamma})/k_{\mathrm{B}}T_0}$ is the Maxwell–Boltzmann distribution. Comparing this to Fourier's law, one can identify the thermal conductivity as

$$\lambda = \frac{1}{2Vk_{\mathrm{B}}T_0^2}\int_{-\tau}^{\tau}\mathrm{d}t\prime\left\langle E_1^0(\boldsymbol{\Gamma})\dot{E}_1^0(\boldsymbol{\Gamma}^0(t\prime|\ \boldsymbol{\Gamma},0)\right\rangle_{\mathrm{equil}}$$

$$= \frac{1}{2Vk_{\mathrm{B}}T_0^2}\left\langle \dot{E}_1^0(t)[E_1(t+\tau) - E_1(t-\tau)]\right\rangle_{\mathrm{equil}}.$$

(6.1
45)

The right-hand side is independent of τ for $\tau \gtrsim \tau_{\mathrm{relax}}$. This can be written in a number of different ways, but all involve the equilibrium time correlation function of the heat flux or an integral thereof. This is a typical example of a Green–Kubo relation (Onsager 1931, Green 1954, Kubo 1966). It is to be noted that the time correlation function in any Green–Kubo relation *always* invokes adiabatic trajectories (Kubo *et al* 1978, Zwanzig 2001, Bellac *et al* 2004, Pottier 2010, Attard 2012).

From this analysis one sees that the general formula for obtaining the Green–Kubo relations for a steady-state thermodynamic system is

$$\left\langle \dot{S}^0_{\mathrm{r,st}}(\boldsymbol{\Gamma}) \right\rangle_{\mathrm{non-equil}} = \int d\boldsymbol{\Gamma}\; \wp(\boldsymbol{\Gamma} \mid x_{\mathrm{r}}, \nabla x_{\mathrm{r}})\, \dot{S}^0_{\mathrm{r,st}}(\boldsymbol{\Gamma})$$

$$= \frac{\displaystyle\int d\boldsymbol{\Gamma}\; e^{[S_{\mathrm{r,st}}(\boldsymbol{\Gamma}\mid x_{\mathrm{r}}, \nabla x_{\mathrm{r}}) + S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}\mid \nabla x_{\mathrm{r}})]/k_{\mathrm{B}}}\, \dot{S}^0_{\mathrm{r,st}}(\boldsymbol{\Gamma})}{\displaystyle\int d\boldsymbol{\Gamma}\; e^{[S_{\mathrm{r,st}}(\boldsymbol{\Gamma}\mid x_{\mathrm{r}}, \nabla x_{\mathrm{r}}) + S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}\mid \nabla x_{\mathrm{r}})]/k_{\mathrm{B}}}}$$

$$= k_{\mathrm{B}}^{-1} \frac{\displaystyle\int d\boldsymbol{\Gamma}\; e^{S_{\mathrm{r,st}}(\boldsymbol{\Gamma}\mid x_{\mathrm{r}})/k_{\mathrm{B}}} S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma} \mid \nabla x_{\mathrm{r}})\, \dot{S}^0_{\mathrm{r,st}}(\boldsymbol{\Gamma})}{\displaystyle\int d\boldsymbol{\Gamma}\; e^{S_{\mathrm{r,st}}(\boldsymbol{\Gamma}\mid x_{\mathrm{r}})/k_{\mathrm{B}}}} + \mathcal{O}(\nabla x_{\mathrm{r}})^2$$

$$= k_{\mathrm{B}}^{-1} \left\langle S^{\mathrm{odd}}_{\mathrm{r,dyn}}(\boldsymbol{\Gamma} \mid \nabla x_{\mathrm{r}})\, \dot{S}^0_{\mathrm{r,st}}(\boldsymbol{\Gamma}) \right\rangle_{\mathrm{equil}}$$

$$\approx k_{\mathrm{B}}^{-1} \left\langle S^{\mathrm{odd};0}_{\mathrm{r,dyn}}(\boldsymbol{\Gamma} \mid \nabla x_{\mathrm{r}})\, \dot{S}^0_{\mathrm{r,st}}(\boldsymbol{\Gamma}) \right\rangle_{\mathrm{equil}}.$$

(6.146)

In the third equality the exponentials have been expanded in powers of the gradient in the reservoir field variable, and second order terms have been neglected. In the denominator the term linear in the gradient vanishes because it has odd spatial symmetry. In the fourth equality the odd temporal parity projection of the dynamic part of the reservoir entropy is the only nonzero contribution because the adiabatic flux has odd temporal parity. In the fifth equality, the most likely trajectory has been replaced by the adiabatic trajectory, equation (6.98). The equilibrium average arises because $S_{\mathrm{r,st}}(\boldsymbol{\Gamma} \mid x_{\mathrm{r}}, \nabla x_{\mathrm{r}})$ is the equilibrium reservoir entropy for the point in phase space. The left-hand side is the flux times the volume averaged in the non-equilibrium system with applied reservoir gradient. The right-hand side is the equilibrium average of that flux multiplied by the temporal integral that is here identified with the dynamic part of the reservoir entropy.

The fact that the theory gives the Green–Kubo relations should give one confidence both in the adiabatic transformation of the dynamic part of the reservoir entropy, equation (6.98), and in the general expression for the phase space probability for non-equilibrium systems, equations (6.4) and (6.14). Of course the Green–Kubo relations are a linear theory, whereas the full expression for the phase space probability for non-equilibrium systems, equations (6.4) and (6.14), applies in all circumstances, linear and non-linear.

6.4 Driven Brownian motion

Consider mechanical non-equilibrium system consisting of a solute that is acted upon by a time-varying potential,

$$U(\mathbf{x}, t) = \frac{\kappa}{2}[\mathbf{x} - \mathbf{b}(t)] \cdot [\mathbf{x} - \mathbf{b}(t)].$$

(6.147)

Here $\mathbf{x}$ is the position of the solute. The potential is in the form of a trap, with $\kappa > 0$. We shall assume that the location of the trap is in uniform motion,

$$\mathbf{b}(t) = \dot{\mathbf{b}}t.$$

(6.148)

All of the solvent is subsumed into the reservoir of temperature T, and so the Hamiltonian of the sub-system is

$$\mathcal{H}(\dot{\mathbf{x}}, \mathbf{x}; t) = \frac{m}{2}\dot{x}^2 + U(\mathbf{x}, t),$$

where m is the mass of the solute. The static part of the reservoir entropy is

$$S_{\mathrm{r,st}}(\dot{\mathbf{x}}, \mathbf{x}; t) = \frac{-\mathcal{H}(\dot{\mathbf{x}}, \mathbf{x}; t)}{T}$$

$$= \frac{-m\dot{x}^2}{2T} - \frac{\kappa}{2T}[\mathbf{x} - \mathbf{b}(t)] \cdot [\mathbf{x} - \mathbf{b}(t)].$$

From physical considerations, for uniform trap velocity, it is most likely that the particle will move with the same velocity as the trap, and that it will lag the trap location such that the mechanical force is equal and opposite to the drag force (see figure 6.7). That is, the most likely configuration can be expected to be

$$\bar{\mathbf{X}}(t) \equiv \{\bar{\mathbf{x}}(t), \bar{\dot{\mathbf{x}}}(t)\} = \{\mathbf{b}(t) - \gamma\dot{\mathbf{b}}/\kappa, \dot{\mathbf{b}}\}.$$

This maximizes the reservoir entropy,

$$\nabla S_{\mathrm{r}}(\mathbf{X}, t)|_{\bar{\mathbf{X}}(t)} = 0.$$

This result will be confirmed by explicit calculation toward the end of the following analysis.
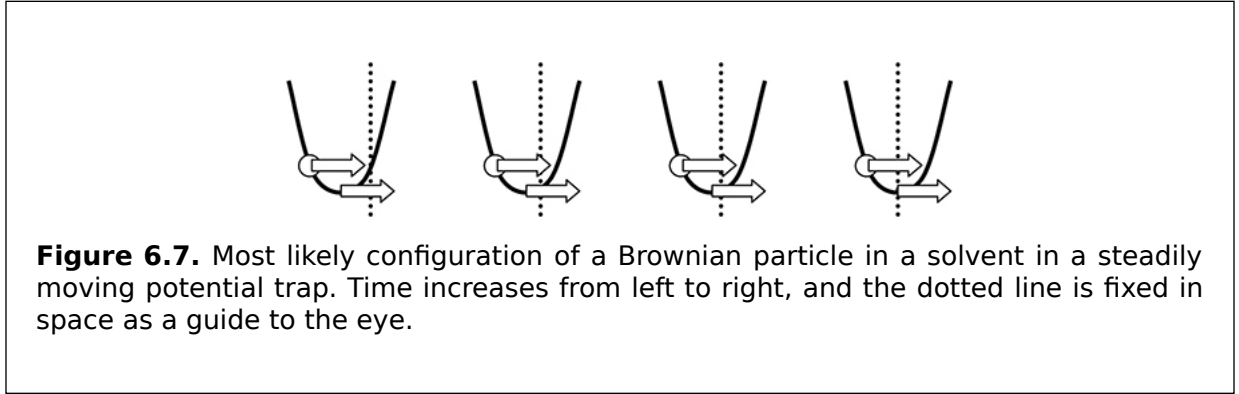
Figure 6.7. Most likely configuration of a Brownian particle in a solvent in a steadily moving potential trap. Time increases from left to right, and the dotted line is fixed in space as a guide to the eye.

The adiabatic rate of static entropy production is

$$\dot{S}^0_{\mathrm{r,st}}(\mathbf{X}, t) = \frac{\partial S_{\mathrm{r,st}}(\mathbf{X}, t)}{\partial t}$$

$$= \frac{\kappa}{T}\dot{\mathbf{b}} \cdot (\mathbf{x} - \mathbf{b}(t)).$$

According to equation (6.117), we should expect that the rate of total entropy production is equal and opposite of the most likely value of this

$$\dot{S}_{\mathrm{tot}}(t) = -\overline{\dot{S}^0_{\mathrm{r,st}}}(t) = -\dot{S}^0_{\mathrm{r,st}}(\bar{\mathbf{X}}(t), t) = \frac{\gamma}{T}\dot{b}^2.$$

From the results of section 6.2.2, the stochastic dissipative equations of motion over an infinitesimal time step are

$$\mathbf{x}(t + \Delta_t) = \mathbf{x}(t) + \Delta_t \dot{\mathbf{x}}(t),$$

$$\dot{\mathbf{x}}(t + \Delta_t) = \dot{\mathbf{x}}(t) - \frac{\Delta_t}{m} \frac{\partial \mathcal{H}(\dot{\mathbf{x}}, \mathbf{x}; t)}{\partial \mathbf{x}} + \mathbf{R}(\dot{\mathbf{x}}, \mathbf{x}; t, \Delta_t)$$

$$= \dot{\mathbf{x}}(t) - \frac{\Delta_t \kappa}{m} [\mathbf{x}(t) - \mathbf{b}(t)] + \bar{\mathbf{R}} + \tilde{\mathbf{R}}.$$

Note that the reservoir 'force' is here in velocity space rather than the more conventional momentum space. Hence it is the actual reservoir force divided by the particle mass.

In the present model all of the solvent is subsumed into the reservoir, and the subsystem consists of a single Brownian particle. In this case it would be an exceedingly poor approximation to replace the dissipative equations of motion by adiabatic equations of motion.

Making contact with Langevin's equation (4.19), specifically in the second entropy form, equation (4.21), the transport matrix can be written in terms of the the drag coefficient, $\Lambda = 2m^{-2}T\gamma\mathbf{I}$. Hence the probability distribution of the stochastic part of the force is

$$\wp(\tilde{\mathbf{R}} \mid \Delta_t) = \frac{e^{-m^2 \tilde{R}^2 / 4\gamma |\Delta_t| k_{\mathrm{B}} T}}{\left(4\pi m^{-2} \gamma \mid \Delta_t \mid k_{\mathrm{B}} T\right)^{3/2}}.$$

The diffusion constant D and the drag coefficient γ are related as $D = k_{\mathrm{B}} T / \gamma$.

According to equation (6.67), the dissipative force is

$$\bar{\mathbf{R}}_v = \frac{\mid \Delta_t \mid}{2} \Lambda \nabla_v S_{\mathrm{r,st}}(\mathbf{X}(t), t) - \frac{\mid \Delta_t \mid - \Delta_t}{2} \Lambda \nabla_v S_{\mathrm{r,st}}(\bar{\mathbf{X}}(t), t)$$

$$= \frac{- \mid \Delta_t \mid \gamma}{m} \dot{\mathbf{x}}(t) + \frac{(\mid \Delta_t \mid - \Delta_t) \gamma}{m} \overline{\dot{\mathbf{x}}}(t)$$

$$= \frac{- \mid \Delta_t \mid \gamma}{m} \left[\dot{\mathbf{x}}(t) - \overline{\dot{\mathbf{x}}}(t)\right] - \frac{\Delta_t \gamma}{m} \overline{\dot{\mathbf{x}}}(t).$$

The usual drag force going forward in time is $\mathbf{F}_{\mathrm{d}} = -\gamma \dot{\mathbf{x}}$, which can be identified in this.

For future comparison, the zeroth order approximation, equation (6.74), which is here $\bar{\mathbf{R}}_v = |\Delta_t| \Lambda \nabla_v S_{\mathrm{r,st}}(\mathbf{\Gamma}, t)/2$, is equivalent to setting the second term of the first and second equalities to zero.

Therefore, the stochastic, dissipative equations of motion for the uniformly driven Brownian particle are explicitly

$$\mathbf{x}(t + \Delta_t) = \mathbf{x}(t) + \Delta_t \dot{\mathbf{x}}(t),$$

$$\dot{\mathbf{x}}(t + \Delta_t) = \dot{\mathbf{x}}(t) - \frac{\Delta_t \kappa}{m} [\mathbf{x}(t) - \mathbf{b}(t)] - \frac{\mid \Delta_t \mid \gamma}{m} \dot{\mathbf{x}} + \frac{(\mid \Delta_t \mid - \Delta_t) \gamma}{m} \dot{\mathbf{b}} + \tilde{\mathbf{R}}_v.$$

Setting $\tilde{\mathbf{R}}_v = 0$, and taking the limit $\Delta_t \to 0$, these dissipative equations may be re-written in the form

$$\frac{\mathrm{d}\bar{\mathbf{x}}(t\prime\mid\mathbf{X}(t),t)}{\mathrm{d}t\prime}\bigg|_{t\prime=t^{\pm}} = \dot{\mathbf{x}}(t),$$

$$\frac{\mathrm{d}\bar{\mathbf{x}}(t\prime\mid\mathbf{X}(t),t)}{\mathrm{d}t\prime}\bigg|_{t\prime=t^{\pm}} = \frac{-\kappa}{m}[\mathbf{x}(t) - \mathbf{b}(t)] - \frac{\hat{\Delta}_t\gamma}{m}\dot{\mathbf{x}}(t) + \frac{(\hat{\Delta}_t - 1)\gamma}{m}\dot{\mathbf{b}},$$

(6.1 59)

where $\hat{\Delta}_t = \mathrm{sign}\ \Delta_t$.

Since $\mathbf{b}(t) = \dot{\mathbf{b}}t$, in vector form this is

$$\dot{\bar{\mathbf{X}}}(t\mid\mathbf{X}(t),t) = \mathbf{A}\mathbf{X}(t) + \mathbf{B}t + \mathbf{C},$$

(6.1 60)

with

$$\mathbf{A} \equiv \begin{pmatrix} 0 & \mathbf{I} \\ (-\kappa/m)\mathbf{I} & (-\hat{\Delta}_t\gamma/m)\mathbf{I} \end{pmatrix},$$

$$\mathbf{B} \equiv \{0, (\kappa/m)\dot{\mathbf{b}}\},$$

$$\text{and} \quad \mathbf{C} \equiv \{0, ((\hat{\Delta}_t - 1)\gamma/m)\dot{\mathbf{b}}\}.$$

(6.1 61)

In these matrices and vectors, the three components of direction are independent and can be considered separately. On a backward trajectory, $\hat{\Delta}_t = -1$, and $\mathbf{C} = \{0, (-2\gamma/m)\dot{\mathbf{b}}\}$.

Again for future comparison, the zeroth order approximation, equation (6.74), takes $\mathbf{C} = 0$.

As can be confirmed by direct substitution, the solution to the differential equation is

$$\bar{\mathbf{X}}(t\prime\mid\mathbf{X},t) = \mathrm{e}^{(t\prime-t)\mathbf{A}}\left[\mathbf{X} + t\mathbf{A}^{-1}\mathbf{B} + \mathbf{A}^{-1}\mathbf{C} + \mathbf{A}^{-2}\mathbf{B}\right]$$
$$- t\prime\mathbf{A}^{-1}\mathbf{B} - \mathbf{A}^{-1}\mathbf{C} - \mathbf{A}^{-2}\mathbf{B}.$$

(6.1 62)

As mentioned, the three components of direction are independent from each other.

For future reference, the eigenvalues of $\mathbf{A}$ are given by the characteristic equation

$$\lambda_{\pm}\left[\lambda_{\pm} + \frac{\hat{\Delta}_t\gamma}{m}\right] + \frac{\kappa}{m} = 0,$$

(6.1 63)

or

$$\lambda_{\pm} = \frac{-\hat{\Delta}_t\gamma}{2m}\left\{1 \pm \sqrt{1 - 4\kappa m/\gamma^2}\right\}$$
$$\sim \left\{\frac{-\hat{\Delta}_t\gamma}{m}, \frac{-\hat{\Delta}_t\kappa}{\gamma}\right\}, \quad \kappa \to 0.$$

On a backwards trajectory, $\hat{\Delta}_t = -1$, both eigenvalues are positive (assuming $4\kappa m/\gamma^2 < 1$).

Using the adiabatic rate of static entropy production, $\dot{S}^0_{\text{r,st}}(\mathbf{X}, t) = \kappa \dot{\mathbf{b}} \cdot [\mathbf{x}(t) - \mathbf{b}(t)]/T$, and the full expression for the most likely backward trajectory, the dynamic part of the reservoir entropy is

$$
\begin{aligned}
S_{\text{r,dyn}}(\mathbf{X}, t) &= -\int_0^t \mathrm{d}t\prime\, \dot{S}^0_{\text{r,st}}(\bar{\mathbf{X}}(t\prime|\,\mathbf{X}, t), t\prime) \\
&= \frac{-\kappa}{T}\dot{\mathbf{b}} \cdot \int_0^t \mathrm{d}t\prime\, \left[\bar{\mathbf{x}}(t\prime|\,\mathbf{X}, t) - \dot{\mathbf{b}}t\prime\right] \\
&= \frac{\kappa \dot{\mathbf{b}}^2 t^2}{2T} - \frac{\kappa}{T}\dot{\mathbf{b}} \cdot \left\{ \mathbf{A}^{-1}\left[\mathbf{I} - \mathrm{e}^{-t\mathbf{A}}\right]\left[\mathbf{X} + t\mathbf{A}^{-1}\mathbf{B} + \mathbf{A}^{-1}\mathbf{C} + \mathbf{A}^{-2}\mathbf{B}\right]\right. \\
&\qquad \left. - \frac{t^2}{2}\mathbf{A}^{-1}\mathbf{B} - t\mathbf{A}^{-1}\mathbf{C} - t\mathbf{A}^{-2}\mathbf{B}\right\}_{\mathbf{x}}.
\end{aligned}
$$

Note that because this is the most likely backward trajectory in the integrand, $\hat{\Delta}_t = -1$ in $\mathbf{A}$ and $\mathbf{C}$. Assuming both eigenvalues are positive, since $t > 0$ the exponential term may be neglected and this becomes

$$
S_{\text{r,dyn}}(\mathbf{X}, t) = \frac{\kappa \dot{\mathbf{b}}^2 t^2}{2T} - \frac{\kappa}{T}\dot{\mathbf{b}} \cdot \left\{ \mathbf{A}^{-1}\mathbf{X} + \mathbf{A}^{-2}\mathbf{C} + \mathbf{A}^{-3}\mathbf{B} - \frac{t^2}{2}\mathbf{A}^{-1}\mathbf{B} - t\mathbf{A}^{-1}\mathbf{C}\right\}_{\mathbf{x}}.
$$

As has been mentioned, the zeroth order approximation, equation (6.74), takes $\mathbf{C} = 0$.

The term quadratic in t in the braces is

$$
\begin{aligned}
\{\mathbf{A}^{-1}\mathbf{B}\}_{\mathbf{x}} &= \left\{ \begin{pmatrix} 0 & \mathbf{I} \\ (-\kappa/m)\mathbf{I} & (\gamma/m)\mathbf{I} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \kappa\dot{\mathbf{b}}/m \end{pmatrix}\right\}_{\mathbf{x}} \\
&= \frac{m}{\kappa}\left\{ \begin{pmatrix} (\gamma/m)\mathbf{I} & -\mathbf{I} \\ (\kappa/m)\mathbf{I} & 0 \end{pmatrix} \begin{pmatrix} 0 \\ \kappa\dot{\mathbf{b}}/m \end{pmatrix}\right\}_{\mathbf{x}} \\
&= -\dot{\mathbf{b}}.
\end{aligned}
$$

This term is equal and opposite the term outside the braces and it therefore cancels with it.

The term linear in time is

$$
\begin{aligned}
\{\mathbf{A}^{-1}\mathbf{C}\}_{\mathbf{x}} &= \frac{m}{\kappa}\left\{ \begin{pmatrix} (\gamma/m)\mathbf{I} & -\mathbf{I} \\ (\kappa/m)\mathbf{I} & 0 \end{pmatrix} \begin{pmatrix} 0 \\ -2\gamma\dot{\mathbf{b}}/m \end{pmatrix}\right\}_{\mathbf{x}} \\
&= \frac{2\gamma}{\kappa}\dot{\mathbf{b}}.
\end{aligned}
$$

One of the terms independent of time and of $\mathbf{X}$ is

$$\{\mathbf{A}^{-3}\mathbf{B}\}_{\mathbf{x}} = \frac{m^3}{\kappa^3}\left\{\begin{pmatrix}(\gamma/m)\mathbf{I} & -\mathbf{I}\\(\kappa/m)\mathbf{I} & 0\end{pmatrix}^3\begin{pmatrix}0\\ \kappa\dot{\mathbf{b}}/m\end{pmatrix}\right\}_{\mathbf{x}}$$

$$= \frac{m^3}{\kappa^3}\left\{\begin{pmatrix}(\gamma/m)\mathbf{I} & -\mathbf{I}\\(\kappa/m)\mathbf{I} & 0\end{pmatrix}^2\begin{pmatrix}-\kappa\dot{\mathbf{b}}/m\\ 0\end{pmatrix}\right\}_{\mathbf{x}}$$

$$= \frac{m^3}{\kappa^3}\left\{\begin{pmatrix}(\gamma/m)\mathbf{I} & -\mathbf{I}\\(\kappa/m)\mathbf{I} & 0\end{pmatrix}\begin{pmatrix}-\gamma\kappa\dot{\mathbf{b}}/m^2\\ -\kappa^2\dot{\mathbf{b}}/m^2\end{pmatrix}\right\}_{\mathbf{x}}$$

$$= \frac{m^3}{\kappa^3}\left[\frac{-\gamma^2\kappa}{m^3} + \frac{\kappa^2}{m^2}\right]\dot{\mathbf{b}}$$

$$= \left[\frac{m}{\kappa} - \frac{\gamma^2}{\kappa^2}\right]\dot{\mathbf{b}}.$$

(6.1
69)

The other term independent of time and of $\mathbf{X}$ is

$$\{\mathbf{A}^{-2}\mathbf{C}\}_{\mathbf{x}} = \frac{m^2}{\kappa^2}\left\{\begin{pmatrix}(\gamma/m)\mathbf{I} & -\mathbf{I}\\(\kappa/m)\mathbf{I} & 0\end{pmatrix}^2\begin{pmatrix}0\\ -2\gamma\dot{\mathbf{b}}/m\end{pmatrix}\right\}_{\mathbf{x}}$$

$$= \frac{m^2}{\kappa^2}\left\{\begin{pmatrix}(\gamma/m)\mathbf{I} & -\mathbf{I}\\(\kappa/m)\mathbf{I} & 0\end{pmatrix}\begin{pmatrix}2\gamma\dot{\mathbf{b}}/m\\ 0\end{pmatrix}\right\}_{\mathbf{x}}$$

$$= \frac{2\gamma^2}{\kappa^2}\dot{\mathbf{b}}.$$

(6.1
70)

The term dependent on $\mathbf{X}$ is

$$\{\mathbf{A}^{-1}\mathbf{X}\}_{\mathbf{x}} = \frac{m}{\kappa}\left\{\begin{pmatrix}(\gamma/m)\mathbf{I} & -\mathbf{I}\\(\kappa/m)\mathbf{I} & 0\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \dot{\mathbf{x}}\end{pmatrix}\right\}_{\mathbf{x}}$$

$$= \frac{\gamma}{\kappa}\mathbf{x} - \frac{m}{\kappa}\dot{\mathbf{x}}.$$

(6.1
71)

Putting these together the dynamic part of the reservoir entropy is

$$S_{\mathrm{r,dyn}}(\mathbf{X}, t) = \frac{\kappa\dot{b}^2 t^2}{2T} - \frac{\kappa}{T}\dot{\mathbf{b}}\cdot\left\{\mathbf{A}^{-1}\mathbf{X} + \mathbf{A}^{-2}\mathbf{C} + \mathbf{A}^{-3}\mathbf{B} - \frac{t^2}{2}\mathbf{A}^{-1}\mathbf{B} - t\mathbf{A}^{-1}\mathbf{C}\right\}_{\mathbf{x}}$$

$$= \left[\frac{\gamma^2}{\kappa T} - \frac{m}{T}\right]\dot{b}^2 - \frac{\gamma}{T}\dot{\mathbf{b}}\cdot\mathbf{x} + \frac{m}{T}\dot{\mathbf{b}}\cdot\dot{\mathbf{x}} - \frac{2\gamma^2}{\kappa T}\dot{b}^2 + t\frac{2\gamma}{T}\dot{b}^2$$

$$= \frac{-\gamma}{T}\dot{\mathbf{b}}\cdot\left[\mathbf{x} - \mathbf{b}(t) + \frac{\gamma}{\kappa}\dot{\mathbf{b}}\right] + \frac{m}{T}\dot{\mathbf{b}}\cdot[\dot{\mathbf{x}} - \dot{\mathbf{b}}] + t\frac{\gamma}{T}\dot{b}^2.$$

(6.1
72)

The final equality uses the fact that $\mathbf{b}(t) = \dot{\mathbf{b}}t$. The gradient of this is constant in phase space, $\nabla_v S_{\mathrm{r,dyn}}(\mathbf{X}, t) = m\dot{\mathbf{b}}/T = -\nabla_v S_{\mathrm{r,st}}(\bar{\mathbf{X}}, t)$. This is consistent with the approximation equation (6.64), which underpins the expression for the most likely reservoir force, equation (6.67). The final constant linear in time is the reservoir entropy produced to date, equation (6.154), as will be confirmed shortly.

The zeroth order approximation, equation (6.74), sets $\mathbf{C} = 0$, in which case this reduces to

$$
\begin{aligned}
S_{\mathrm{r,dyn}}(\mathbf{X}, t) &= \frac{\kappa \dot{b}^2 t^2}{2T} - \frac{\kappa}{T}\dot{\mathbf{b}} \cdot \left\{ \mathbf{A}^{-1}\mathbf{X} + \mathbf{A}^{-3}\mathbf{B} - \frac{t^2}{2}\mathbf{A}^{-1}\mathbf{B} \right\}_{\mathbf{x}} \\
&= \left[\frac{\gamma^2}{\kappa T} - \frac{m}{T} \right]\dot{b}^2 - \frac{\gamma}{T}\dot{\mathbf{b}} \cdot \mathbf{x} + \frac{m}{T}\dot{\mathbf{b}} \cdot \dot{\mathbf{x}} \\
&= \frac{-\gamma}{T}\dot{\mathbf{b}} \cdot \left[\mathbf{x} - \mathbf{b}(t) - \frac{\gamma}{\kappa}\dot{\mathbf{b}} \right] + \frac{m}{T}\dot{\mathbf{b}} \cdot [\dot{\mathbf{x}} - \dot{\mathbf{b}}].
\end{aligned}
$$

(6.1 73)

Compared to the full result, this has lost the term linear in time, which represents the important physical quantity of the entropy produced to date, and it has changed the first term in brackets. Both of these changes are constants in phase space, so in fact the zeroth order approximation, equation (6.74), gives the same relative probability for points in phase space as the full expression, equation (6.67).

The full result for the dynamic part of the reservoir entropy is *not* odd with respect to velocity reversal, nor with respect to reflection in the trap minimum. The even terms that would be projected out are significant in these two individual projections. It is not *exactly* odd with respect to both operations simultaneously, but the even part in this case is immaterial. That is, the odd projection of this with respect to both operations is

$$
\begin{aligned}
S_{\mathrm{r,dyn}}^{\mathrm{odd}}(\mathbf{X}, t) &\equiv \frac{1}{2}\left[S_{\mathrm{r,dyn}}(\mathbf{X}, t) - S_{\mathrm{r,dyn}}(\mathbf{X}^{\dagger\ddagger}, t) \right] \\
&= -\frac{\gamma}{T}\dot{\mathbf{b}} \cdot [\mathbf{x} - \mathbf{b}(t)] + \frac{m}{T}\dot{\mathbf{b}} \cdot \dot{\mathbf{x}}.
\end{aligned}
$$

(6.1 74)

This differs from the full expression by constant terms (one of which is the even asymptotic contribution). One can see that it is necessary to perform both operations simultaneously otherwise one or other of $\mathbf{x}$ or $\dot{\mathbf{x}}$ is projected out by the operation. As far as determining the relative probability of phase space points is concerned, this odd projection with respect to both operations is equally as good as the full expression.

The static part of the reservoir entropy,

$$
S_{\mathrm{r,st}}(\mathbf{X}, t) = \frac{-m}{2T}\dot{\mathbf{x}} \cdot \dot{\mathbf{x}} - \frac{\kappa}{2T}[\mathbf{x} - \mathbf{b}(t)] \cdot [\mathbf{x} - \mathbf{b}(t)],
$$

(6.1 75)

added to the dynamic part of the reservoir entropy gives the reservoir entropy,

$$S_r(\mathbf{X}, t) = \frac{-m}{2T}\dot{\mathbf{x}} \cdot \dot{\mathbf{x}} - \frac{\kappa}{2T}[\mathbf{x} - \mathbf{b}(t)] \cdot [\mathbf{x} - \mathbf{b}(t)]$$
$$- \frac{\gamma}{T}\dot{\mathbf{b}} \cdot \left[\mathbf{x} - \mathbf{b}(t) + \frac{\gamma}{\kappa}\dot{\mathbf{b}}\right] + \frac{m}{T}\dot{\mathbf{b}} \cdot [\dot{\mathbf{x}} - \dot{\mathbf{b}}] + t\frac{\gamma}{T}\dot{b}^2$$
$$= \frac{-m}{2T}[\dot{\mathbf{x}} - \dot{\mathbf{b}}] \cdot [\dot{\mathbf{x}} - \dot{\mathbf{b}}]$$
$$- \frac{\kappa}{2T}\left[\mathbf{x} - \mathbf{b}(t) + \frac{\gamma}{\kappa}\dot{\mathbf{b}}\right] \cdot \left[\mathbf{x} - \mathbf{b}(t) + \frac{\gamma}{\kappa}\dot{\mathbf{b}}\right]$$
$$+ t\frac{\gamma}{T}\dot{b}^2 - \frac{m}{2T}\dot{b}^2 - \frac{\kappa}{2T}\frac{\gamma^2}{\kappa^2}\dot{b}^2.$$

(6.176)

By inspection, one can see that this is a maximum with respect to $\mathbf{x}$ when $\bar{\mathbf{x}}(t) = \mathbf{b}(t) - \gamma\dot{\mathbf{b}}/\kappa$, and it is a maximum with respect to $\dot{\mathbf{x}}$ when $\bar{\dot{\mathbf{x}}}(t) = \dot{\mathbf{b}}$. This confirms the result deduced on physical grounds in equation (6.151).

The partial time derivative of this in the most likely state is

$$\frac{\partial S_r(\mathbf{X}, t)}{\partial t}\Big|_{\mathbf{X}=\bar{\mathbf{X}}(t)} = \frac{\gamma}{T}\dot{b}^2.$$

(6.177)

This is equal to the rate of total entropy production, and is equal and opposite to the adiabatic rate of static entropy production, equation (6.154).

The $\mathbf{X}$-dependent part of the reservoir entropy is in the form of a fluctuation entropy about the most likely state. It would give a Gaussian probability distribution. The fluctuation coefficients (i.e. $-m/2T$ for the velocity, $-\kappa/2T$ for the position) come from the second derivatives of the static entropy, and they would be the same in the corresponding equilibrium system.

6.5 Path entropy and trajectory probability

6.5.1 Path entropy

For a transition $\{\boldsymbol{\Gamma}_1, t_1\} \to \{\boldsymbol{\Gamma}_2, t_2\}$, the actual change in reservoir entropy is

$$\vec{\Delta} S_r(\boldsymbol{\Gamma}_2, t_2; \boldsymbol{\Gamma}_1, t_1) \equiv S_r(\boldsymbol{\Gamma}_2, t_2 \mid \boldsymbol{\Gamma}_1, t_1) - S_r(\boldsymbol{\Gamma}_1, t_1)$$
$$= \mathbf{R}_p \cdot \nabla_p S_{r,st}(\boldsymbol{\Gamma}, t)$$
$$= \left[\boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_1 - t_{21}\dot{\boldsymbol{\Gamma}}^0\right] \cdot \nabla S_{r,st}(\boldsymbol{\Gamma}, t)$$
$$= S_{r,st}(\boldsymbol{\Gamma}_2, t_2) - S_{r,st}(\boldsymbol{\Gamma}_1, t_1) - t_{21}\dot{S}^0_{r,st}(\boldsymbol{\Gamma}, t).$$

(6.178)

Here $S_r(\boldsymbol{\Gamma}_2, t_2 \mid \boldsymbol{\Gamma}_1, t_1)$ is the conditional reservoir entropy, which is the entropy for the point $\{\boldsymbol{\Gamma}_2, t_2\}$ given that the system had previously been at $\{\boldsymbol{\Gamma}_1, t_1\}$. Here and in what follows, the forward trajectory is considered, $t_2 > t_1$. Also, $\mathbf{R}$ is the actual force from the reservoir, and so by the conservation laws, $\mathbf{R} \cdot \nabla S_{r,st}(\boldsymbol{\Gamma}, t)$ is the actual change in reservoir entropy. We always work to linear order in the time step and so $\{\boldsymbol{\Gamma}, t\}$ can be anywhere on the interval. The final two equalities are simply rearrangements that make the adiabatic contributions explicit.

Note that this, $\vec{\Delta}S_\mathrm{r}$ the actual change in entropy on an actual transition, is not the same as the difference in entropy, $\Delta S_\mathrm{r} \equiv S_\mathrm{r}(\mathbf{\Gamma}_2, t_2) - S_\mathrm{r}(\mathbf{\Gamma}_1, t_1)$. Both quantities in the latter are calculated as if the phase space point had been reached via the most likely trajectory (see figure 6.8).

Figure 6.8. The actual transition between two points in phase space (dashed arrow), and the two most likely trajectories leading to the points (solid arrows). The change in entropy, $\vec{\Delta}S_\mathrm{r}$, is calculated on the former, whereas the difference in entropy, ΔS_r, is calculated using the latter.
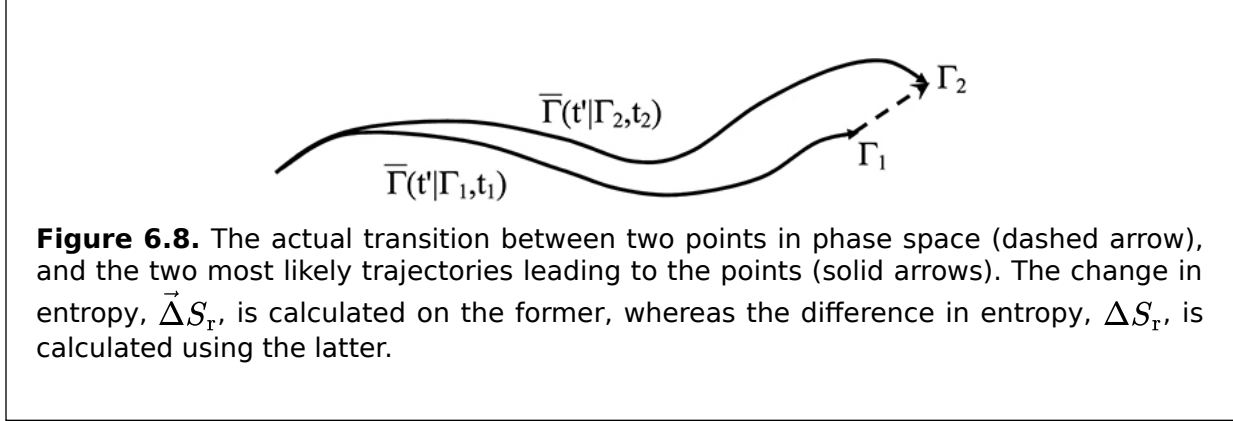
From equation (6.67), the most likely reservoir force for this transition is

$$\bar{\mathbf{R}}_p = \frac{\mid t_{21} \mid}{2} \Lambda \left[\nabla_p S_\mathrm{r,st}(\mathbf{\Gamma}, t) - \nabla_p S_\mathrm{r,st}(\bar{\mathbf{\Gamma}}(t), t) \right] + \frac{t_{21}}{2} \Lambda \nabla_p S_\mathrm{r,st}(\bar{\mathbf{\Gamma}}(t), t)$$

$$= \frac{t_{21}}{2} \Lambda \nabla_p S_\mathrm{r,st}(\mathbf{\Gamma}, t),$$

(6.179)

since $t_{21} > 0$.

In view of equations (6.52) and (6.53), the second entropy for the transition may be written

$$S^{(2)}(\mathbf{\Gamma}_2, t_2; \mathbf{\Gamma}_1, t_1) = \frac{-\Lambda^{-1} : \tilde{\mathbf{R}}_p \tilde{\mathbf{R}}_p}{2 \mid t_{21} \mid} + S_\mathrm{r}(\mathbf{\Gamma}_1, t_1) + \frac{1}{2}[S_\mathrm{tot}(t_2) - S_\mathrm{tot}(t_1)].$$

(6.180)

This neglects the (constant) contribution from the fluctuations, the normalizing factor for the Gaussian. Writing the random part of the reservoir force as $\tilde{\mathbf{R}}_p = \mathbf{R}_p - \bar{\mathbf{R}}_p$, the conditional second entropy can be defined and re-written as

$$S^{(2)}(\mathbf{\Gamma}_2, t_2 \mid \mathbf{\Gamma}_1, t_1) \equiv S^{(2)}(\mathbf{\Gamma}_2, t_2; \mathbf{\Gamma}_1, t_1) - S_\mathrm{r}(\mathbf{\Gamma}_1, t_1) - \frac{1}{2}[S_\mathrm{tot}(t_2) - S_\mathrm{tot}(t_1)]$$

$$= \frac{-1}{2 \mid t_{21} \mid} \Lambda^{-1} : \tilde{\mathbf{R}}_p \tilde{\mathbf{R}}_p$$

$$= \frac{-1}{2 \mid t_{21} \mid} \Lambda^{-1} : \left[\mathbf{R}_p \mathbf{R}_p + \bar{\mathbf{R}}_p \bar{\mathbf{R}}_p \right] + \frac{1}{\mid t_{21} \mid} \Lambda^{-1} : \mathbf{R}_p \bar{\mathbf{R}}_p$$

$$= \frac{-1}{2 \mid t_{21} \mid} \Lambda^{-1} : \left[\mathbf{R}_p \mathbf{R}_p + \bar{\mathbf{R}}_p \bar{\mathbf{R}}_p \right] + \frac{1}{2} \mathbf{R}_p \cdot \nabla_p S_\mathrm{r,st}(\mathbf{\Gamma}, t)$$

$$= \frac{-1}{2 \mid t_{21} \mid} \Lambda^{-1} : \left[\mathbf{R}_p \mathbf{R}_p + \bar{\mathbf{R}}_p \bar{\mathbf{R}}_p \right] + \frac{1}{2} \vec{\Delta} S_\mathrm{r}(\mathbf{\Gamma}_2, t_2; \mathbf{\Gamma}_1, t_1).$$

Since $\bar{\mathbf{R}}_p$ and $\mathbf{R}_p$ scale with the size of the time step, the conditional second entropy itself also scales with the size of the time step, as can also be seen directly from the definition.

This result for a single transition is now extended to a trajectory, which is a series of such transitions. Denote the trajectory by $\underline{\mathbf{\Gamma}}$. Divide the interval $[t_0, t_f]$ of the trajectory into uniformly spaced nodes, $t_n = t_0 + n\Delta_t$, $n = 0, 1, \ldots, f$, with $\mathbf{\Gamma}_n = \mathbf{\Gamma}(t_n)$ and the time step being $\Delta_t = [t_f - t_0]/f > 0$. We assume that the change in entropy for each nodal transition is independent of the prior trajectory of the system. With these, the actual change in entropy for this trajectory is

$$
\begin{aligned}
\vec{\Delta} S_{\mathrm{r}}(\underline{\mathbf{\Gamma}}) &= \sum_{n=0}^{f-1} \vec{\Delta} S_{\mathrm{r}}(\mathbf{\Gamma}_{n+1}, t_{n+1}; \mathbf{\Gamma}_n, t_n) \\
&= \sum_{n=0}^{f-1} \mathbf{R}_n \cdot \nabla_p S_{\mathrm{r,st}}(\mathbf{\Gamma}_n, t_n) \\
&= \sum_{n=0}^{f-1} \left\{ S_{\mathrm{r,st}}(\mathbf{\Gamma}_{n+1}, t_{n+1}) - S_{\mathrm{r,st}}(\mathbf{\Gamma}_n, t_n) - \Delta_t \dot{S}_{\mathrm{r,st}}^0(\mathbf{\Gamma}_n, t_n) \right\} \\
&= S_{\mathrm{r,st}}(\mathbf{\Gamma}_f, t_f) - S_{\mathrm{r,st}}(\mathbf{\Gamma}_0, t_0) - \int_{t_0}^{t_f} \mathrm{d}t\prime\, \dot{S}_{\mathrm{r,st}}^0(\mathbf{\Gamma}(t\prime), t\prime) \\
&\equiv S_{\mathrm{r,st}}(\mathbf{\Gamma}_f, t_f) - S_{\mathrm{r,st}}(\mathbf{\Gamma}_0, t_0) + S_{\mathrm{r,dyn}}(\underline{\mathbf{\Gamma}}).
\end{aligned}
$$

Here $\mathbf{R}_n$ is the actual reservoir force applied at the nth node, and the final equality defines the dynamic part of the trajectory entropy. This is sketched in figure 6.9.
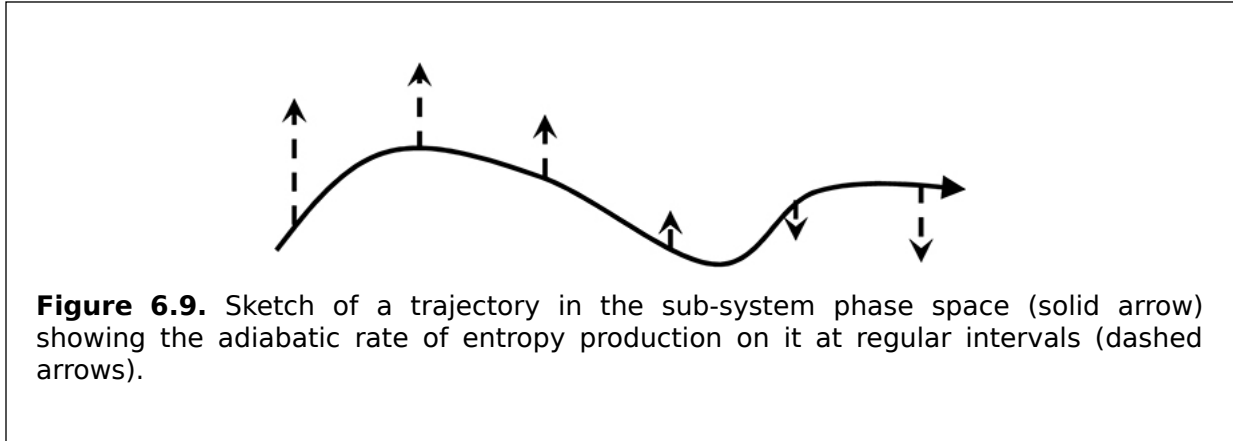
Figure 6.9. Sketch of a trajectory in the sub-system phase space (solid arrow) showing the adiabatic rate of entropy production on it at regular intervals (dashed arrows).

It is a significant assumption that the change in entropy for each nodal interval is independent of the preceding trajectory. This assumes that on the present infinitesimal time steps, Hamilton's equations apply to both the sub-system and the reservoir, and these are strictly Markovian. Of course a certain statistical re-summation of the reservoir perturbation has taken place in the formulations of $\bar{\mathbf{R}}$ and of $\tilde{\mathbf{R}}$. But this is not same as, for example, the integration of fast modes in Brownian dynamics, where memory effects can be important.

In view of the Markov assumption, the trajectory entropy is just the sum of the second entropy for the transition between each node,

$$S_{\mathrm{r}}(\underline{\boldsymbol{\Gamma}} \mid \boldsymbol{\Gamma}_0, t_0) = \sum_{n=0}^{f-1} S^{(2)}(\boldsymbol{\Gamma}_{n+1}, t_{n+1} \mid \boldsymbol{\Gamma}_n, t_n)$$

$$= \sum_{n=0}^{f-1} \frac{-1}{2 \mid \Delta_t \mid} \Lambda^{-1} : \tilde{\mathbf{R}}_n \tilde{\mathbf{R}}_n$$

$$= \sum_{n=0}^{f-1} \left\{ \frac{-\Lambda^{-1} :}{2 \mid \Delta_t \mid} \left[\mathbf{R}_n \mathbf{R}_n + \bar{\mathbf{R}}_n \bar{\mathbf{R}}_n \right] + \frac{1}{2} \vec{\Delta} S_{\mathrm{r}}(\boldsymbol{\Gamma}_{n+1}, t_{n+1}; \boldsymbol{\Gamma}_n, t_n) \right\}$$

$$= \frac{1}{2} \vec{\Delta} S_{\mathrm{r}}(\underline{\boldsymbol{\Gamma}}) - \frac{1}{2 \mid \Delta_t \mid} \sum_{n=0}^{f-1} \Lambda^{-1} : \left[\mathbf{R}_n \mathbf{R}_n + \bar{\mathbf{R}}_n \bar{\mathbf{R}}_n \right].$$

(6.1
83)

Here the total reservoir force at the nth node is $\mathbf{R}_n = \boldsymbol{\Gamma}_{n+1} - \boldsymbol{\Gamma}_n - \Delta_t \dot{\boldsymbol{\Gamma}}_n^0$.

The probability of the trajectory conditioned on the starting point is just the exponential of this trajectory entropy. This is usefully written as the product of Gaussians, the second equality above,

$$\wp(\underline{\boldsymbol{\Gamma}} \mid \boldsymbol{\Gamma}_0, t_0) = \prod_{n=0}^{f-1} \wp(\boldsymbol{\Gamma}_{n+1}, t_{n+1} \mid \boldsymbol{\Gamma}_n, t_n)$$

$$= \prod_{n=0}^{f-1} \frac{\mathrm{e}^{S^{(2)}(\boldsymbol{\Gamma}_{n+1}, t_{n+1} \mid \boldsymbol{\Gamma}_n, t_n)/k_{\mathrm{B}}}}{Z(t_{n+1}, t_n)}$$

$$= \prod_{n=0}^{f-1} \frac{\Theta_\Lambda(\tilde{\mathbf{R}}_n)}{\left[\mathrm{Det} \ 2\pi \mid \Delta_t \mid k_{\mathrm{B}} \Lambda \right]^{1/2}}.$$

(6.1
84)

Here and below, the un-normalized Gaussian distribution is

$$\Theta_\Lambda(\mathbf{R}_p) \equiv \mathrm{e}^{-\Lambda_{pp}^{-1} : \mathbf{R}_p \mathbf{R}_p / 2 k_{\mathrm{B}} \mid \Delta_t \mid}.$$

(6.1
85)

By inspection, one sees that this expression for the conditional trajectory probability is correctly normalized.

This result for the conditional trajectory probability is based on the second equality for the trajectory entropy. The conditional trajectory probability can also be written in terms of the change in entropy over the trajectory, which occurs in the final equality for the trajectory entropy,

$$\wp(\underline{\boldsymbol{\Gamma}} \mid \boldsymbol{\Gamma}_0, t_0) = \frac{1}{Z(\underline{t})} \mathrm{e}^{S_{\mathrm{r}}(\underline{\boldsymbol{\Gamma}} \mid \boldsymbol{\Gamma}, t_0)/k_{\mathrm{B}}}$$

$$= \mathrm{e}^{[S_{\mathrm{r,st}}(\boldsymbol{\Gamma}_f, t_f) - S_{\mathrm{r,st}}(\boldsymbol{\Gamma}_0, t_0)]/2 k_{\mathrm{B}}} \mathrm{e}^{S_{\mathrm{r,dyn}}(\underline{\boldsymbol{\Gamma}})/2 k_{\mathrm{B}}}$$

$$\times \prod_{n=0}^{f-1} \frac{\Theta_\Lambda(\mathbf{R}_n) \Theta_\Lambda(\bar{\mathbf{R}}_n)}{\left[\mathrm{Det} \ 2\pi \mid \Delta_t \mid k_{\mathrm{B}} \Lambda \right]^{1/2}}.$$

(6.1
86)

Because the exponent here is equal to the previous Gaussian form, this must also be correctly normalized.

It is usually the case that the initial probability distribution is that of the non-equilibrium system at time t_0,

$$\wp(\mathbf{\Gamma}, t_0) = \frac{1}{Z(t_0)} e^{S_r(\mathbf{\Gamma}, t_0)/k_B} = \frac{1}{Z(t_0)} e^{[S_{r,st}(\mathbf{\Gamma}, t_0) + S_{r,dyn}(\mathbf{\Gamma}, t_0)]/k_B}.$$

<div align="right">(6.1 87)</div>

For convenience, sometimes one instead assumes a dynamically disordered system, namely one that is initially distributed according to the static part of the reservoir entropy alone,

$$\wp_{st}(\mathbf{\Gamma}, t_0) = \frac{1}{Z_{st}(t_0)} e^{S_{r,st}(\mathbf{\Gamma}, t_0)/k_B}.$$

<div align="right">(6.1 88)</div>

With this latter assumption the unconditional trajectory probability in the second form is

$$\begin{aligned} \wp(\underline{\mathbf{\Gamma}}) &= \wp(\underline{\mathbf{\Gamma}} \mid \mathbf{\Gamma}_0, t_0) \wp_{st}(\mathbf{\Gamma}_0, t_0) \\ &= \frac{1}{Z_{st}(t_0)} e^{[S_{r,st}(\mathbf{\Gamma}_f, t_f) + S_{r,st}(\mathbf{\Gamma}_0, t_0)]/2k_B} e^{S_{r,dyn}(\underline{\mathbf{\Gamma}})/2k_B} \\ &\quad \times \prod_{n=0}^{f-1} \frac{\Theta_\Lambda(\mathbf{R}_n) \Theta_\Lambda(\bar{\mathbf{R}}_n)}{[\text{Det } 2\pi \mid \Delta_t \mid k_B \Lambda]^{1/2}}. \end{aligned}$$

<div align="right">(6.1 89)</div>

For a long trajectory, the actual initial distribution of states makes little difference to the unconditional trajectory probability.

6.5.1.1 Conjugate trajectory

We now wish to analyse the properties of the conjugate trajectory upon which the velocities are reversed. To this end, in the event that the Hamiltonian is explicitly time dependent, we define the conjugate Hamiltonian on the trajectory time interval as $\tilde{\mathcal{H}}(\mathbf{\Gamma}, t) = \mathcal{H}(\mathbf{\Gamma}, t_0 + t_f - t)$. This is the same conjugation operation as was defined in equation (5.16) for the discussion of microscopic reversibility for a time-dependent potential. With † representing velocity reversal, the phase space point $\mathbf{\Gamma} = \{\mathbf{q}, \mathbf{p}\}$ has conjugate $\mathbf{\Gamma}^\dagger = \{\mathbf{q}, -\mathbf{p}\}$. The trajectory in the conjugate system, $\underline{\tilde{\mathbf{\Gamma}}}$, has the property that $\tilde{\mathbf{\Gamma}}(t) = \mathbf{\Gamma}(t_0 + t_f - t)^\dagger$. Equivalently, $\tilde{\mathbf{\Gamma}}_n = \mathbf{\Gamma}_{f-n}^\dagger$. Time proceeds in the positive direction on the conjugate trajectory, with the initial point being the conjugate of the original final point, $\tilde{\mathbf{\Gamma}}_0 = \mathbf{\Gamma}_f^\dagger$.

The static entropy is the instantaneous equilibrium entropy, and as such it has even parity, $S_{r,st}(\mathbf{\Gamma}^\dagger, t) = S_{r,st}(\mathbf{\Gamma}, t)$. Hence $\tilde{S}_{r,st}(\tilde{\mathbf{\Gamma}}(t), t) = S_{r,st}(\mathbf{\Gamma}(t_0 + t_f - t), t_0 + t_f - t)$. In the conjugate system this is a function of the conjugate Hamiltonian, $\tilde{\mathcal{H}}(\mathbf{\Gamma}, t)$. The adiabatic derivatives are equal and opposite in the two systems,

$$\dot{\tilde{S}}_{r,st}^0(\tilde{\mathbf{\Gamma}}(t), t) = -\dot{S}_{r,st}^0(\mathbf{\Gamma}(t_0 + t_f - t), t_0 + t_f - t).$$

<div align="right">(6.1 90)</div>

This result only holds for the conjugate Hamiltonian as defined here. Of course there are some systems where the conjugate Hamiltonian is the same as the original, such as equilibrium and thermodynamic steady-state systems, where the Hamiltonian does not depend on time. Also, harmonic systems with time dependence of the form $\cos(2\pi l t/(t_0 + t_f))$, with l an integer.

The conditional trajectory probability in the conjugate system is

$$\wp(\tilde{\underline{\boldsymbol{\Gamma}}} \mid \tilde{\boldsymbol{\Gamma}}_0; \tilde{\mathcal{H}}) = \mathrm{e}^{[\tilde{S}_{\mathrm{r,st}}(\tilde{\boldsymbol{\Gamma}}_f, \tilde{t}_f) - \tilde{S}_{\mathrm{r,st}}(\tilde{\boldsymbol{\Gamma}}_0, \tilde{t}_0)]/2k_{\mathrm{B}}} \mathrm{e}^{\tilde{S}_{\mathrm{r,dyn}}(\tilde{\underline{\boldsymbol{\Gamma}}})/2k_{\mathrm{B}}}$$

$$\times \prod_{n=0}^{f-1} \frac{\Theta_\Lambda(\mathbf{R}_n^{\mathrm{conj}}) \Theta_\Lambda(\bar{\mathbf{R}}_n^{\mathrm{conj}})}{[\mathrm{Det}\ 2\pi \mid \Delta_t \mid k_{\mathrm{B}}\Lambda]^{1/2}}$$

$$= \mathrm{e}^{[S_{\mathrm{r,st}}(\boldsymbol{\Gamma}_0, t_0) - S_{\mathrm{r,st}}(\boldsymbol{\Gamma}_f, t_f)]/2k_{\mathrm{B}}} \mathrm{e}^{-S_{\mathrm{r,dyn}}(\underline{\boldsymbol{\Gamma}})/2k_{\mathrm{B}}}$$

$$\times \prod_{n=0}^{f-1} \frac{\Theta_\Lambda(\mathbf{R}_n) \Theta_\Lambda(\bar{\mathbf{R}}_n)}{[\mathrm{Det}\ 2\pi \mid \Delta_t \mid k_{\mathrm{B}}\Lambda]^{1/2}}.$$

(6.1
The second equality follows from three facts. First, the dynamic trajectory entropy is 91)
negated on the conjugate trajectory,

$$\tilde{S}_{\mathrm{r,dyn}}(\tilde{\underline{\boldsymbol{\Gamma}}}) = -\int_{t_0}^{t_f} \mathrm{d}t\ \dot{\tilde{S}}_{\mathrm{r,st}}^0(\tilde{\boldsymbol{\Gamma}}(t), t)$$

$$= \int_{t_0}^{t_f} \mathrm{d}t\ \dot{S}_{\mathrm{r,st}}^0(\boldsymbol{\Gamma}(t_0 + t_f - t), t_0 + t_f - t)$$

$$= \int_{t_0}^{t_f} \mathrm{d}t\prime\ \dot{S}_{\mathrm{r,st}}^0(\boldsymbol{\Gamma}(t\prime), t\prime)$$

$$= -S_{\mathrm{r,dyn}}(\underline{\boldsymbol{\Gamma}}).$$

(6.1
Second, the reservoir force on the conjugate trajectory is the negative conjugate of 92)
that on the original trajectory,

$$\mathbf{R}_n^{\mathrm{conj}} = \tilde{\boldsymbol{\Gamma}}_{n+1} - \tilde{\boldsymbol{\Gamma}}_n - \Delta_t \dot{\tilde{\boldsymbol{\Gamma}}}_n^0$$

$$= \boldsymbol{\Gamma}_{f-n-1}^\dagger - \boldsymbol{\Gamma}_{f-n}^\dagger - \Delta_t \dot{\boldsymbol{\Gamma}}^0(\boldsymbol{\Gamma}_{f-n-1}^\dagger, t_{f-n-1})$$

$$= \boldsymbol{\Gamma}_{f-n-1}^\dagger - \boldsymbol{\Gamma}_{f-n}^\dagger + \Delta_t \dot{\boldsymbol{\Gamma}}^0(\boldsymbol{\Gamma}_{f-n-1}, t_{f-n-1})^\dagger$$

$$= -\mathbf{R}_{f-n-1}^\dagger.$$

(6.1
Hence the Gaussian exponent is unchanged, 93)
$\Lambda_{pp} : [-\mathbf{R}_{f-n-1}^\dagger][-\mathbf{R}_{f-n-1}^\dagger] = \Lambda_{pp} : \mathbf{R}_{f-n-1}\mathbf{R}_{f-n-1}$. The sum of the Gaussian exponents in both cases adds together all the reservoir forces. Third, the most likely reservoir also simply changes sign,

$$\bar{\mathbf{R}}_n^{\mathrm{conj}} = \frac{\mid \Delta_t \mid}{2} \Lambda_{pp} \tilde{\nabla}_p \tilde{S}_{\mathrm{r,st}}(\tilde{\boldsymbol{\Gamma}}(t_n), t_n)$$

$$= \frac{-\mid \Delta_t \mid}{2} \Lambda_{pp} \nabla_p S_{\mathrm{r,st}}(\boldsymbol{\Gamma}_{f-n}, t_{f-n}).$$

(6.1
Recall that the time step is positive on both trajectories, and so the most likely force is 94)
proportional to the gradient of the static part of the reservoir entropy alone. And so again the Gaussian factor of these is unchanged.

The ratio of the conditional trajectory probabilities for the two systems is

$$\frac{\wp(\underline{\boldsymbol{\Gamma}} \mid \boldsymbol{\Gamma}_0)}{\wp(\underline{\tilde{\boldsymbol{\Gamma}}} \mid \tilde{\boldsymbol{\Gamma}}_0; \tilde{\mathcal{H}})} = e^{[S_{\mathrm{r,st}}(\boldsymbol{\Gamma}_f, t_f) - S_{\mathrm{r,st}}(\boldsymbol{\Gamma}_0, t_0)]/k_\mathrm{B}} e^{S_{\mathrm{r,dyn}}(\underline{\boldsymbol{\Gamma}})/k_\mathrm{B}}$$

$$= e^{\vec{\nabla} S_\mathrm{r}(\underline{\boldsymbol{\Gamma}})/k_\mathrm{B}}$$

$$\approx e^{S_{\mathrm{r,dyn}}(\underline{\boldsymbol{\Gamma}})/k_\mathrm{B}}.$$

(6.195)

The time interval is implicit in the trajectory, and for simplicity it is not shown explicitly as a conditioning argument. The exponent on the right-hand side of the second equality is the actual change in reservoir entropy over the original trajectory. This obviously scales with the length of the trajectory, and so for a long enough trajectory the initial conditions can be neglected. This means that this result also holds approximately for the ratio of unconditional probabilities.

Alternatively, for the unconditional probability, one assumption that one can make is that the system is initially and finally in the non-equilibrium state given by the usual expression for the reservoir entropy. In this case the ratio of the unconditional probabilities is

$$\frac{\wp(\underline{\boldsymbol{\Gamma}} \mid \boldsymbol{\Gamma}_0)\wp(\boldsymbol{\Gamma}_0, t_0)}{\wp(\underline{\tilde{\boldsymbol{\Gamma}}} \mid \tilde{\boldsymbol{\Gamma}}_0; \tilde{\mathcal{H}})\wp(\tilde{\boldsymbol{\Gamma}}_0, t_0; \tilde{\mathcal{H}})} = \frac{Z(t_0; \tilde{\mathcal{H}})}{Z(t_0)} \frac{e^{S_{\mathrm{r,dyn}}(\boldsymbol{\Gamma}_0, t_0)/k_\mathrm{B}}}{e^{S_{\mathrm{r,dyn}}(\tilde{\boldsymbol{\Gamma}}_0, t_0; \tilde{\mathcal{H}})/k_\mathrm{B}}} e^{S_{\mathrm{r,dyn}}(\underline{\boldsymbol{\Gamma}})/k_\mathrm{B}}$$

$$\approx e^{S_{\mathrm{r,dyn}}(\underline{\boldsymbol{\Gamma}})/k_\mathrm{B}}.$$

(6.196)

The final approximation assumes that the total unconstrained entropy does not increase monotonically over time.

A different set of initial conditions is for the system to be in the static state initially and finally. Now the ratio of the unconditional probabilities is

$$\frac{\wp(\underline{\boldsymbol{\Gamma}} \mid \boldsymbol{\Gamma}_0)\wp_{\mathrm{st}}(\boldsymbol{\Gamma}_0, t_0)}{\wp(\underline{\tilde{\boldsymbol{\Gamma}}} \mid \tilde{\boldsymbol{\Gamma}}_0; \tilde{\mathcal{H}})\wp_{\mathrm{st}}(\tilde{\boldsymbol{\Gamma}}_0, t_0; \tilde{\mathcal{H}})} = \frac{Z_{\mathrm{st}}(t_f)}{Z_{\mathrm{st}}(t_0)} e^{S_{\mathrm{r,dyn}}(\underline{\boldsymbol{\Gamma}})/k_\mathrm{B}}$$

$$\approx e^{S_{\mathrm{r,dyn}}(\underline{\boldsymbol{\Gamma}})/k_\mathrm{B}}.$$

(6.197)

In general, the logarithm of the partition function is the total unconstrained entropy, which when multiplied by the negative of the temperature gives a free energy. Hence the first factor in the first equality is essentially the exponential of the difference in the static total entropy or static free energy. For a cyclic trajectory, $\mathcal{H}(\boldsymbol{\Gamma}, t_f) = \mathcal{H}(\boldsymbol{\Gamma}, t_0)$, the final approximation is exact as $Z_{\mathrm{st}}(t_f) = Z_{\mathrm{st}}(t_0)$.

In any case these examples confirm the general conclusion that for a large enough time interval one can ignore the state of the system beyond it. The ratio of probabilities, conditional or unconditional, of the original and conjugate trajectories is the exponential of the change in the reservoir entropy over the interval.

6.5.2 Fluctuation and work theorems

Although the Second Law of Thermodynamics asserts that the entropy increases with time, from the discussion in preceding sections of this book, specifically sections 1.2.4, 1.3, and 3.1.5, it is clear that this is meant in a probabilistic sense only. Hence the question arises: how much more probable is the entropy to increase than to decrease?

The expressions given above for the trajectory probability give a quantitative answer to this question. The probability of observing the reservoir entropy change by $\vec{\Delta} S_\mathrm{r}$ over the

given time interval can be related to the probability of observing the opposite change by elementary manipulations of the above results. For a specified time interval one has

$$\wp(\vec{\Delta} S_r \mid t_{21}) = \int d\underline{\Gamma} \ \delta(\vec{\Delta} S_r - \vec{\Delta} S_r(\underline{\Gamma}))\wp(\underline{\Gamma})$$

$$= \int d\underline{\tilde{\Gamma}} \ \delta(\vec{\Delta} S_r - \vec{\Delta} S_r(\underline{\Gamma}))\wp(\underline{\tilde{\Gamma}} \mid \tilde{\mathcal{H}})e^{\vec{\Delta} S_r(\underline{\Gamma})/k_B}$$

$$= e^{\vec{\Delta} S_r/k_B} \int d\underline{\tilde{\Gamma}} \ \delta(\vec{\Delta} S_r + \vec{\Delta} S_r(\underline{\tilde{\Gamma}}))\wp(\underline{\tilde{\Gamma}} \mid \tilde{\mathcal{H}})$$

$$= e^{\vec{\Delta} S_r/k_B}\wp(-\vec{\Delta} S_r \mid t_{21}, \tilde{\mathcal{H}}).$$

(6.198)

Here $\vec{\Delta} S_r$ is just a number, the specified change in entropy over the given time interval. All possible trajectories are integrated over in obtaining this result. Again, the time interval is implicit in the trajectory, and for simplicity it is not shown explicitly in the trajectory probability.

This result relies upon the conjugate system as defined above. As has been mentioned, in some cases the conjugate system is the same as the original system. Examples include equilibrium systems, steady-state thermodynamic systems, and mechanical non-equilibrium systems with harmonic Hamiltonian.

This result is exact for the conditional trajectory probability, which is to say that the initial and final points of the trajectory are specified. In any case for a sufficiently long trajectory, end effects can be neglected and it also holds approximately for the unconditional probability.

The result says that the probability of a particular entropy increase is exponentially more likely than the same decrease. This shows why the Second Law of Thermodynamics can be considered exact for a macroscopic system. In essence this result is the fluctuation theorem that was first derived by Bochkov and Kuzovlev (1981). Different versions and applications of the theorem have subsequently been presented by others (Evans *et al* 1993, Jarzynski 1997, Evans 2003, Attard 2006b). A feature of the present derivation is that it includes the stochastic, dissipative forces from the reservoir (i.e. it is not restricted to adiabatic trajectories), and it applies also to non-equilibrium thermodynamic systems (i.e. it is not restricted to mechanical non-equilibrium systems).

The work theorem is closely related to the fluctuation theorem. This is derived from the average over trajectories of the exponential of the negative of the entropy change on each trajectory. For a specified time interval one has

$$\left\langle e^{-\vec{\Delta} S_r(\underline{\Gamma})/k_B} \right\rangle = \int d\underline{\Gamma} \ \wp(\underline{\Gamma})e^{-\vec{\Delta} S_r(\underline{\Gamma})/k_B}$$

$$= \int d\underline{\tilde{\Gamma}} \ \wp(\underline{\tilde{\Gamma}} \mid \tilde{\mathcal{H}})$$

$$= 1.$$

(6.199)

On the left-hand side appears the negative of the change in reservoir entropy over the trajectory. End effects have again been neglected, which is valid for long enough trajectories. What is interesting about this result is that the exponent on the left-hand side scales linearly with the length of the time interval, but the average itself is independent of the time interval.

As an illustration of this, consider the average of the exponential of the negative of the dynamic part of the reservoir entropy. This is just the thermodynamic work. Assuming static

equilibrium initially and finally, this is

$$
\begin{aligned}
\left\langle e^{-S_{\text{r,dyn}}(\boldsymbol{\Gamma})/k_{\text{B}}} \right\rangle &= \int d\underline{\boldsymbol{\Gamma}} \; \wp(\underline{\boldsymbol{\Gamma}} \mid \boldsymbol{\Gamma}_0) \wp_{\text{st}}(\boldsymbol{\Gamma}_0, t_0) e^{-S_{\text{r,dyn}}(\boldsymbol{\Gamma})/k_{\text{B}}} \\
&= \int d\underline{\tilde{\boldsymbol{\Gamma}}} \; \wp(\underline{\tilde{\boldsymbol{\Gamma}}} \mid \tilde{\boldsymbol{\Gamma}}_0; \tilde{\mathcal{H}}) \wp_{\text{st}}(\tilde{\boldsymbol{\Gamma}}_0, t_0; \tilde{\mathcal{H}}) \frac{\wp_{\text{st}}(\boldsymbol{\Gamma}_0, t_0)}{\wp_{\text{st}}(\tilde{\boldsymbol{\Gamma}}_0, t_0; \tilde{\mathcal{H}})} \\
&\quad \times e^{-S_{\text{r,st}}(\boldsymbol{\Gamma}_0, t_0)/k_{\text{B}}} e^{S_{\text{r,st}}(\boldsymbol{\Gamma}_f, t_f)/k_{\text{B}}} \\
&= \frac{Z_{\text{st}}(t_f)}{Z_{\text{st}}(t_0)} \\
&= e^{-\beta \Delta F_{\text{st}}}.
\end{aligned}
$$

(6.200)

This uses the fact that $Z_{\text{st}}(t_0; \tilde{\mathcal{H}}) = Z_{\text{st}}(t_f)$, and also the fact that the free energy is $-k_{\text{B}}T$ times the logarithm of the partition function. The dynamic trajectory entropy, equation (6.182), is in essence the work done on the sub-system (at least it is for a mechanical non-equilibrium system; the present definition also applies to a thermodynamic non-equilibrium system), and so one sees that the left-hand side is the exponential of the negative of the work done. On the right-hand side, the exponent is the negative of the difference in the static free energy divided by $k_{\text{B}}T$.

The work theorem was first given by Bochkov and Kuzovlev (1981), who assumed a long cyclic trajectory. A similar result was also given by Jarzynski (1997), who also treated a mechanical non-equilibrium system and assumed a Boltzmann distribution at the beginning and end of the trajectory. The present result is valid both for applied thermodynamic gradients and for external time-dependent mechanical fields.

6.6 Path entropy for mechanical work

6.6.1 Stochastic dissipative equations of motion

We now analyze a mechanical non-equilibrium system, which is a sub-system with a time-varying Hamiltonian, $\mathcal{H}(\boldsymbol{\Gamma}, t) = \mathcal{K}(\mathbf{p}) + U(\mathbf{q}, t)$. Energy can be exchanged with a thermal reservoir of temperature T. The kinetic energy may be taken to be $\mathcal{K}(\mathbf{p}) = \sum_{j\alpha} p_{j\alpha}^2/2m$, where $j = 1, 2 \ldots, N$ labels the atoms and $\alpha = x, y, z$ labels the axes. The time dependence of the potential is generally contained in an externally applied potential, and the initial aim is to calculate the mechanical work that this does on the sub-system over time.

The static part of the reservoir entropy is the usual equilibrium expression, $S_{\text{r,st}}(\boldsymbol{\Gamma}, t) = -\mathcal{H}(\boldsymbol{\Gamma}, t)/T$. Hence the change in entropy for a transition, $\boldsymbol{\Gamma}_2 \xrightarrow{\Delta_t} \boldsymbol{\Gamma}_1$, $\Delta_t \equiv t_2 - t_1$, is explicitly

$$
\begin{aligned}
\vec{\Delta} S_{\text{r}}(\boldsymbol{\Gamma}_2, t_2; \boldsymbol{\Gamma}_1, t_1) &= \mathbf{R} \cdot \nabla_{\text{p}} S_{\text{r,st}}(\boldsymbol{\Gamma}, t) \\
&= \frac{-1}{mT} \sum_{j\alpha} R_{j\alpha} \, p_{j\alpha}.
\end{aligned}
$$

(6.201)

The total reservoir force is the force required beyond the adiabatic force to give the designated transition, $\mathbf{R} = \boldsymbol{\Gamma}_2 - \boldsymbol{\Gamma}_1 - \Delta_t \dot{\boldsymbol{\Gamma}}^0$. As usual, it comprises a dissipative part and

a stochastic part, $\mathbf{R} = \bar{\mathbf{R}} + \tilde{\mathbf{R}}$.

We shall take the second entropy fluctuation matrix to be proportional to the identity matrix in momentum space, $\Lambda = \lambda \mathbf{I}_{pp}$. In this case the variance of the stochastic part of the force is

$$\left\langle \tilde{\mathbf{R}}(t) \, \tilde{\mathbf{R}}(t) \right\rangle = k_{\mathrm{B}} \mid \Delta_t \mid \lambda \, \mathbf{I}_{pp},$$

(6.202)

and the most likely reservoir force is

$$\bar{\mathbf{R}}(\Delta_t, \mathbf{\Gamma}, t) = \frac{\mid \Delta_t \mid \lambda}{2} \left[\nabla_{\mathrm{p}} S_{\mathrm{r,st}}(\mathbf{\Gamma}, t) + [\hat{\Delta}_t - 1] \nabla_p S_{\mathrm{r,st}}(\bar{\mathbf{\Gamma}}(t), t) \right]$$

$$= \frac{- \mid \Delta_t \mid \lambda}{2mT} \left[\mathbf{\Gamma}_{\mathrm{p}} + [\hat{\Delta}_t - 1] \bar{\mathbf{\Gamma}}_{\mathrm{p}}(t) \right].$$

(6.203)

Here and throughout, $\hat{\Delta}_t \equiv \mathrm{sign} \, \Delta_t$.

The scalar λ represents the influence of the reservoir on the sub-system. In practical computations, the magnitude of this is to a large extent arbitrary. (One does have to consider the twin effects of the length of the time step and the variance of the stochastic force in solving the equations of motion by simple time stepping. These are not independent, and the statistical averages must be insensitive to the chosen values.) Of course the same constant λ must be used for the variance of the stochastic force as for the magnitude of the dissipative force, which is just the fluctuation–dissipation theorem. For a positive time step, $\Delta_t > 0$, which is usually the case in the results analyzed below, the most likely momenta are not required because $\hat{\Delta}_t - 1 = 0$.

The stochastic, dissipative equations of motion over a single time step for the present mechanical non-equilibrium mechanical system are explicitly

$$q_{j\alpha}(t + \Delta_t) \;=\; q_{j\alpha}(t) + \Delta_t \dot{q}^0_{j\alpha}(t),$$

$$p_{j\alpha}(t + \Delta_t) \;=\; p_{j\alpha}(t) + \Delta_t \dot{p}^0_{j\alpha}(t) - \frac{\mid \Delta_t \mid \lambda}{2 \, mT} \left[p_{j\alpha} + [\hat{\Delta}_t - 1] \bar{p}_{j\alpha}(t) \right] + \tilde{R}_{j\alpha}.$$

(6.204)

As usual, the adiabatic rates of change are

$$\dot{q}^0_{j\alpha}(t) = \frac{\partial \mathcal{H}(\mathbf{\Gamma}, t)}{\partial p_{j\alpha}} = \frac{p_{j\alpha}}{m} \text{ and } \dot{p}^0_{j\alpha}(t) = \frac{-\partial \mathcal{H}(\mathbf{\Gamma}, t)}{\partial q_{j\alpha}}.$$

(6.205)

In these equations of motion, the stochastic forces at each time step are uncorrelated. The dissipative force, $-(\mid \Delta_t \mid \lambda/2 \, mT) p_{j\alpha}$, has the form of a drag force. The final term that appears here, $-(\mid \Delta_t \mid \lambda/2 \, mT)[\hat{\Delta}_t - 1] \bar{p}_{j\alpha}(t)$, is only nonzero for a backward time step. As mentioned, that equality of the drag coefficient and the variance of the stochastic force is known as the fluctuation–dissipation theorem. These stochastic, dissipative equations of motion have the same form as the Langevin equation for Brownian motion, section 4.1.2, but of course the justification for the present molecular application is statistical rather than hydrodynamic.

The conditional second entropy for the transition is

$$S^{(2)}(\mathbf{\Gamma}_2, t_2 \mid \mathbf{\Gamma}_1, t_1) = \frac{-1}{2 \mid \Delta_t \mid \lambda} \tilde{\mathbf{R}} \cdot \tilde{\mathbf{R}}$$

$$= \frac{-1}{2 \mid \Delta_t \mid \lambda} \left[\mathbf{R} \cdot \mathbf{R} + \bar{\mathbf{R}} \cdot \bar{\mathbf{R}} \right] + \frac{1}{2} \vec{\Delta} S_{\mathrm{r}}(\mathbf{\Gamma}_2, t_2; \mathbf{\Gamma}_1, t_1)$$

$$= \frac{-1}{2 \mid \Delta_t \mid \lambda} \sum_{j\alpha} \left[R_{j\alpha}^2 + \frac{\mid \Delta_t \mid^2 \lambda^2}{4 m^2 T^2} p_{j\alpha}^2 \right] - \frac{1}{2mT} \sum_{j\alpha} R_{j\alpha} \, p_{j\alpha}, \quad \Delta$$

(6.2

With this the conditional transition probability is $\qquad$ 06)

$$\wp(\mathbf{\Gamma}_2, t_2 \mid \mathbf{\Gamma}_1, t_1) = \frac{1}{Z(t_2, t_1)} \mathrm{e}^{S^{(2)}(\mathbf{\Gamma}_2, t_2 \mid \mathbf{\Gamma}_1, t_1)/k_{\mathrm{B}}}$$

$$= \frac{1}{Z(t_2, t_1)} \mathrm{e}^{-\tilde{\mathbf{R}} \cdot \tilde{\mathbf{R}}/2 k_{\mathrm{B}} \mid \Delta_t \mid \lambda} \delta\left(\mathbf{\Gamma}_{\mathrm{q}2} - \mathbf{\Gamma}_{\mathrm{q}1} - \Delta_t \dot{\mathbf{\Gamma}}_{\mathrm{q}}^0 \right)$$

$$= \frac{1}{Z(t_2, t_1)} \delta\left(\mathbf{\Gamma}_{\mathrm{q}2} - \mathbf{\Gamma}_{\mathrm{q}1} - \Delta_t \dot{\mathbf{\Gamma}}_{\mathrm{q}}^0 \right)$$

$$\times \prod_{j\alpha} \mathrm{e}^{-\left[R_{j\alpha}^2 + \mid \Delta_t \mid^2 \lambda^2 p_{j\alpha}^2 / 4 m^2 T^2 \right]/2 k_{\mathrm{B}} \mid \Delta_t \mid \lambda} \prod_{j\alpha} \mathrm{e}^{-\beta R_{j\alpha} p_{j\alpha}/2\, m}.$$

(6.2

The normalization factor is $Z(t_2, t_1) = (2\pi k_{\mathrm{B}} \mid \Delta_t \mid \lambda)^{3N/2}$, and $\beta \equiv 1/k_{\mathrm{B}}T$. $\qquad$ 07)

The final exponent, $-\beta \sum_{j\alpha} R_{j\alpha} \, p_{j\alpha}/2m = -\beta \mathbf{R} \cdot \nabla_{\mathrm{p}} \mathcal{H}(\mathbf{\Gamma}, t)/2$, is essentially half the change in entropy for the transition. It is interesting to note that the present second entropy analysis yields the same factor of $\beta/2$ as in Glauber and Kawasaki dynamics, which are stochastic approaches to dynamical systems that neglect adiabatic evolution (Glauber 1963, Kawasaki 1966, Langer 1969, Metiu *et al* 1975). The present result for the conditional stochastic transition probability satisfies detailed balance for an equilibrium Boltzmann distribution. It has been applied with a stochastic molecular dynamics algorithm for equilibrium systems (Attard 2002, Boinepalli and Attard 2003), It has also been used as the basis for a non-equilibrium molecular dynamics algorithm for steady heat flow and for driven Brownian motion (Attard 2006b, 2009a).

6.6.2 Work done and free energy difference

As in section 6.5.1, denote a trajectory by the vector $\underline{\mathbf{\Gamma}}$, with elements $\mathbf{\Gamma}_n = \mathbf{\Gamma}(t_n)$, and nodes $t_n = t_0 + n\Delta_t$, $n = 0, 1, \ldots, f$, the time step being $\Delta_t = [t_f - t_0]/f > 0$. The change in reservoir entropy for a mechanical non-equilibrium system is

$$\vec{\Delta} S_{\mathrm{r}}(\underline{\mathbf{\Gamma}}) = S_{\mathrm{r,st}}(\mathbf{\Gamma}_f, t_f) - S_{\mathrm{r,st}}(\mathbf{\Gamma}_0, t_0) - \int_{t_0}^{t_f} \mathrm{d}t\prime \, \dot{S}_{\mathrm{r,st}}^0(\mathbf{\Gamma}(t\prime), t\prime)$$

$$= \frac{1}{T} \left[\mathcal{H}(\mathbf{\Gamma}_0, t_0) - \mathcal{H}(\mathbf{\Gamma}_f, t_f) \right] + \frac{1}{T} \int_{t_0}^{t_f} \mathrm{d}t\prime \, \frac{\partial U(\mathbf{\Gamma}\prime, t\prime)}{\partial t\prime},$$

(6.2

where $\mathbf{\Gamma}\prime \equiv \mathbf{\Gamma}(t\prime)$. The dynamic part of the reservoir entropy, the final term, is the work 08) done over the actual trajectory divided by temperature, $S_{\mathrm{r,dyn}}(\underline{\mathbf{\Gamma}}) = W(\underline{\mathbf{\Gamma}})/T$, with the work done being

$$W(\underline{\boldsymbol{\Gamma}}) = \int_{t_0}^{t_f} \mathrm{d}t\prime \, \frac{\partial U(\boldsymbol{\Gamma}\prime, t\prime)}{\partial t\prime}$$

$$= \int_{t_0}^{t_f} \mathrm{d}t\prime \, \dot{\mathcal{H}}^0(\boldsymbol{\Gamma}(t\prime), t\prime)$$

$$= \Delta_t \sum_{n=0}^{f-1} \dot{\mathcal{H}}^0(\boldsymbol{\Gamma}_n, t_n).$$

(6.2 09)

The earlier figure 6.9 serves also to illustrate this. The adiabatic rate of change of the energy is just the partial time derivative of the potential, $\dot{\mathcal{H}}^0(\boldsymbol{\Gamma}, t) = \partial U(\boldsymbol{\Gamma}, t)/\partial t$. Other recipes for discretizing the integral give the same result to leading order in the time step. As has been mentioned in section 6.5.1, the work term dominates the change in entropy for a sufficiently long trajectory and the contributions of the initial and final states may be neglected. The exception to this is if the time-dependent part of the potential grossly changes in proportion to the length of the time interval.

The conditional trajectory entropy is

$$S(\underline{\boldsymbol{\Gamma}} \mid \boldsymbol{\Gamma}_0, t_0) = \frac{1}{2}\vec{\Delta}S_{\mathrm{r}}(\underline{\boldsymbol{\Gamma}}) - \frac{1}{2 \mid \Delta_t \mid \lambda} \sum_{n=0}^{f-1} \sum_{j\alpha} \left[R_{n;j\alpha}^2 + \frac{\mid \Delta_t \mid^2 \lambda^2}{4m^2T^2} p_{n;j\alpha}^2 \right]$$

$$= \frac{1}{2T}[\mathcal{H}(\boldsymbol{\Gamma}_0, t_0) - \mathcal{H}(\boldsymbol{\Gamma}_f, t_f)] + \frac{W(\underline{\boldsymbol{\Gamma}})}{2T}$$

$$- \frac{1}{2 \mid \Delta_t \mid \lambda} \sum_{n=0}^{f-1} \sum_{j\alpha} \left[R_{n;j\alpha}^2 + \frac{\mid \Delta_t \mid^2 \lambda^2}{4m^2T^2} p_{n;j\alpha}^2 \right].$$

(6.2 10)

The total reservoir force at the nth node is $\mathbf{R}_n \equiv \boldsymbol{\Gamma}_{\mathrm{p},n+1} - \boldsymbol{\Gamma}_{\mathrm{p},n} - \Delta_t \dot{\boldsymbol{\Gamma}}_{\mathrm{p},n}^0$, with the adiabatic velocity being evaluated anywhere on the nth interval.

The conditional trajectory probability is the exponential of this divided by Boltzmann's constant,

$$\wp(\underline{\boldsymbol{\Gamma}} \mid \boldsymbol{\Gamma}_0, t_0) = \prod_{n=0}^{f-1} \frac{\Theta_\lambda(\tilde{\mathbf{R}}_n)}{(2\pi k_{\mathrm{B}} \mid \Delta_t \mid \lambda)^{3N/2}}$$

$$= \mathrm{e}^{\vec{\Delta}S_{\mathrm{r}}(\underline{\boldsymbol{\Gamma}})/2k_{\mathrm{B}}} \prod_{n=0}^{f-1} \frac{\Theta_\lambda(\mathbf{R}_n)\Theta_\lambda(\bar{\mathbf{R}}_n)}{(2\pi k_{\mathrm{B}} \mid \Delta_t \mid \lambda)^{3N/2}}$$

$$= \mathrm{e}^{-\beta[\mathcal{H}(\boldsymbol{\Gamma}_f, t_f) - \mathcal{H}(\boldsymbol{\Gamma}_0, t_0)]/2} \mathrm{e}^{\beta W(\underline{\boldsymbol{\Gamma}})/2} \prod_{n=0}^{f-1} \frac{\Theta_\lambda(\mathbf{R}_n)\Theta_\lambda(\bar{\mathbf{R}}_n)}{(2\pi k_{\mathrm{B}} \mid \Delta_t \mid \lambda)^{3N/2}},$$

(6.2 11)

where the un-normalized Gaussian probability distribution is

$$\Theta_\lambda(\mathbf{R}_n) \equiv \mathrm{e}^{-\mathbf{R}_n \cdot \mathbf{R}_n / 2k_{\mathrm{B}} \mid \Delta_t \mid \lambda}.$$

(6.2 12)

As in section 6.5.1, define a conjugate system that on the interval $t \in [t_0, t_f]$ has Hamiltonian $\tilde{\mathcal{H}}(\mathbf{\Gamma}, t) = \mathcal{H}(\mathbf{\Gamma}, t_f + t_0 - t)$. The conjugate trajectory satisfies $\tilde{\mathbf{\Gamma}}(t) = \mathbf{\Gamma}(t_f + t_0 - t)^{\dagger}$, where the dagger denotes velocity reversal. The respective conditional probabilities are in the ratio

$$\frac{\wp(\underline{\mathbf{\Gamma}} \mid \mathbf{\Gamma}_0)}{\wp(\underline{\tilde{\mathbf{\Gamma}}} \mid \tilde{\mathbf{\Gamma}}_0; \tilde{\mathcal{H}})} = e^{\vec{\Delta} S_r(\underline{\mathbf{\Gamma}})/k_B} = e^{-\beta[\mathcal{H}(\mathbf{\Gamma}_f, t_f) - \mathcal{H}(\mathbf{\Gamma}_0, t_0)]} e^{\beta W(\underline{\mathbf{\Gamma}})}.$$

(6.2 13)

As in the general case, the Gaussians cancel in this ratio.

Assuming that the system was in the non-equilibrium state at the start of the original trajectory, $\wp(\mathbf{\Gamma}_0, t_0) = Z(t_0)^{-1} e^{S_r(\mathbf{\Gamma}_0, t_0)/k_B}$, and also of the conjugate trajectory, $\wp(\tilde{\mathbf{\Gamma}}_0, t_0; \tilde{\mathcal{H}}) = Z(\tilde{t}_0)^{-1} e^{S_r(\tilde{\mathbf{\Gamma}}_0, t_0; \tilde{\mathcal{H}})/k_B}$, with $S_{r,st}(\tilde{\mathbf{\Gamma}}_0, t_0; \tilde{\mathcal{H}}) = S_{r,st}(\mathbf{\Gamma}_f, t_f)$, the ratio of the unconditional trajectory probabilities is

$$\frac{\wp(\underline{\mathbf{\Gamma}} \mid \mathbf{\Gamma}_0)\wp(\mathbf{\Gamma}_0, t_0)}{\wp(\underline{\tilde{\mathbf{\Gamma}}} \mid \tilde{\mathbf{\Gamma}}_0; \tilde{\mathcal{H}})\wp(\tilde{\mathbf{\Gamma}}_0, t_0; \tilde{\mathcal{H}})} = \frac{Z(t_0; \tilde{\mathcal{H}})}{Z(t_0)} \frac{e^{\beta \bar{W}(\mathbf{\Gamma}_0, t_0)}}{e^{\beta \bar{W}(\tilde{\mathbf{\Gamma}}_0, t_0; \tilde{\mathcal{H}})}} e^{\beta W(\underline{\mathbf{\Gamma}})}.$$

(6.2 14)

This follows because the dynamic part of the reservoir entropy is essentially the most likely work done,

$$\bar{W}(\mathbf{\Gamma}, t) \equiv T S_{r,dyn}(\mathbf{\Gamma}, t)$$
$$= -T \int_0^t dt\prime \, \dot{S}_r^0(\bar{\mathbf{\Gamma}}(t\prime \mid \mathbf{\Gamma}, t), t\prime)$$
$$= \int_0^t dt\prime \, \frac{U(\mathbf{\Gamma}\prime, t\prime)}{\partial t\prime}, \quad \mathbf{\Gamma}\prime \equiv \bar{\mathbf{\Gamma}}(t\prime \mid \mathbf{\Gamma}, t).$$

(6.2 15)

These require the definition of the Hamiltonian to be extended beyond the time interval. For a cyclic system, $\mathcal{H}(\mathbf{\Gamma}, t_0) = \mathcal{H}(\mathbf{\Gamma}, t_f)$, with extended Hamiltonian $\tilde{\mathcal{H}}(\mathbf{\Gamma}, t) = \mathcal{H}(\mathbf{\Gamma}, t)$, $t \leqslant t_0$, the dynamic part of the entropies are equal, $S_{r,dyn}(\mathbf{\Gamma}, t; \tilde{\mathcal{H}}) = S_{r,dyn}(\mathbf{\Gamma}, t)$. In this case $Z(t_0; \tilde{\mathcal{H}}) = Z(t_0)$ and

$$\frac{\wp(\underline{\mathbf{\Gamma}} \mid \mathbf{\Gamma}_0)\wp(\mathbf{\Gamma}_0, t_0)}{\wp(\underline{\tilde{\mathbf{\Gamma}}} \mid \tilde{\mathbf{\Gamma}}_0; \tilde{\mathcal{H}})\wp(\tilde{\mathbf{\Gamma}}_0, t_0; \tilde{\mathcal{H}})} = e^{\beta W(\underline{\mathbf{\Gamma}})}.$$

(6.2 16)

This result also holds for a system in which the pre-factors do not change grossly over the time interval, provided that the time interval is long enough.

A similar result holds for a dynamically disordered system prior to and after the trajectory. This corresponds to the instantaneous static distribution which is just the Maxwell–Boltzmann distribution,

$$\wp_{st}(\mathbf{\Gamma}, t) \equiv \frac{1}{Z_{st}(t)} e^{-\beta \mathcal{H}(\mathbf{\Gamma}, t)}, \quad t = t_0 \text{ or } t_f.$$

(6.2 17)

In this case the unconditional probability ratio is given exactly by,

$$\frac{\wp(\underline{\pmb{\Gamma}} \mid \pmb{\Gamma}_0)\wp_{\rm st}(\pmb{\Gamma}_0, t_0)}{\wp(\underline{\tilde{\pmb{\Gamma}}} \mid \tilde{\pmb{\Gamma}}_0; \tilde{\mathcal{H}})\wp_{\rm st}(\tilde{\pmb{\Gamma}}_0, t_0; \tilde{\mathcal{H}})} = \frac{Z_{\rm st}(t_0; \tilde{\mathcal{H}})}{Z_{\rm st}(t_0)}{\rm e}^{\beta W(\pmb{\Gamma})}$$

$$= {\rm e}^{-\beta[F_{\rm st}(t_f) - F_{\rm st}(t_0)]}{\rm e}^{\beta W(\pmb{\Gamma})}.$$

Here the static or instantaneous Helmholtz free energy is the logarithm of the total entropy, which is just the partition function,

$$F_{\rm st}(t) = -k_{\rm B}T \ \ln \ Z_{\rm st}(t) = -k_{\rm B}T \ \ln \ \frac{1}{h^{3N}N!} \int {\rm d}\pmb{\Gamma} \ {\rm e}^{-\beta \mathcal{H}(\pmb{\Gamma}, t)}.$$

For a cyclic system, $\mathcal{H}(\pmb{\Gamma}, t_0) = \mathcal{H}(\pmb{\Gamma}, t_f)$, the static partition functions are equal, and the Helmholtz free energy is the same at the start as at the end of the time interval.

The fluctuation and work theorems were given above for the general non-equilibrium case, equations (6.198), and (6.199). Those results obviously can be applied to the present case of mechanical work. To be explicit, equation (6.200) is the average Boltzmann factor of the work done. Assuming thermal equilibrium at the beginning and at the end of the interval, this is

$$\left\langle {\rm e}^{-\beta W(\pmb{\Gamma})} \right\rangle = \int {\rm d}\underline{\pmb{\Gamma}} \ \wp(\underline{\pmb{\Gamma}} \mid \pmb{\Gamma}_0)\wp_{\rm MB}(\pmb{\Gamma}_0, t_0){\rm e}^{-\beta W(\pmb{\Gamma})}$$

$$= \int {\rm d}\underline{\tilde{\pmb{\Gamma}}} \ \tilde{\wp}(\underline{\tilde{\pmb{\Gamma}}} \mid \tilde{\pmb{\Gamma}}_0)\tilde{\wp}_{\rm MB}(\tilde{\pmb{\Gamma}}_0, t_0){\rm e}^{-\beta[F_{\rm st}(t_f) - F_{\rm st}(t_0)]}$$

$$= {\rm e}^{-\beta[F_{\rm st}(t_f) - F_{\rm st}(t_0)]}.$$

The left-hand side is the average of the exponential of the negative of the mechanical work actually done over the time interval, divided by $k_{\rm B}T$. This average is over all possible trajectories. The exponent on the right-hand side of the final equality is the negative of the difference in the Helmholtz free energy divided by $k_{\rm B}T$. The assumption that the system is in an equilibrium state at the termini of the trajectory is valid if the time dependence of the Hamiltonian is negligible near the termini of the trajectory: $\mathcal{H}(\pmb{\Gamma}, t) \approx \mathcal{H}(\pmb{\Gamma}, t_0)$, for $t \in [t_0, t_0 - \tau]$, $\mathcal{H}(\pmb{\Gamma}, t) \approx \mathcal{H}(\pmb{\Gamma}, t_f)$, for $t \in [t_f, t_f + \tau]$, where τ is some relaxation time.

For example, if

$$\mathcal{H}(\pmb{\Gamma}, t) \rightarrow \begin{cases} \mathcal{H}_1(\pmb{\Gamma}), & t \lesssim t_0 \\ \mathcal{H}_2(\pmb{\Gamma}), & t \gtrsim t_f, \end{cases}$$

then the exponent on the right-hand side is the difference in Helmholtz free energy for the two equilibrium systems with the respective Hamiltonians. Most commonly, $\mathcal{H}(\pmb{\Gamma}, t)$ is chosen to smoothly transform from one to the other. Also common is for the trajectories to be calculated adiabatically, and the average of the work done over them is taken with respect to a Maxwell–Boltzmann distribution of the initial point and initial Hamiltonian. The present formulation allows the average to be more correctly taken over stochastic dissipative trajectories with the correct Gaussian weight, which is the right-hand side of the first equality above.

As has been mentioned, this mechanical work theorem was first given by Bochkov and Kuzovlev (1981) (for the case of a long cyclic trajectory), and was later given by Jarzynski (1997). The advantage of the present derivation is that it explicitly accounts for the

exchange of heat between the sub-system and the reservoir during the performance of the work.

Summary

- The non-equilibrium probability in the sub-system phase space is the exponential of the reservoir entropy, which is the sum of a static or instantaneous part, and a dynamic part. The latter is the time integral over the most likely backward trajectory of the adiabatic rate of static reservoir entropy production. In the mechanical case it may be interpreted as the negative of the most likely work done by the external potential on the sub-system.
- The stochastic, dissipative equations of motion for a non-equilibrium system can be derived from the second entropy for the transition probability. The stochastic reservoir force is Gaussian distributed, and the dissipative force is dominated by an irreversible term proportional to the gradient of the static part of the reservoir entropy. That a single transport matrix gives the variance of the former and the strength of the latter is a non-equilibrium version of the fluctuation–dissipation theorem in phase space.
- The dynamic part of the reservoir entropy can be approximated by its odd projection, which in turn can often be approximated by an integral over adiabatic trajectories.
- The Green–Kubo relations for the transport coefficients can be derived from the average flux by linearizing the non-equilibrium probability with respect to the dynamic part of the reservoir entropy.
- A Brownian particle in a potential trap in uniform motion serves to illustrate results for a mechanical non-equilibrium system simply and explicitly. From the stochastic, dissipative equations of motion, which have Langevin form, the dynamic part of the reservoir entropy may be obtained. Maximizing the full reservoir entropy of the non-equilibrium system shows that the most likely configuration has the particle moving with the same velocity as the trap, and lagging the trap such that the drag force is equal and opposite to the mechanical restoring force.
- Analysis of the path entropy in general gives the probability of a trajectory, from which the average of various function of the change in entropy may be deduced. These give the fluctuation and work theorems for both mechanical and thermodynamic non-equilibrium systems.

References

Attard P 2002 Stochastic molecular dynamics: a combined monte carlo and molecular dynamics technique for isothermal simulations *J. Chem. Phys.* **116** 9616

Attard P 2006a Statistical mechanical theory for steady state systems. IV. Transition probability and simulation algorithm demonstrated for heat flow *J. Chem. Phys.* **124** 024109

Attard P 2006b Statistical mechanical theory for steady state systems. V. Non-equilibrium probability density *J. Chem. Phys.* **124** 224103

Attard P and Gray-Weale A 2008 Statistical mechanical theory for steady state systems. VIII. General theory for a brownian particle driven by a time- and space-varying force *J. Chem. Phys.* **128** 114509

Attard P 2009a Statistical mechanical theory for non-equilibrium systems. IX. Stochastic molecular dynamics *J. Chem. Phys.* **130** 194113

Attard P 2009b Non-equilibrium Monte Carlo simulation for a driven Brownian particle *Phys. Rev.* E **80** 041126

Attard P 2009c The second entropy: a general theory for non-equilibrium thermodynamics and statistical mechanics *Annu. Rep. Prog. Chem.* **105** 63 (section C)

Attard P 2012 *Non-Equilibrium Thermodynamics and Statistical Mechanics: Foundations and Applications* (Oxford: Oxford University Press)

Attard P 2014 Simplified derivation of the non-equilibrium probability distribution

Le Bellac M, Mortessagne F and Batrouni G G 2004 *Equilibrium and Non-Equilibrium Statistical Thermodynamics* (Cambridge: Cambridge University Press)

Bochkov G N and Kuzovlev Y E 1981 Non-linear fluctuation-dissipation relations and stochastic models in non-equilibrium thermodynamics *Physica* **106A** 443

Boinepalli S and Attard P 2003 Grand canonical molecular dynamics *J. Chem. Phys.* **119** 12769

Einstein A 1933 *On the Method of Theoretical Physics* (The Herbert Spencer Lecture, delivered at Oxford, 10 June)

Evans D J, Cohen E G D and Morriss G P 1993 Probability of second law violations in shearing steady states *Phys. Rev. Lett.* **71** 2401

Evans D J 2003 A non-equilibrium free energy theorem for deterministic systems *Mol. Phys.* **101** 1551

Glauber R J 1963 Time-dependent statistics of the Ising model *J. Math. Phys.* **4** 294

Green M S 1954 Markoff random processes and the statistical mechanics of time-dependent phenomena. II. Irreversible processes in fluids *J. Chem. Phys.* **22** 398

Jarzynski C 1997 Non-equilibrium equality for free energy differences *Phys. Rev. Lett.* **78** 2690

Kawasaki K 1966 Diffusion constants near the critical point for time-dependent Ising models. I *Phys. Rev.* **145** 244

Kubo R 1966 The fluctuation-dissipation theorem *Rep. Prog. Phys.* **29** 255

Kubo R, Toda M and Hashitsume N 1978 Statistical physics II *Non-Equilibrium Statistical Mechanics* (Berlin: Springer)

Langer J S 1969 Statistical theory of the decay of metastable states *Ann. Phys.* **54** 258

Metiu H, Kitahara K and Ross J 1975 Derivation of stochastic equations for non-equilibrium Ising mean field model *J. Chem. Phys.* **63** 5116

Onsager L 1931 Reciprocal relations in irreversible processes. I *Phys. Rev.* **37** 405 and 38 2265

Pottier N 2010 *Non-Equilibrium Statistical Physics: Linear Irreversible Processes* (Oxford: Oxford University Press)

Slessor K 1939 *Five Bells: XX Poems* (Sydney: F C Johnson)

Weinberg S 2015 *To Explain the World: The Discovery of Modern Science* (London: Penguin)

Zwanzig R 2001 *Non-Equilibrium Statistical Mechanics* (Oxford: Oxford University Press)

Entropy Beyond the Second Law (Second Edition)

Thermodynamics and statistical mechanics for equilibrium, non-equilibrium, classical, and quantum systems

Phil Attard

Chapter 7

Quantum statistical mechanics as entropy collapse in classical phase space

'Inspiration and aesthetic judgement are important in the development of scientific theories, but the verification of these theories relies finally on impartial experimental tests of their predictions. Though mathematics is used in the formulation of physical theories and in working out their consequences, science is not a branch of mathematics, and scientific theories cannot be deduced by purely mathematical reasoning'

<div align="right">Weinberg (2015 p xiii)</div>

'Let me tell you the secret that has led me to my goal. My only strength lies in my tenacity'

<div align="right">Pasteur (Dubos 1950)</div>

The main aim of this chapter is to formulate quantum statistical mechanics from first principles. Specifically, the von Neumann trace (i.e. mixture of states) for the partition function and statistical averages is derived from the postulates of quantum mechanics. A dual expansion for these that accounts for non-commutativity and particle symmetry is developed, with the leading order term giving classical statistical mechanics, and hence classical mechanics.

Perhaps the key equation in quantum statistical mechanics is the trace over the density operator form of the partition function,

$$
Z = \mathrm{TR}\ \hat{\rho} = \sum_n \mathrm{e}^{-\beta \mathcal{H}_n}.
$$

(7.1)

This is due to von Neumann (1932) and it is said to give the canonical partition function of a quantum system as the trace of the density operator, which in turn is said to equal the Maxwell–Boltzmann weighted sum over all the energy eigenstates. The challenge is to actually derive both of these from first principles, and to give their correct form and interpretation.

7.1 Conventional quantum statistical average

7.1.1 Expectation value

We begin by establishing the basic notation and by setting out the conventional approach to quantum statistical mechanics.

The state of a quantum system may be specified by a particular wave function ψ. A wave function will usually be denoted in the position representation, $\psi(\mathbf{r})$, or else in the bra-ket notation, $|\psi\rangle$ and $\langle\psi|$. The squared norm (or squared modulus) of the wave function is

$$
N(\psi) \equiv \langle \psi \mid \psi \rangle = \int \mathrm{d}\mathbf{r}\ \psi(\mathbf{r})^* \psi(\mathbf{r}),
$$

where the asterisk denotes the complex conjugate. It is usually the case that the wave (7.2) function is normalized, $N(\psi) = 1$.

The expectation value of an operator $\hat{O}$ when the system is in the wave state ψ is (Messiah 1961, Merzbacher 1970)

$$O(\psi) = \frac{1}{N(\psi)} \langle \psi \mid \hat{O} \mid \psi \rangle = \frac{1}{N(\psi)} \int d\mathbf{r} \; \psi(\mathbf{r})^* \hat{O}(\mathbf{r})\psi(\mathbf{r}).$$

(7.3)

The system can also be described as being in a particular quantum state $n = 0, 1, 2, \ldots$, which are the eigenstates of some operator. The corresponding eigenfunctions ζ_n are orthonormal, $\langle \zeta_n \mid \zeta_m \rangle = \delta_{n,m}$, and they form a complete basis for the Hilbert space. Hence the wave function can be written

$$\psi(\mathbf{r}) = \sum_n \psi_n \zeta_n(\mathbf{r}), \quad \psi_n = \langle \zeta_n \mid \psi \rangle.$$

(7.4)

An arbitrary operator can also be represented in this basis, $O_{mn} = \langle \zeta_m \mid \hat{O} \mid \zeta_n \rangle$, and its expectation value then takes the form (Messiah 1961, Merzbacher 1970)

$$O(\psi) = \frac{1}{N(\psi)} \sum_{m,n} \psi_m^* O_{mn} \psi_n.$$

(7.5)

Suppose $\zeta_n^O(\mathbf{r})$ is a normalized eigenfunction of the operator with eigenvalue O_n, $\hat{O}\zeta_n^O = O_n \zeta_n^O$. The eigenstates may possibly be degenerate, but nevertheless the eigenfunctions can be organized to form a complete set, $\psi = \sum_n \psi_n^O \zeta_n^O$. In the eigenfunction basis the operator is diagonal, $O_{mn}^O = O_n \delta_{mn}$, and the expectation value in the wave state ψ is

$$O(\psi) = \frac{1}{N(\psi)} \sum_n \psi_n^{O*} \psi_n^O O_n.$$

(7.6)

The quantity $\psi_n^{O*}\psi_n^O/N(\psi)$ has the interpretation as the probability of the system being in the quantum state n (given that it is in the wave state ψ).

In this form, the eigenvalues O_n are the possible outcomes of the application of the operator $\hat{O}$, and the square of the amplitude $\psi_n^{O*}\psi_n^O$ is proportional to the probability that that particular outcome will occur. (Many believe that a measurement is the application of an operator. I think that this is too simplistic, and that the outcome of a measurement involves statistical effects of the type enunciated below.) In the more general form for the expectation value as the double sum in an arbitrary basis, equation (7.5), the complex numbers $\psi_m^*\psi_n$ cannot be interpreted in terms of classical probability theory because, as discussed in section 1.1.2, the latter is predicated upon real, non-negative weights. The general form exhibits the phenomenon of quantum superposition and interference: the quantum system can be in more than one state at a time, and the different states affect each other.

The wave function collapses upon operator application, $\psi \Rightarrow \zeta_n^O$, which is to say that the system no longer exists in a superposition of states but rather becomes a pure state. The wave function could be said to be decoherent in that there is no interference between states. In general a decoherent system is a mixture of pure states, and its expectation value is formally identical to a classical average. The specifically quantum aspects of the system (interference, superposition) are no longer present.

One may use the bra-ket notation for the wave function to construct the (single wave function) density operator,

$$\hat{\rho} \equiv \frac{1}{N(\psi)} \mid \psi \rangle \langle \psi \mid.$$

(7.7)

With this the expectation value may be written in the form of the trace of the operator product,

$$O(\psi) = \mathrm{TR} \ \hat{\rho} \ \hat{O} = \frac{1}{N(\psi)} \sum_{m,n} \psi_n \psi_m^* O_{mn}.$$

(7.8)

Because the density operator constructed from a single wave function has a dyadic form, it cannot be diagonalized. To see this, consider some representation in which it is diagonal. If $\rho_{n_1 n_1} = \mid \psi_{n_1} \mid^2$ and $\rho_{n_2 n_2} = \mid \psi_{n_2} \mid^2$ are two distinct nonzero diagonal elements, then the corresponding off-diagonal elements, $\rho_{n_1 n_2} = \rho_{n_2 n_1}^* = \psi_{n_1}^* \psi_{n_2}$, would also be nonzero. This contradiction proves that the single wave function density operator cannot be diagonal in any representation. The exception to this is if it has precisely one nonzero element, as only occurs upon collapse due to measurement.

7.1.2 Statistical average

In the context of the statistical average in quantum statistical mechanics, the non-diagonal nature of the single wave function density operator is directly relevant. This is because conventionally the statistical average is expressed as a trace over the density operator (von Neumann 1932). We shall return to this matter shortly, but for the present let us assume the existence of a probability operator, $\hat{\wp}$. Let us further assume that the statistical average can be expressed as the trace of it and the observable operator

$$\left\langle \hat{O} \right\rangle_{\mathrm{stat}} = \mathrm{TR} \ \hat{\wp} \ \hat{O}$$

$$= \sum_{m,n} \wp_{mn} \ O_{nm}$$

$$= \sum_{n} \wp_{nn}^O \ O_{nn}^O$$

$$= \sum_{n} \wp_{nn}^S \ O_{nn}^S.$$

(7.9)

In the second last equality the eigenstates of the observable operator have been used. This form only requires the diagonal elements of the probability operator, whether or not the latter is diagonalizable. This is identical to the classical probability form for a statistical average, namely the sum over states of the operator eigenvalue, $O_{nn}^O = O_n$, times the

probability of the state, $\wp_{nn}^O$. From the point of view of classical probability this is rather appealing, but it does raise the question of how it arises from the quantum expectation value form that reflects a superposition of states and interference between states.

The final equality for the statistical average is also problematic. Here the eigenstates of the probability operator are invoked. Although not essential for the present discussion, and anticipating somewhat results to come, these can be assumed to be entropy eigenstates and denoted by the superscript S, $\hat{\wp} \, | \, \zeta_n^S \rangle = \wp_n \, | \, \zeta_n^S \rangle$, and $\wp_{mn}^S = \wp_n \delta_{m,n}$. Again this is apparently dissonant with the expectation value for a wave state, where the density operator cannot be diagonalized. This demands a more careful examination of the origin of the trace form for the statistical average and of the properties of the probability operator. One can already see that any such diagonalizable probability operator cannot be a single wave function density operator.

The challenge is to get from the expectation value form to the statistical average form. In the former, the density operator always includes off-diagonal contributions, (except in the observable operator representation itself). In the latter the off-diagonal contributions vanish in both the entropy and the observable operator representations. The qualitative difference between these two expressions can be seen in figure 7.1.
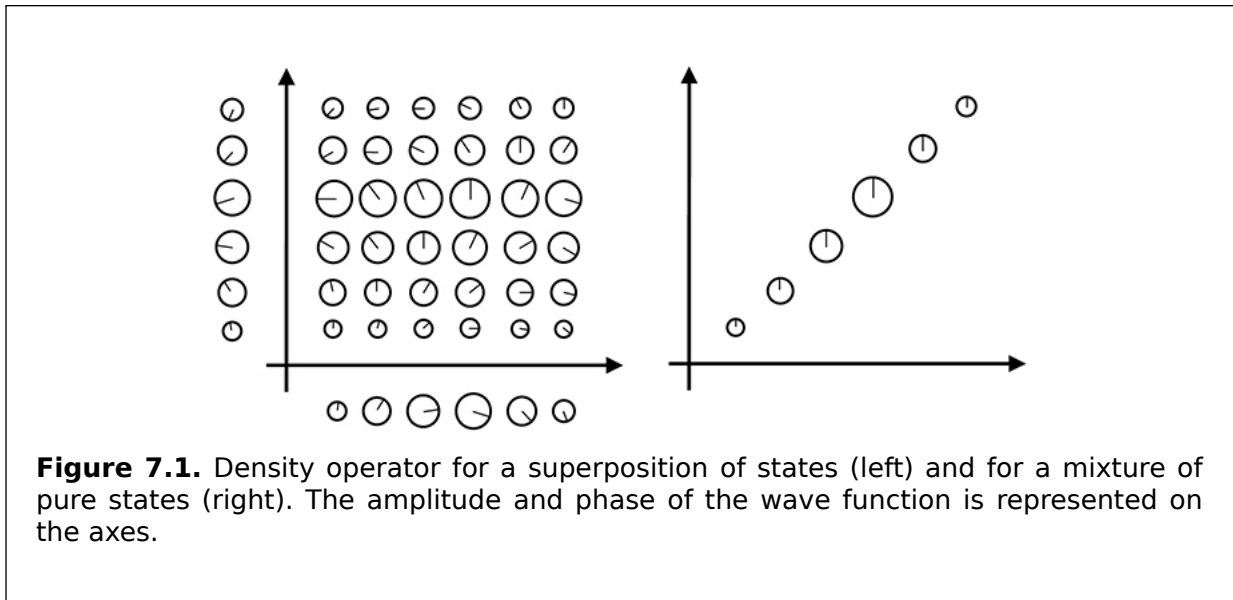

Figure 7.1. Density operator for a superposition of states (left) and for a mixture of pure states (right). The amplitude and phase of the wave function is represented on the axes.

The conventional view of quantum statistical mechanics is in terms of the density operator rather than the probability operator (von Neumann 1932, Messiah 1961, Merzbacher 1970). Because the density operator of a single-wave function cannot be diagonalized, most workers instead imagine an ensemble of systems, labeled $a = 1, 2, \ldots, M$, each with its own wave function ψ_a. The ensemble average of the single-wave function density operators defines the many-wave function density operator,

$$\hat{\bar{\rho}} \equiv \frac{1}{M} \sum_{a=1}^{M} \frac{| \, \psi_a \rangle \, \langle \psi_a \, |}{\langle \psi_a \, | \, \psi_a \rangle}.$$

(7.10)

(Almost always the wave functions are taken to be normalized, $\langle \psi_a \, | \, \psi_a \rangle = 1$, and the denominator is redundant.) This can also be called the averaged density operator. This does not factorize into a dyadic product, which is the crucial difference from the single-wave function density operator.

There is no reason to suppose that this cannot be diagonalized. Hence the many-wave function density operator can be equated to the probability operator introduced above. In the representation in which it is diagonal (presumably the entropy representation), the entries equal the state probabilities,

$$\bar{\rho}^{\mathrm{S}}_{mn} = \frac{1}{M} \sum_{a=1}^{M} \frac{\psi^{S*}_{a,m} \psi^{S}_{a,n}}{\langle \psi_a \mid \psi_a \rangle} = \wp_n \delta_{m,n}.$$

(7.11)

One can see that the off-diagonal terms average to zero if the phase angles of the coefficients are uncorrelated (decoherent). The diagonal terms all have zero phase angle exactly and so they cannot cancel with each other. This shows that the crucial difference between the one- and many-wave function density operators is this averaging process.

In view of this, in so far as one or other exists, the averaged density operator is equal to the probability operator, $\hat{\bar{\rho}} = \hat{\wp}$. This operator relationship holds in any representation, not just the entropy representation in which both were said to be diagonal. If the probability operator is an ordinary operator, then it can only be diagonal in the basis formed from the eigenfunctions of the entropy operator (or of any operator that commutes with the entropy operator). This would mean that the off-diagonal phase cancelation just discussed for the averaged density operator only occurred in the entropy basis and not in any other basis. It remains to prove this, either for the averaged density operator in the ensemble picture, or else for the probability operator directly.

With this definition of the many-wave function density operator, and its equality with the probability operator, the statistical average can be written as the trace of it and the observable operator,

$$\left\langle \hat{O} \right\rangle_{\mathrm{stat}} = \mathrm{TR}\ \hat{\bar{\rho}}\ \hat{O} = \frac{1}{M} \sum_{a=1}^{M} \frac{\langle \psi_a \mid \hat{O} \mid \psi_a \rangle}{\langle \psi_a \mid \psi_a \rangle}.$$

(7.12)

The conventional presentation of the ensemble picture is that the wave functions ψ_a are each a pure quantum state, in which case the ensemble comprises a mixture of pure quantum states. This means that each system in the ensemble has a fully collapsed wave function. The above discussion departs slightly from this in that it is sufficient for equation (7.11) to hold, which is to say that the phases of the entropy states are decoherent. In other words we need not insist that each member of the ensemble be in a pure quantum state.

In the presentation of classical statistical mechanics, the ensemble picture was criticized, sections 1.4.1 and 1.4.2. Since the many-wave particle density operator is a direct manifestation of the ensemble, it is similarly unsatisfactory. The problem is that such a collection of independent replica systems has no physical reality. The fact of the matter is that an actual measurement is made on a single system. This is not to say that the trace form of the quantum statistical average (with a probability operator equivalent to an averaged density operator) is incorrect; in fact this trace form has long been successfully used. Rather, it says that the explanation and justification for it does not lie in the ensemble picture. And of course, without the ensemble picture there is no need or basis for the many-wave function density operator.

Instead of an ensemble average one might consider that the decoherent system arises from a time average for the density operator. Let us imagine that a measurement extends over a brief time interval, τ. In this period the wave function evolves, $\psi(t)$, and one can define the averaged density operator as

$$\hat{\bar{\rho}} \equiv \frac{1}{\tau N_0} \int_0^\tau dt \ | \psi(t) \rangle \langle \psi(t) |,$$

(7.1
3)

assuming unitary evolution of the wave function, $N(\psi(t)) \equiv \langle \psi(t) | \psi(t) \rangle = N_0$. Again, assuming that this is the same as to the probability operator, $\hat{\bar{\rho}} = \hat{\wp}$, it must be diagonal in the entropy representation,

$$\bar{\rho}_{ln}{}^S = \frac{1}{\tau N_0} \int_0^\tau dt \ \psi_l^{S*}(t) \psi_n^S(t) = \wp_n \delta_{l,n}.$$

(7.1
4)

Presumably, the phase factors of the entropy modes evolve independently, which makes them effectively random. This means that the off-diagonal elements average to zero over the measurement time.

It is not clear that this putative temporal averaging over the measurement time is any more realistic than the ensemble averaging. Both explanations have the appearance of *post facto* rationalization. Under the circumstances it would perhaps be more honest not to pretend to justify equation (7.9), but rather simply to assert it as the definition of the statistical average, taking the existence of the probability operator as a given. In my opinion it is better not to introduce or to invoke the averaged density operator as it has no physical reality.

In the following two sections, a derivation from first principles is given of the von Neumann trace form for the statistical average equation (7.9). The derivation gives the explicit form for the probability operator and shows that it is diagonal in the entropy representation. It also demonstrates the passage from the wave function and expectation value formulation of quantum mechanics to the statistical average formulation as a weighted sum over quantum states.

7.2 Uniform weight density of wave space

In chapter 5 equilibrium classical statistical mechanics was derived. It was shown in section 5.2 that the entropy of a point in the phase space of an isolated system was uniform. This result was first discussed in terms of the conventional ergodic hypothesis, section 5.2.1, then in terms of constant probability on a trajectory, section 5.2.2, and finally, and most satisfactorily, a proof was given based on uniformity in time and the density of energy hypersurfaces, section 5.2.3.

For the present quantum problem one can proceed analogously. For an isolated system, in section 7.2.1 it is simply asserted that quantum states are equally probable. In section 7.2.2 it is shown that the probability density is constant on a trajectory. And in section 7.2.3 it is proven that the probability density is uniform in wave space based on time uniformity and energy and norm hypersurface density.

7.2.1 Equal state probability hypothesis

The simplest presentation of the probability operator focuses on the quantum states, and, as in the classical case, it begins with the microstates of an isolated system. From their weight the required quantum state probabilities follow, first for an isolated system, and thence for a sub-system interacting with a reservoir. By analogy with the ergodic hypothesis of classical statistical mechanics, one can take it as an axiom that

all quantum states of an isolated system have equal weight.

(7.1
5)

As will be seen, the quantum states are the eigenstates of any complete operator. These are the diagonal elements in any representation.

In the two following subsections, this hypothesis is actually derived from more formal considerations; it appears as equation (7.40). Instead of quantum states, the following analysis in sections 7.2.2 and 7.2.3 is based on the Hilbert space of wave functions. The aim is to show that the wave space probability density for an isolated system is uniform.

7.2.2 Trajectory uniformity

Consider an isolated system with wave function ψ. Schrödinger's equation for the time rate of change of the latter is

$$\mathrm{i}\hbar\dot{\psi}^0 = \hat{\mathcal{H}}\psi.$$

(7.16)

Here $\hat{\mathcal{H}}$ is the Hamiltonian or energy operator. This is here taken to be independent of time. As in the earlier classical analysis, the superscript 0 is used for adiabatic quantities, which are those of the isolated system. In this section only the total isolated system appears.

One can construct a trajectory by integrating this

$$\psi(t) \equiv \psi^0(t \mid \psi_0, t_0) = \hat{U}^0(t - t_0)\psi_0.$$

(7.17)

Here the adiabatic time propagator is $\hat{U}^0(t) = \mathrm{e}^{t\hat{\mathcal{H}}/\mathrm{i}\hbar}$. For a time-dependent Hamiltonian, this is $\hat{U}^0(t - t_0) = \mathrm{e}^{\int_{t_0}^{t} \mathrm{d}t\prime\, \hat{\mathcal{H}}(t\prime)/\mathrm{i}\hbar}$, with time-ordering assumed.

Each point ψ in wave space has a real, non-negative weight density $\omega(\psi, t)$. As usual, the normalized form of this is the probability density $\wp(\psi, t)$. In section 7.3 these will be related to the probability operator and to the statistical average.

As for any density, $\mathrm{d}\psi\, \wp(\psi, t)$ is the probability of the system being within $\mid \mathrm{d}\psi \mid$ of ψ. This is a real non-negative number: $\mathrm{d}\psi\, \wp(\psi, t) = |\mathrm{d}\psi\, \wp(\psi, t)| = |\mathrm{d}\psi|\, |\wp(\psi, t)|$. Since it is always the product that occurs, without loss of generality one may take each to be individually real.

The probability density is normalized

$$\int \mathrm{d}\psi\, \wp(\psi, t) = \int \mathrm{d}\underline{\psi}\, \wp(\underline{\psi}, t) = 1.$$

(7.18)

Here has been invoked an arbitrary representation, $\underline{\psi} = \{\psi_n\}, \ n = 1, 2, \ldots$. The coefficients are complex, $\psi_n = \psi_n^{\mathrm{r}} + \mathrm{i}\psi_n^{\mathrm{i}}$, and the infinitesimal volume element can be written as

$$\mathrm{d}\underline{\psi} = \mathrm{d}\underline{\psi}^{\mathrm{r}}\, \mathrm{d}\underline{\psi}^{\mathrm{i}} \equiv \mathrm{d}\psi_1^{\mathrm{r}}\, \mathrm{d}\psi_1^{\mathrm{i}}\, \mathrm{d}\psi_2^{\mathrm{r}}\, \mathrm{d}\psi_2^{\mathrm{i}} \cdots$$

(7.19)

With this all the integrations are over the real line, $\psi_n^{\mathrm{r}} \in [-\infty, \infty]$, and $\psi_n^{\mathrm{i}} \in [-\infty, \infty]$, $n = 1, 2, \ldots$.

As is the case for Hamilton's classical equations of motion, Schrödinger's equation gives an incompressible trajectory,

$$\frac{\mathrm{d}\dot{\psi}^0}{\mathrm{d}\psi} = \underline{\partial}_{\psi^{\mathrm{r}}} \cdot \dot{\underline{\psi}}^{\,0,\mathrm{r}} + \underline{\partial}_{\psi^{\mathrm{i}}} \cdot \dot{\underline{\psi}}^{\,0,\mathrm{i}}$$

$$= \underline{\partial}_{\psi} \cdot \dot{\underline{\psi}}^{\,0} + \underline{\partial}_{\psi^*} \cdot \dot{\underline{\psi}}^{\,0*}$$

$$= \frac{1}{\mathrm{i}\hbar}\mathrm{TR}\,\underline{\underline{\mathcal{H}}} - \frac{1}{\mathrm{i}\hbar}\mathrm{TR}\,\underline{\underline{\mathcal{H}}}$$

$$= 0.$$

(7.20)

The first two equalities are in fact the general expression for the trajectory compressibility. They hold as well for the total time derivative when the sub-system is open to a reservoir, heat bath, or environment. The final two equalities only hold when the evolution is given by Schrödinger's equation.

Again as in classical phase space, the compressibility gives the logarithmic rate of change of the volume element. Since the compressibility vanishes, the volume element is a constant of the motion of the isolated system,

$$\mathrm{d}\psi^0(t) = \mathrm{d}\psi_0.$$

(7.21)

The total time derivative of the probability density on the adiabatic trajectory is

$$\frac{\mathrm{d}^0 \wp(\psi,t)}{\mathrm{d}t} = \frac{\partial \wp(\psi,t)}{\partial t} + \dot{\underline{\psi}}^{\,0,\mathrm{r}} \cdot \underline{\partial}_{\psi^{\mathrm{r}}} \wp(\psi,t) + \dot{\underline{\psi}}^{\,0,\mathrm{i}} \cdot \underline{\partial}_{\psi^{\mathrm{i}}} \wp(\psi,t)$$

$$= \frac{\partial \wp(\psi,t)}{\partial t} + \dot{\underline{\psi}}^{\,0} \cdot \underline{\partial}_{\psi} \wp(\psi,t) + \dot{\underline{\psi}}^{\,0*} \cdot \underline{\partial}_{\psi^*} \wp(\psi,t)$$

$$= \frac{\partial \wp(\psi,t)}{\partial t} + \underline{\partial}_{\psi} \cdot \left[\dot{\underline{\psi}}^{\,0} \wp(\psi,t) \right] + \underline{\partial}_{\psi^*} \cdot \left[\dot{\underline{\psi}}^{\,0*} \wp(\psi,t) \right].$$

(7.22)

The first and second equalities are just the general definition of the total derivative. The final equality is valid on an adiabatic trajectory, upon which the compressibility vanishes. The probability flux can be seen to be $J_\wp(\psi,t) \equiv \dot{\psi}\wp(\psi,t)$.

For the isolated system, the probability density evolves adiabatically,

$$\wp(\psi_1,t_1) = \int \mathrm{d}\psi_0\, \wp(\psi_0,t_0)\, \delta\big(\psi_1 - \psi^0(t_1 \mid \psi_0,t_0)\big).$$

(7.23)

This invokes Bayes' theorem for the transition probability, essentially equation (3.14), $\wp(\psi_1,t_1;\psi_0,t_0) = \wp(\psi_1,t_1 \mid \psi_0,t_0)\,\wp(\psi_0,t_0)$, together with the type 1 reduction condition, equation (3.17). The conditional transition probability in the present adiabatic case is $\wp(\psi_1,t_1 \mid \psi_0,t_0) = \delta\big(\psi_1 - \psi^0(t_1 \mid \psi_0,t_0)\big)$, which reflects the deterministic nature of Schrödinger's equation.

From the incompressibility of wave space under Schrödinger's equation, $\mathrm{d}\psi_0 = \mathrm{d}\psi_1$. With $t_1 = t_0 + \Delta_t$, expanding to linear order in the time step yields

$$\wp(\psi_1, t_1) = \int \mathrm{d}\psi_0 \; \wp(\psi_0, t_0) \, \delta\big(\psi_1 - \psi^0(t_1 \mid \psi_0, t_0)\big)$$

$$= \wp\Big(\psi_1 - \Delta_t \dot{\psi}^0, t_0\Big)$$

$$= \wp(\psi_1, t_0) - \Delta_t \underline{\dot{\psi}}^0 \cdot \underline{\partial}_\psi \wp(\psi, t_0) - \Delta_t \underline{\dot{\psi}}^{0*} \cdot \underline{\partial}_{\psi^*} \wp(\psi, t_0).$$

(7.24)

Rearranging, this gives the partial time derivative as

$$\frac{\partial \wp(\psi, t)}{\partial t} = -\underline{\dot{\psi}}^0 \cdot \underline{\partial}_\psi \wp(\psi, t) - \underline{\dot{\psi}}^{0*} \cdot \underline{\partial}_{\psi^*} \wp(\psi, t).$$

(7.25)

This holds for an isolated system evolving under Schrödinger's equation.

Combining this with the second equality of equation (7.22), one sees that the total adiabatic time derivative of the probability density vanishes, $\mathrm{d}^0 \wp(\psi, t)/\mathrm{d}t = 0$. Hence for an isolated system, the probability density is a constant of the motion,

$$\wp(\psi(t), t) = \wp(\psi_0, t_0),$$

(7.26)

where $\psi(t) \equiv \psi^0(t \mid \psi_0, t_0)$.

This is the quantum analogue of equation (5.20), in which the total time derivative of the probability density in the phase space of an isolated system vanishes under Hamilton's equations. This is not surprising as the physical picture underlying the original classical result is the same as that which holds for the present quantum result. Like Hamilton's equations, Schrödinger's equation is deterministic, which means that trajectories do not cross, and they are neither created nor destroyed. In both cases trajectories within a given volume remain inside throughout its evolution. In so far as the probability of the volume is proportional to the number of trajectories it contains, this is conserved by both Hamilton's and Schrödinger's equation. What the two equations additionally have in common is that they are incompressible, which means that the volume is also a constant of the motion. These two facts mean that the probability density must be conserved moving along the trajectory.

For the present equilibrium system, the Hamiltonian operator does not depend upon time. In this case the probability density must be independent of time, $\wp(\psi)$. One therefore has the stronger result

$$\wp(\psi^0(t \mid \psi_0, t_0)) = \wp(\psi_0).$$

(7.27)

This says that the probability density is the same everywhere on a trajectory.

Finally, by analogy with the classical ergodic hypothesis, one may suppose that a single trajectory passes sufficiently close to all relevant points of the state space. The hypothesis says that one can take any two points in state space, ψ_1 and ψ_2, with the same norm and energy expectation to lie on a single trajectory. (It will be shown explicitly shortly that Schrödinger's equation conserves the norm and energy of the wave function.) By the above, they therefore have the same probability density

$$\wp(\psi_2) = \wp(\psi_1), \quad \text{if } E(\psi_2) = E(\psi_1) \quad \text{and} \quad N(\psi_2) = N(\psi_1).$$

(7.28)

Hence the wave space probability density must be of the form

$$\wp(\psi \mid E, N) = \frac{\delta(E(\psi) - E)\,\delta(N(\psi) - N)}{W(E, N)},$$

with the normalizing factor being

$$W(E, N) = \int \mathrm{d}\psi\; \delta(E(\psi) - E)\,\delta(N(\psi) - N).$$

These are the quantum analogues of equations (5.22) and (5.23). And so one is faced with an identical problem: in order to transform from the conditional wave space probability, $\wp(\psi \mid E, N)$, to the unconditional wave space probability, $\wp(\psi)$, one has to make some hypothesis about the dependence of the wave space weight density on the hypersurface, $w(E, N)$. If one assumes that this is constant or negligibly varying with energy and norm, then one can conclude that an isolated system has uniform weight in wave space,

$$\wp(\psi) = \frac{1}{W}, \quad W = \int \mathrm{d}\psi.$$

Obviously it is a little unsatisfactory to simply assume this. It is also unsatisfactory to assume that a single trajectory completely fills each hypersurface. Now an alternative derivation of the result is given, which avoids these assumptions. The approach is the direct analogue of the classical derivation of section 5.2.3.

7.2.3 Time average and hypersurface density

Schrödinger's equations conserve energy and norm. These are well-known results which are straightforward to prove directly.

The norm squared of the wave function is $N(\psi) \equiv \langle \psi \mid \psi \rangle$. Its adiabatic rate of change is

$$\begin{aligned}
\dot{N}^0(\psi) &= \langle \dot{\psi}^0 \mid \psi \rangle + \langle \psi \mid \dot{\psi}^0 \rangle \\
&= \left\langle \frac{1}{i\hbar}\hat{\mathcal{H}}\psi \mid \psi \right\rangle + \left\langle \psi \mid \frac{1}{i\hbar}\hat{\mathcal{H}}\psi \right\rangle \\
&= \frac{-1}{i\hbar}\langle \psi \mid \hat{\mathcal{H}}^\dagger \mid \psi \rangle + \frac{1}{i\hbar}\langle \psi \mid \hat{\mathcal{H}} \mid \psi \rangle \\
&= 0.
\end{aligned}$$

The final equality follows because the Hamiltonian operator, like all physical operators, is Hermitian conjugate, $\hat{\mathcal{H}}^\dagger = \hat{\mathcal{H}}$. In fact the Hamiltonian operator obeys a stronger symmetry: it is both real, $\hat{\mathcal{H}}^* = \hat{\mathcal{H}}$, and symmetric, $\hat{\mathcal{H}}^{\mathrm{T}} = \hat{\mathcal{H}}$.

The energy of the wave state ψ is $E(\psi) \equiv \langle \psi \mid \hat{\mathcal{H}} \mid \psi \rangle / N(\psi)$. Hence by a similar argument it is also a constant of the adiabatic motion,

$$\dot{\mathrm{E}}^0(\psi) = \frac{1}{N(\psi)}\langle \dot{\psi}^0 \mid \hat{\mathcal{H}} \mid \psi\rangle + \frac{1}{N(\psi)}\langle \psi \mid \hat{\mathcal{H}} \mid \dot{\psi}^0\rangle$$

$$= \frac{1}{N(\psi)}\left\langle \frac{1}{i\hbar}\hat{\mathcal{H}}\psi \mid \hat{\mathcal{H}} \mid \psi\right\rangle + \frac{1}{N(\psi)}\left\langle \psi \mid \hat{\mathcal{H}} \mid \frac{1}{i\hbar}\hat{\mathcal{H}}\psi\right\rangle$$

$$= \frac{-1}{i\hbar N(\psi)}\langle \psi \mid \hat{\mathcal{H}}^\dagger\hat{\mathcal{H}} \mid \psi\rangle + \frac{1}{i\hbar N(\psi)}\langle \psi \mid \hat{\mathcal{H}}\hat{\mathcal{H}} \mid \psi\rangle$$

$$= 0.$$

$$(7.33)$$

These results show that the trajectory of an isolated system, $\psi(t)$, is confined to a hypersurface of constant norm and energy. In view of this let $\tilde{\psi}$ denote a normalized wave function. Let χ denote a wave function on a hypersurface of constant energy E and norm squared N. This hypersurface can be labeled $\{N, E\}$, and it is obviously a sub-space of the Hilbert space of all wave functions. In this notation the normalized wave function is $\tilde{\psi} = \tilde{\psi}(\chi, E)$, and the full wave function is $\psi = \psi(\tilde{\psi}, N) = \psi(\chi, E, N)$.

Ultimately, the weight density of the full wave space of the isolated system, $\omega(\psi)$, is sought. Initially the weight density on the hypersurface, $\omega(\chi \mid N, E)$, is derived. This is then combined with the density of the $\{N, E\}$-hypersurface in the wave space to obtain the full weight density.

As in section 5.2.3, it is taken as axiomatic that time is homogeneous, which means that a statistical average is a simple time average over a trajectory. This implies that the weight density on the $\{N, E\}$-hypersurface must be inversely proportional to the speed of the trajectory,

$$\omega(\chi \mid N, E) \propto \mid \dot{\psi}^0 \mid^{-1}$$
$$= \langle \psi \mid \hat{\mathcal{H}}\hat{\mathcal{H}} \mid \psi\rangle^{-1/2},$$

$$(7.34)$$

since the Hamiltonian operator is Hermitian. Notice that this depends upon ψ. This is just the time that the system spends in a volume element $\mid \mathrm{d}\chi \mid$, which is the same as the time spent in the volume element of the full wave space $\mid \mathrm{d}\psi \mid$. The proportionality factor is an immaterial constant on the hypersurface. As argued in section 5.2.3, this constant cannot vary with energy E or with norm squared N because this would violate the axiom that a statistical average is a simple time average over the trajectory irrespective of energy or norm squared.

Now we transform this hypersurface weight density to the full wave space. The Jacobean for the transformation $\chi \Rightarrow \tilde{\psi}$ is

$$|\tilde{\nabla}E| = \left[\frac{\partial E(\psi)}{\partial |\tilde{\psi}\rangle}\frac{\partial E(\psi)}{\partial \langle\tilde{\psi}|}\right]^{1/2} = \langle \tilde{\psi} \mid \hat{\mathcal{H}}\hat{\mathcal{H}} \mid \tilde{\psi}\rangle^{1/2},$$

$$(7.35)$$

and that for the transformation $\tilde{\psi} \Rightarrow \psi$,

$$|\nabla N| = \left[\frac{\partial N(\psi)}{\partial |\psi\rangle}\frac{\partial N(\psi)}{\partial \langle\psi|}\right]^{1/2} = \langle \psi \mid \psi\rangle^{1/2}.$$

With these the full weight density is

$$\omega(\psi \mid N, E) = \omega(\chi \mid N, E) \frac{\mid \tilde{\nabla} E \mid}{\Delta_E} \frac{\mid \nabla N \mid}{\Delta_N}, \quad \mid N(\psi) - N \mid < \Delta_N, \ \mid E(\psi) - E \mid < \Delta_E$$

$$\propto \frac{\langle \tilde{\psi} \mid \hat{\mathcal{H}}\hat{\mathcal{H}} \mid \tilde{\psi} \rangle^{1/2} \langle \psi \mid \psi \rangle^{1/2}}{\langle \psi \mid \hat{\mathcal{H}}\hat{\mathcal{H}} \mid \psi \rangle^{1/2}} \ \delta(N(\psi) - N) \ \delta(E(\psi) - E)$$

$$= \ \delta(N(\psi) - N) \ \delta(E(\psi) - E).$$

Integrating this over all energies and norm squareds yields the unconditional weight density for the wave space of the isolated system,

$$\omega(\psi) = 1.$$

The interpretation of the origin of this result is essentially the same as that for the result for classical phase space in section 5.2.3. The weight density is inversely proportional to the speed of the trajectory (i.e. low speed equates to more time in a volume element), and proportional to the hypersurface density, (i.e. for fixed spacing between discrete hypersurfaces, Δ_E and Δ_N, large gradients correspond to more hypersurfaces per unit wave space volume). These ideas are the same as those depicted in the classical case in figure 5.3.

Just as in the classical case for Hamilton's equations of motion, it is a remarkable consequence of Schrödinger's equation that the speed is identical to the magnitude of the gradient of the hypersurface, so that these two cancel to give a weight density that is uniform in wave space. This result is the same as that given above as equation (7.31), which was obtained only by assuming that the hypersurface weight $w(E, N)$ was constant or negligibly varying.

This weight density is a physical observable, and as such there must be a corresponding weight operator whose expectation value it is. Hence one must have

$$\omega(\psi) \equiv \frac{\langle \psi \mid \hat{\omega} \mid \psi \rangle}{\langle \psi \mid \psi \rangle} = 1.$$

Since this must hold for all wave states, the weight operator for the isolated system must be the identity operator, and one must have

$$\hat{\omega} = \hat{\mathrm{I}}, \quad \text{or} \quad \omega_{mn} = \delta_{m,n}.$$

The result holds in any representation.

One conclusion that follows from this is that the quantum microstates of the system are not unique. Rather, any complete operator gives a collective of microstates. Also, a particular set of microstates is the set of diagonal elements of a particular representation. Equivalently, they are the eigenstates of any complete operator. This result implies that these microstates all have equal weight, which is the axiom given as equation (7.15) at the start of this section.

A second conclusion is that the off-diagonal elements in any representation have zero weight. It should be clear from this that the weight operator is of no use for the expectation value, since the off-diagonal elements of the single wave function density matrix, the dyadic $\psi_m^* \psi_n$, $m \neq n$, certainly contribute to the expectation value. Nevertheless, it will be

shown that for two interacting systems, the expectation value of an operator on one system has to be weighted by the states of the other system, and for this a weight operator is required (albeit one derived from the weight operator of the present isolated system).

7.3 Canonical equilibrium system

Now the canonical equilibrium system is analyzed in detail. As usual, the total isolated system consists of a sub-system and a thermal reservoir that can exchange energy with each other. In the literature the reservoir is often called the bath, or, particularly in the quantum field, the environment. Also the phrase 'the system' is often used to mean what is here called 'the sub-system'. Here and throughout this book the terms used are total system, sub-system, and reservoir. These are denoted by superscripts or subscripts tot, s, and r where any ambiguity exists.

7.3.1 Entropy of energy states

As outlined in section 1.1.2, the classical theory of probability is formulated in terms of microstates and macrostates, and these now need to be defined for a quantum system. A microstate, which is labeled by a single lower case Roman letter, usually corresponds to an eigenstate of a complete set of commuting operators. In this case the eigenvalue is unique. However, it can also correspond to an eigenstate of an incomplete operator. In such a case the microstates are degenerate and microstates with different labels can share the same eigenvalue of the incomplete operator. A macrostate, which is labeled by a lower case Greek letter, corresponds to the principal quantum number of an operator. Each macrostate has a unique label and corresponds to a unique eigenvalue of the operator. A Greek and a Roman letter paired together are used to signify a microstate, with the Greek letter labeling the principal quantum number, and the Roman letter labeling the degenerate quantum states.

For example, writing the eigenfunctions with a single microstate label, one has $\hat{O} \, | \, \zeta_n^{\mathrm{O}} \rangle = O_n \, | \, \zeta_n^{\mathrm{O}} \rangle$. In this case different values of n may yield the same eigenvalue (assuming that $\hat{O}$ is incomplete operator). Alternatively, the eigenfunctions may be written as the combination of principal and degenerate state labels, $\hat{O} \, | \, \zeta_{\alpha g}^{\mathrm{O}} \rangle = O_\alpha \, | \, \zeta_{\alpha g}^{\mathrm{O}} \rangle$. In this case different values of α necessarily yield different eigenvalues.

For the present canonical equilibrium system, it is energy that is exchanged between the sub-system and the reservoir. Hence the focus is here on energy states. Since the microstates of an isolated system must have equal weight, equation (7.40) the weight of the energy microstates is $w_{\alpha g}^{\mathrm{E}} = 1$, since without loss of generality this can be set to unity. This means that the energy microstates of the total system have no internal entropy, $S_{\alpha g}^{\mathrm{E}} = k_{\mathrm{B}} \, \ln w_{\alpha g}^{\mathrm{E}} = 0$.

In the general formulation of classical probability, section 1.1.2, the weight of a macrostate is the sum of the weights of the microstates contained within it. In the present case the energy macrostate weight is

$$w_\alpha^{\mathrm{E}} = \sum_g {}^{(\mathrm{E}, \alpha)} w_{\alpha g}^{\mathrm{E}} \equiv N_\alpha^{\mathrm{E}}.$$

(7.41)

This is just the number of degenerate states with energy E_α. Hence the entropy of an energy macrostate is the logarithm of the degeneracy,

$$S_\alpha^{\mathrm{E}} = k_{\mathrm{B}} \, \ln N_\alpha^{\mathrm{E}}.$$

Analogous to the analysis for classical thermodynamics, section 2.3.1, the entropy of the reservoir when the sub-system is in a particular energy microstate will be required. Since energy is conserved one must have

$$\mathrm{E^{tot} = E^s + E^r}.$$

As mentioned, tot means the total system, s means the sub-system, and r means the reservoir. The interaction energy between the sub-system and the reservoir is included in the reservoir energy.

Now the reservoir entropy can be written in two different ways, which will be equated to each other below. The first expression directly invokes the energy degeneracy discussed above,

$$S^{\mathrm{r}}(E^{\mathrm{r}}_{\alpha}) = k_{\mathrm{B}} \ln N^{\mathrm{Er}}_{\alpha}, \quad \text{with } N^{\mathrm{Er}}_{\alpha} \equiv \sum_{g \in \alpha}{}^{(\mathrm{Er})}.$$

This is the entropy of the isolated reservoir in this energy macrostate.

The second expression for the reservoir entropy invokes the definition of the thermodynamic temperature, equation (2.11), $T^{-1} \equiv \partial S(E)/\partial E$. By definition the reservoir is much larger than the sub-system, and so the sub-system energy may be used as an expansion variable. The Taylor expansion yields

$$S^{\mathrm{r}}(E^{\mathrm{r}}) = S^{\mathrm{r}}(E^{\mathrm{tot}} - E^{\mathrm{s}}) = \text{const.} - \frac{E^{\mathrm{s}}}{T^{\mathrm{r}}}.$$

The higher order terms are negligible in the thermodynamic limit. This gives the reservoir entropy as a function of the sub-system energy. Henceforth the superscript on the reservoir temperature will be dropped as it is the only temperature that will appear.

7.3.2 Wave function entanglement

For the total system, the wave function lies in the Hilbert space that comprises that of the sub-system and the reservoir. It is assumed that the sub-system and the reservoir interact so weakly that they may be treated as quasi-independent. This means that for any state of one all permitted states of the other are available. The total number of states is essentially the product of the two individual totals, subject to the conservation laws. Also the basis functions for the total system are the product of the basis functions of the sub-system and of the reservoir, each considered as isolated.

With $\{|\,\zeta^{\mathrm{s}}_n\rangle\}$ an orthonormal basis for the sub-system and $\{|\,\zeta^{\mathrm{r}}_m\rangle\}$ an orthonormal basis for the reservoir, the most general wave state of the total system can be expanded in terms of these basis functions,

$$|\,\psi_{\mathrm{tot}}\rangle = \sum_{n,m} c_{nm} \,|\,\zeta^{\mathrm{s}}_n, \zeta^{\mathrm{r}}_m\rangle.$$

If the coefficient matrix is dyadic, $c_{nm} = c^{\mathrm{s}}_n c^{\mathrm{r}}_m$, then the wave state is separable, $|\,\psi_{\mathrm{tot}}\rangle = |\,\psi_{\mathrm{s}}, \psi_{\mathrm{r}}\rangle$ with $|\,\psi_{\mathrm{s}}\rangle = \sum_n {}^{(\mathrm{s})}c^{\mathrm{s}}_n \,|\,\zeta^{\mathrm{s}}_n\rangle$, and $|\,\psi_{\mathrm{r}}\rangle = \sum_m {}^{(\mathrm{r})}c^{\mathrm{r}}_m \,|\,\zeta^{\mathrm{r}}_m\rangle$. Alternatively, if the coefficients are not dyadic, then the wave state is inseparable, which is called an entangled state (Messiah 1961, Merzbacher 1970).

In general terms entangled states arise from the conservation laws. This can be seen explicitly for the present case of energy exchange, in which the total energy is fixed,

$E^{\mathrm{tot}} = E^{\mathrm{s}} + E^{\mathrm{r}}$. (Recall that the interaction energy is included in the reservoir energy.) Using the respective energy eigenfunctions as a basis for the sub-system and the reservoir, the most general expansion of the total wave function is

$$| \, \psi_{\mathrm{tot}} \rangle = \sum_{\alpha g, \beta h} c_{\alpha g, \beta h} \, | \, \zeta_{\alpha g}^{\mathrm{Es}}, \zeta_{\beta h}^{\mathrm{Er}} \rangle.$$

(7.47)

From energy conservation one must have that

$$c_{\alpha g, \beta h} = 0 \quad \text{if} \ E_{\alpha}^{\mathrm{s}} + E_{\beta}^{\mathrm{r}} \neq E_{\mathrm{tot}}.$$

(7.48)

This condition cannot be satisfied for a system able to exchange energy if the coefficient is in dyadic from. To see this, suppose the contrary, that $c_{\alpha g, \beta h} = c_{\alpha g}^{\mathrm{s}} c_{\beta h}^{\mathrm{r}}$. Denote one of the occupied sub-system energy macrostates by α_1, so that $c_{\alpha_1 g}^{\mathrm{s}} \neq 0$. Because the macrostate labels are distinct, there can be only one nonzero reservoir macrostate, say β_1, that conserves energy, $E_{\alpha_1}^{\mathrm{s}} + E_{\beta_1}^{\mathrm{r}} = E_{\mathrm{tot}}$. Hence one must have that $c_{\beta h}^{\mathrm{r}} = 0$, $\beta \neq \beta_1$. But a unique occupied reservoir energy macrostate implies that α_1 must also be unique, which is to say that there can be no other occupied sub-system energy macrostate.

This shows that if the total wave function is in product form, then the sub-system and reservoir each can only have a single fixed energy. This in turn means that neither wave function, ψ_{s} or ψ_{r}, can contain a superposition of principal energy states (because these are non-degenerate). It follows from this that the sub-system is unable to change energy by exchange with the reservoir, because during such an exchange it would be in a superposition of the initial and final energy states. Therefore, a dyadic form for the expansion coefficient corresponds to the sub-system and the reservoir being isolated from each other. But this contradicts the basis of a canonical equilibrium system, namely that the energy of the sub-system fluctuates over time as it exchanges energy with the reservoir. One concludes from this contradiction that if the sub-system and the reservoir can exchange energy, then the total wave function must be entangled.

Accordingly, for the present canonical equilibrium system, the total wave function cannot be factorized. Instead it must have a representation of the form

$$| \, \psi_{\mathrm{tot}} \rangle = \sum_{\alpha, g, h} c_{\alpha g, h} \, | \, \zeta_{\alpha g}^{\mathrm{Es}}, \zeta_{\beta_\alpha h}^{\mathrm{Er}} \rangle.$$

(7.49)

Here the reservoir principal energy label β_α depends upon the sub-system principal energy α and is defined implicitly by $E_{\alpha}^{\mathrm{s}} + E_{\beta_\alpha}^{\mathrm{r}} = E_{\mathrm{tot}}$. For brevity, the coefficients are written, $c_{\alpha g, h} \equiv c_{\alpha g}^{\mathrm{s}} c_{\beta_\alpha h}^{\mathrm{r}}$. Because β_α depends upon α, this is entangled (i.e. the sum of the products is not equal to the product of the sums).

The most well-known case of entanglement is that of particles with spin that result from radioactive decay (figure 7.2). It is the conservation of spin that is responsible for the entanglement. Measurement of the spin of either daughter particle collapses the superposed wave function into one or other of the two wave functions that form it. If it were not for spin conservation, the daughter wave function would be the superposition of four possible wave functions, which are the independent (i.e. unentangled) product of two spinors for each particle (figure 7.3). It is spin conservation that entangles the two daughter particles and that reduces the superposition basis set from four to two possibilities.
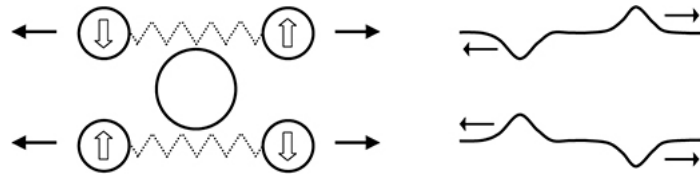
Figure 7.2. The two possible decays (upper and lower left) of a spin 0 particle (center left) into two spin half particles. On the right are the two wave functions that are superposed for the final state. Each represents the entanglement of two single-particle wave functions.
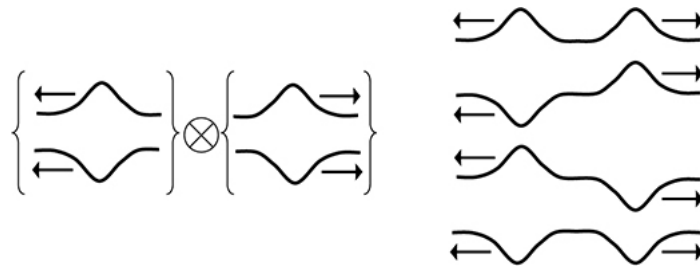

Figure 7.3. The dyadic product of two spin half-particle wave functions (left), and the four wave functions this gives for superposition, each representing two unentangled single-particle wave functions (right).

It should be clear in this example that the entangled and unentangled wave functions are very different. For the entangled case, a measurement of the product of the spins of the two daughter particles yields $\langle \psi \mid \hat{\sigma}_l \hat{\sigma}_r \mid \psi \rangle = -1$. The same measurement on the unentangled wave function yields $\langle \psi \mid \hat{\sigma}_l \hat{\sigma}_r \mid \psi \rangle = 0$. This example is but one of many that could be offered, since quite generally entanglement has far-reaching consequences.

7.3.3 Expectation values and wave function collapse

For the present canonical equilibrium system, conservation of energy causes the entanglement of the sub-system and reservoir wave functions. It will now be shown that this causes the principal energy states of the sub-system to collapse, which is to say that a sub-system expectation value is diagonal in these. This is quite significant for quantum statistical mechanics in view of the discussion in section 7.1 that the single wave function density operator cannot be diagonalized. In contrast, if a probability operator existed in the usual sense of an observable operator, then it can be diagonalized. Therefore, the collapse of the principal energy states due to entanglement that will now be established already goes a long way to proving the existence of the probability operator.

In view of the expansion for the wave function of the total system, equation (7.49), and assuming a normalized wave function, $\langle \psi_{\text{tot}} \mid \psi_{\text{tot}} \rangle = 1$, the expectation value of an operator on the sub-system is

$$\langle\psi_{\text{tot}}|\hat{O}^{\text{s}}|\psi_{\text{tot}}\rangle = \sum_{\alpha\prime g\prime;h\prime}\sum_{\alpha g;h} c^{*}_{\alpha\prime g\prime,h\prime}\, c_{\alpha g,h}\Big\langle\zeta^{\text{Es}}_{\alpha\prime g\prime},\zeta^{\text{Er}}_{\beta_{\alpha\prime}h\prime}|\hat{O}^{\text{s}}|\zeta^{\text{Es}}_{\alpha g},\zeta^{\text{Er}}_{\beta_{\alpha}h}\Big\rangle$$

$$= \sum_{\alpha\prime g\prime;h\prime}\sum_{\alpha g;h} c^{*}_{\alpha\prime g\prime,h\prime}\, c_{\alpha g,h}\big\langle\zeta^{\text{Es}}_{\alpha\prime g\prime}|\hat{O}^{\text{s}}|\zeta^{\text{Es}}_{\alpha g}\big\rangle\Big\langle\zeta^{\text{Er}}_{\beta_{\alpha\prime}h\prime}|\,\zeta^{\text{Er}}_{\beta_{\alpha}h}\Big\rangle$$

$$= \sum_{\alpha,g,g\prime}\sum_{h\in\beta_{\alpha}}{}^{(\text{Er})}\, c^{*}_{\alpha g\prime,h}\, c_{\alpha g,h}\big\langle\zeta^{\text{Es}}_{\alpha g\prime}|\hat{O}^{\text{s}}|\zeta^{\text{Es}}_{\alpha g}\big\rangle$$

$$= \sum_{\alpha,g,g\prime}\sum_{h\in\beta_{\alpha}}{}^{(\text{Er})}\, c^{*}_{\alpha g\prime,h}\, c_{\alpha g,h}\, O^{\text{s,E}}_{\alpha g\prime,\alpha g}.$$

(7.50)

The third equality follows because the relationship between α and β_{α} is bijective (the macrostates are non-degenerate), $\big\langle\zeta^{\text{Er}}_{\beta_{\alpha\prime}h\prime}\,|\,\zeta^{\text{Er}}_{\beta_{\alpha}h}\big\rangle = \delta_{\alpha\prime,\alpha}\,\delta_{h\prime,h}$. Only the principal diagonal elements of the sub-system matrix operator in the energy representation contribute to this. This can be equivalently cast in terms of the diagonal elements of the density matrix. The projection onto the sub-system in this case has the effect of converting a superposition of the sub-system basis states into a mixture of pure sub-system basis states, as arises from a collapsed wave function.

In this result energy conservation and entanglement have reduced and collapsed the original four sums over independent principal energy states into a single sum over principal energy states, from $\alpha\prime\alpha,\beta\prime\beta$ then to two sums, $\alpha\prime\alpha$, and then to a single sum, α. The degenerate energy states of the reservoir have also collapsed, since the orthogonality of the reservoir basis functions have reduced the sums over h and $h\prime$ to a single sum over h. Specifically, whereas for an isolated system the non-diagonal entries of the operator matrix in the energy representation would contribute to the expectation value, when energy exchange with a reservoir occurs, orthogonality of the reservoir energy basis functions together with energy conservation eliminate the off-diagonal sub-system energy terms. The principal energy states have at this stage collapsed, whilst the degenerate energy states of the sub-system remain in superposition form.

Here collapse refers to the collapse of the wave function and to the collapse into principal energy states. Collapse means that the number of possible states has been reduced, and also that the interference between superposition states has been reduced by the elimination of off-diagonal energy terms.

7.3.3.1 Random phase approach

The microstates of the total isolated system have equal weight. This can be taken as an hypothesis, equation (7.15), or as a theorem that follows from the uniform density of wave space, equation (7.40). In view of this, the coefficients in the representation of the total wave function all have the same magnitude, which can be set to unity, $|\,c_{\alpha g,h}\,|=1$. (With this choice, the expectation value, which is actually the statistical average, has to be explicitly normalized.) The coefficients therefore are of the form

$$c_{\alpha g,h} = \mathrm{e}^{\mathrm{i}\theta_{\alpha g,h}},$$

(7.51)

with the phase θ being real and $\mathrm{i}\equiv\sqrt{-1}$.

This form is reminiscent of the so-called EPR (Einstein–Podolsky–Rosen) state in which entangled qubits are often described in similar terms (cf figure 7.2).

A given total wave functions correspond to a specific sets of phases, $\{\theta_{\alpha g,h}\}$. Therefore, averaging over the phases makes the product of coefficients that appears in equation

([7.50](#)), zero unless $g\prime = g$,

$$\langle c^*_{\alpha g\prime,h}\, c_{\alpha g,h} \rangle_{\text{stat}} = \langle e^{i[\theta_{\alpha g,h} - \theta_{\alpha g\prime,h}]} \rangle_{\text{stat}} = \delta_{g,g\prime}.$$

(7.52)

This will be explicitly shown in the next section by averaging over the total wave space. With this, the reservoir sum in the expectation value becomes

$$\left\langle \sum_{h\in\beta_\alpha}^{(\text{Er})} c^*_{\alpha g\prime,h}\, c_{\alpha g,h} \right\rangle_{\text{stat}} = \delta_{g,g\prime} \sum_{h\in\beta_\alpha}^{(\text{Er})}$$

$$= \delta_{g,g\prime}\, N^{\text{Er}}_{\beta_\alpha}$$

$$= \delta_{g,g\prime}\, e^{S^{\text{r}}_{\beta_\alpha}/k_{\text{B}}}$$

$$= \delta_{g,g\prime}\, e^{-E^{\text{s}}_\alpha/k_{\text{B}}T}.$$

(7.53)

The statistical average on the left-hand side means the average over total wave functions, which means the average over randomly and uniformly distributed phases. The penultimate equality gives the entropy of the reservoir macrostate β_α as the logarithm of the number of microstates it contains, equation ([7.44](#)), and the final equality uses the thermodynamic relationship, equation ([7.45](#)), with the constant part effectively incorporated into the normalization factor.

The effect of averaging the expectation value over the wave function is to collapse the degenerate sub-system energy states, $g = g\prime$. This both reduces the number of elements contributing to the expectation value and removes the interference between the degenerate energy states (i.e. they are no longer in superposition). Hence both the principal and degenerate states in the energy representation of an operator on the sub-system are in diagonal form. It will be recalled that this result had to be established in order to derive the mixture form for the statistical average, equation ([7.9](#)). Equivalently, the probability operator is fully diagonal in the energy representation. Since in the canonical equilibrium system, the entropy operator is proportional to the energy operator, $\hat{S} = -\hat{\mathcal{H}}/T$, this means that the probability operator is fully diagonal in the entropy representation. (In general the entropy operator is derived from the variables that are exchangeable between the sub-system and the reservoir, and so in all equilibrium and non-equilibrium systems the probability operator will be diagonal in the entropy representation.)

The diagonal nature of the probability operator signifies the elimination of the interference between energy states, both principal and degenerate. This transforms the sub-system from a superposition of states to a mixture of states.

The statistical average also counts the degenerate reservoir energy states. The sum of these gives a weight to each sub-system energy state that is just the number of degenerate reservoir microstates. But by equation ([7.44](#)), this is just the exponential of the corresponding reservoir entropy. It is this second effect, combined with equation ([7.45](#)), that gives the Maxwell–Boltzmann form for the probability operator.

The average over the phases of the expectation value, equation ([7.50](#)), is the statistical average of the operator. With the above result it is

$$\left\langle \hat{O}^{\text{s}} \right\rangle_{\text{stat}} = \frac{1}{Z} \sum_{\alpha,g} e^{-E_\alpha/k_{\text{B}}T} O^{\text{s,E}}_{\alpha g,\alpha g}.$$

(7.54)

Here Z is just the normalization constant (actually partition function), and the sum is over the primary and degenerate energy states of the sub-system. This has the same functional form as the final equality in equation (7.9), which is the von Neumann trace form for the statistical average. This gives explicitly the probability matrix as the Maxwell–Boltzmann form, $\wp^{S}_{\alpha g,\alpha g} = Z^{-1}e^{-E_\alpha/k_B T}$.

7.3.3.2 Wave space approach

The statistical average of an operator on the sub-system can also be expressed as an integral over the total wave space of the expectation value with respect to the entangled wave function. The weight density of the total wave space is uniform, section 7.2. The integral over wave space is carried out by integrating the coefficients in the energy basis state expansion. The allowed coefficients respect energy conservation, equation (7.49), which means that $\mathrm{d}\psi_{\mathrm{tot}} \equiv \mathrm{d}\underline{c} \equiv \prod_{\alpha g,h} \mathrm{d}c^r_{\alpha g,h}\,\mathrm{d}c^i_{\alpha g,h}$. The real and imaginary parts of the coefficient each belong to the real line, $\in (-\infty, \infty)$. Entanglement has reduced the number of indeces from four to three. With the expectation value given by equation (7.50), $O^s(\psi_{\mathrm{tot}}) \equiv \langle \psi_{\mathrm{tot}} \mid \hat{O}^s \mid \psi_{\mathrm{tot}}\rangle/\langle \psi_{\mathrm{tot}} \mid \psi_{\mathrm{tot}}\rangle$, the statistical average is

$$
\begin{aligned}
\left\langle \hat{O}^s \right\rangle_{\mathrm{stat}} &= \frac{1}{Z\prime} \int \mathrm{d}\psi_{\mathrm{tot}}\, O^s(\psi_{\mathrm{tot}}) \\
&= \frac{1}{Z\prime} \sum_{\alpha,g,g\prime} \sum_{h\in\beta_\alpha} {}^{(\mathrm{Er})} O^{s,E}_{\alpha g\prime,\alpha g} \int \mathrm{d}\underline{c}\, \frac{c^{*}_{\alpha g\prime,h}\, c_{\alpha g,h}}{N(\psi_{\mathrm{tot}})} \\
&= \frac{1}{Z\prime} \sum_{\alpha,g} \sum_{h\in\beta_\alpha} {}^{(\mathrm{Er})} O^{s,E}_{\alpha g,\alpha g} \int \mathrm{d}\underline{c}\, \frac{\mid c_{\alpha g,h} \mid^2}{N(\psi_{\mathrm{tot}})} \\
&= \frac{1}{Z\prime} \sum_{\alpha,g} e^{S^r_{\beta_\alpha}/k_B} O^{s,E}_{\alpha g,\alpha g} \times \mathrm{const.} \\
&= \frac{1}{Z} \sum_{\alpha,g} e^{-E^s_\alpha/k_B T} O^{s,E}_{\alpha g,\alpha g}.
\end{aligned}
$$

(7.5
5)

Here the norm squared is $N(\psi_{\mathrm{tot}}) = \sum_{\alpha g,h} \mid c_{\alpha g,h} \mid^2$. The third equality follows since the terms in which the integrand is odd with respect to any coefficient vanish, and so the only non-vanishing terms have $g = g\prime$. The fourth equality follows because all the integrations are identical and give a constant that is independent of α, g, and h, which can be incorporated into the normalization factor. The fourth and fifth equalities use the two forms for the reservoir entropy, equations (7.44) and (7.45). The partition function ensures that $\langle \hat{I}^s \rangle_{\mathrm{stat}} = 1$.

This agrees with the random phase approach, equation (7.54), which should not be surprising since the uniformity of wave space of an isolated system implies the equal weight of isolated system microstates. Both methods show that the reduction and collapse of the energy macrostates is due to energy conservation and the consequent entanglement of the total wave function. They both also show that the collapse of the sub-system degenerate energy states is due to the fact that the off-diagonal contributions average to zero.

This equivalence demonstrates that quantum statistical mechanics can be as well formulated in terms of the wave function as in terms of quantum states.

7.3.4 Maxwell–Boltzmann probability operator

The statistical average given by both approaches agree with the von Neumann trace form, specifically the final equality in equation (7.9). This is in the form of a classical average: it is the sum over states of the probability of a state times the value of the observable in the state. It is essential to the result that the off-diagonal elements in the entropy representation of the probability (density) matrix vanish, which is to say that the wave function in this basis is decoherent.

One can easily convert the present result for the average to the generic von Neumann trace form, the first equality in equation (7.9). Of course, the probability operator replaces the conventional many-wave function density operator. In the energy basis, the representation of any function of the Hamiltonian operator is diagonal, and the diagonal entries are just the function of the eigenvalues,

$$
\mathrm{e}^{-E_\alpha^\mathrm{s}/k_\mathrm{B}T}\delta_{\alpha',\alpha}\delta_{g',g} = \left\{\mathrm{e}^{-\hat{\mathcal{H}}/k_\mathrm{B}T}\right\}_{\alpha'g',\alpha g}^{\mathrm{E}},
$$

(7.56)

with $\hat{\mathcal{H}}$ being the sub-system Hamiltonian operator. For the canonical equilibrium system, the reservoir entropy operator is $\hat{S} = -\hat{\mathcal{H}}/T$, a consequence of equation (7.45). This means that in the present canonical equilibrium system entropy eigenfunctions are identical to the energy eigenfunctions. More generally, the entropy representation is always the representation in which an open quantum system is diagonal. With this the quantum statistical average of a sub-system operator can be written

$$
\begin{aligned}
\left\langle \hat{O} \right\rangle_\mathrm{stat} &= \frac{1}{Z}\sum_{\alpha,g}\mathrm{e}^{-E_\alpha/k_\mathrm{B}T}O_{\alpha g,\alpha g}^{\mathrm{s,E}} \\
&= \frac{1}{Z}\sum_{\alpha g,\alpha'g'}\left\{\mathrm{e}^{-\hat{\mathcal{H}}/k_\mathrm{B}T}\right\}_{\alpha'g',\alpha g}^{\mathrm{E}}O_{\alpha g,\alpha'g'}^{\mathrm{s,E}} \\
&= \mathrm{TR}\ \hat{\wp}\ \hat{O}.
\end{aligned}
$$

(7.57)

The final form is independent of any specific representation or basis. One can conclude that the probability operator for the canonical equilibrium system is just the Maxwell–Boltzmann operator,

$$
\hat{\wp} \equiv \frac{1}{Z}\mathrm{e}^{\hat{S}/k_\mathrm{B}} \equiv \frac{1}{Z}\mathrm{e}^{-\hat{\mathcal{H}}/k_\mathrm{B}T}.
$$

(7.58)

The normalizing partition function is $Z = \mathrm{TR}\ \mathrm{e}^{-\hat{\mathcal{H}}/k_\mathrm{B}T}$. The first equality holds in general, whereas the second equality is specific for the canonical equilibrium system. This result for the canonical equilibrium probability operator (or many-wave function density operator) is well known, for example Feynman (1998), but seldom derived.

The Maxwell–Boltzmann probability operator is the direct quantitative consequence of the sum over the degenerate energy microstates of the reservoir. This gives the exponential of the reservoir entropy for each particular sub-system energy macrostate. This is just the matrix representation of the Maxwell–Boltzmann probability operator in the energy basis. Although the energy representation was used to derive this result, the final expression as a trace of the product of the two operators is invariant with respect to the representation.

One sees in this derivation that it is the statistical average that causes the superposition of the degenerate energy microstates of the sub-system to collapse into a mixture. The

reduction and collapse of the energy macrostates occurred at the level of the expectation value due to the energy conservation law and the entanglement of the reservoir and the sub-system. As has been mentioned above, both collapses are necessary for the mixture form for the statistical average, equation (7.9). The superposition entropy states, both principle and degenerate, had to collapse so that the probability operator was diagonal in the entropy representation. Although we do not work with a density operator, if we were to, the result (i.e. equating it to the probability operator) would imply that the density operator is an average over multiple wave functions.

The Maxwell–Boltzmann probability operator can be written as a Feynman path integral for the time propagator (Feynman and Hibbs 1965). In this approach the temperature is interpreted as an imaginary time. Although this technique has been successfully exploited in many applications (Schulman 1981, Kleinert 2009) it is not analyzed or invoked in what follows.

For the present formulation, the statistical average can be written as an integral of the expectation value over the sub-system wave space

$$\left\langle \hat{O} \right\rangle_{\text{stat}} = \int \mathrm{d}\psi \, \frac{\langle \psi \mid \hat{\wp} \, \hat{O} \mid \psi \rangle}{\langle \psi \mid \psi \rangle}.$$

(7.59)

It is no longer necessary to show that these belong to the sub-system, the superscript s, because the reservoir has been integrated out. This is a continuum version of the trace. The equivalence of this with the above discrete expression for the trace can be shown by expanding the sub-system wave function in any basis, $\mid \psi \rangle = \sum_n \psi_n \mid \zeta_n \rangle$, and integrating,

$$\int \mathrm{d}\psi \, \frac{\langle \psi \mid \hat{\wp} \, \hat{O} \mid \psi \rangle}{\langle \psi \mid \psi \rangle} = \int \mathrm{d}\underline{\psi} \, \frac{\sum_{m,n,l} \psi_m^* \psi_l \wp_{mn} O_{nl}}{\langle \psi \mid \psi \rangle}$$

$$= \frac{1}{Z''} \sum_{m,n} \left\{ \mathrm{e}^{\hat{S}/k_{\mathrm{B}}} \right\}_{mn} O_{nm} \int \mathrm{d}\underline{\psi} \, \frac{\psi_m^* \psi_m}{\langle \psi \mid \psi \rangle}$$

$$= \frac{\text{const.}}{Z''} \sum_{m,n} \left\{ \mathrm{e}^{\hat{S}/k_{\mathrm{B}}} \right\}_{mn} O_{nm}$$

$$= \mathrm{TR} \, \hat{\wp} \, \hat{O}.$$

(7.60)

In the second equality $l = m$ because odd powers of ψ_l vanish. Since the integration is the same for all m, it is a constant that can be taken outside of the sums and incorporated into the partition function, as in the third equality.

In section 7.2, the probability density in wave space for an isolated total system was shown to be uniform in wave space. In contrast, the probability density in the wave space of a sub-system of a total system is non-uniform. For the present canonical equilibrium sub-system that can exchange energy with a reservoir, it is given by the expectation value of the Maxwell–Boltzmann probability operator. The probability density is the expectation value of the probability operator,

$$\wp(\psi) = \frac{\langle \psi \mid \hat{\wp} \mid \psi \rangle}{\langle \psi \mid \psi \rangle},$$

(7.61)

where ψ is the wave function of the sub-system. It should be clear that the statistical average is *not* the integral over the sub-system wave space of the probability density and the expectation value of the operator,

$$\left\langle \hat{O} \right\rangle_{\text{stat}} \neq \int \mathrm{d}\psi \; \wp(\psi) O(\psi).$$

(7.62)

The physical reason that this does not hold is that in general statistical averages are the value of a variable in a state times the probability of the state, summed over the states. The problem with this expression, is that both $\wp(\psi)$ and $O(\psi)$ are expectation values, which include the superposition of states.

7.3.4.1 Environmental selection

In the quantum literature, the present sub-system that can exchange energy with a thermal reservoir is generally called an open quantum system. These have been treated in detail from a number of different perspectives (Davies 1976, Breuer and Petruccione 2002, Weiss 2008). One approach that is related to that developed here is environmental selection, or einselection (Zeh 2001, Zurek 2003 Schlosshauer 2005). Einselection includes the influence of the environment on the system of interest, and this is one obvious point in common with the above analysis of the sub-system and reservoir. The specific goal of einselection is to explain the apparent collapse and decoherence of the wave function upon measurement. Wave function collapse occurred several times in the derivation of equilibrium quantum statistical mechanics just given, and this is another point of similarity. Although the quantitative details differ, the basic conclusion of einselection—that the reservoir or environment suppresses the superposition states of the open sub-system rendering it diagonal—is essentially the same as that drawn here. A more detailed discussion of einselction in the context of the present approach is given in section 1.4 of Attard (2015).

7.4 Quantum statistical mechanics in classical phase space

7.4.1 Partition function

The probability operator derived above for the canonical equilibrium system is

$$\hat{\wp} = \frac{\mathrm{e}^{-\beta\hat{\mathcal{H}}}}{Z(T)},$$

(7.63)

where the canonical partition function that normalizes this is

$$Z(T) = \mathrm{TR}\prime \; \mathrm{e}^{-\beta\hat{\mathcal{H}}} = \sum_n \prime \mathrm{e}^{-\beta\mathcal{H}_n}.$$

(7.64)

In general, the partition function is the total number of states of the total system, and its logarithm gives the total entropy. The prime that has been added here signifies the restriction on the counting of states, so that forbidden states are not included, and so that the same state is not counted more than once. This restriction is essential, but it is usually not explicitly shown or averted to in the conventional expression for the partition function, equation (7.1).

The symmetrization of the wave function plays a rôle in the correct counting, as states that differ only by the permutation of the particle labels are not distinct. Further, states with repeated label values may be allowed for bosons but forbidden for fermions. The way to properly formulate the partition function to account for these effects is to define a symmetrization factor for the microstates.

7.4.1.1 Symmetrization factor

A basis wave function formed from the orthonormal set of unsymmetrized wave functions $\{\phi_{\mathbf{n}}\}$ has symmetrized form

$$\phi_{\mathbf{n}}^{\pm} = \frac{1}{\sqrt{N! \chi_{\mathbf{n}}^{\pm}}} \sum_{\hat{P}} (\pm 1)^p \phi_{\hat{P}\mathbf{n}}.$$

(7.65)

Here $\hat{P}$ is the permutation operator, p is its parity, the upper sign is for bosons, and the lower sign is for fermions. The symmetrization factor for the state, $\chi_{\mathbf{n}}^{\pm}$, is inversely proportional to the number of nonzero distinct permutations of the wave function. Specifically, normalization, $\langle \phi_{\mathbf{n}}^{\pm} \mid \phi_{\mathbf{n}}^{\pm} \rangle = 1$, gives

$$\chi_{\mathbf{n}}^{\pm} = \sum_{\hat{P}} (\pm 1)^p \langle \phi_{\mathbf{n}} \mid \phi_{\hat{P}\mathbf{n}} \rangle.$$

(7.66)

With the symmetrization factor, the sum over distinct, allowed states of some symmetric function $f_{\hat{P}\mathbf{n}} = f_{\mathbf{n}}$ can be written as a sum over all states,

$$\sum_{\mathbf{n}}{}' f_{\mathbf{n}} = \frac{1}{N!} \sum_{\mathbf{n}} \chi_{\mathbf{n}}^{\pm} f_{\mathbf{n}}.$$

(7.67)

7.4.1.2 Example: two particles in two states

As an example, consider a two-particle system, with symmetrization factor $\chi_{k_1 k_2}^{\pm}$, where k_1 is the state of the first particle and k_2 is that of the second. Suppose that there are just two one-particle states, $k_i \in \{1, 2\}$, $i = 1, 2$. One has

$$\chi_{11}^{\pm} = \langle \phi_{11} \mid \phi_{11} \rangle \pm \langle \phi_{11} \mid \phi_{11} \rangle$$
$$= \begin{cases} 2 \\ 0 \end{cases}$$
$$= \chi_{22}^{\pm}.$$

(7.68)

Since two fermions cannot be in the same state, the 11 and 22 states are forbidden for a fermionic system. Also,

$$\chi_{12}^{\pm} = \langle \phi_{12} \mid \phi_{12} \rangle \pm \langle \phi_{12} \mid \phi_{21} \rangle$$
$$= 1$$
$$= \chi_{21}^{\pm}.$$

(7.69)

The state 12 is the same as the state 21.

Hence the sum over distinct, allowed states of some symmetric function $f_{\hat{P}\mathbf{n}} = f_{\mathbf{n}}$ can be written as the sum over all states by using the symmetrization factor,

$$\sideset{}{'}\sum_{\mathbf{n}} f_{\mathbf{n}} = \begin{cases} f_{11} + f_{12} + f_{22} \\ f_{12} \end{cases}$$

$$= \begin{cases} f_{11} + \dfrac{1}{2}[f_{12} + f_{21}] + f_{22} \\ \dfrac{1}{2}[f_{12} + f_{21}] \end{cases}$$

$$= \frac{1}{2}\left\{ \chi_{11}^{\pm} f_{11} + \chi_{12}^{\pm} f_{12} + \chi_{21}^{\pm} f_{21} + \chi_{22}^{\pm} f_{22} \right\}$$

$$= \frac{1}{N!} \sum_{\mathbf{n}} \chi_{\mathbf{n}}^{\pm} f_{\mathbf{n}}.$$

(7.70)

The upper line is for bosons, and the lower line is for fermions. For fermions, the 11 and 22 states are excluded. For bosons the 11 and 22 states are each counted with weight 1. The $12 = 21$ state is counted once for both bosons and fermions.

7.4.1.3 Grand partition function

Let $\phi_{\mathbf{n}}$ be an entropy eigenfunction. In the canonical equilibrium case, entropy eigenfunctions are the same as energy eigenfunctions. With $\hat{\mathcal{H}}$ the Hamiltonian or energy operator, in the canonical equilibrium case the eigenvalue equation is $\hat{\mathcal{H}} \mid \phi_{\mathbf{n}}\rangle = \mathcal{H}_{\mathbf{n}} \mid \phi_{\mathbf{n}}\rangle$. Here the $\{\phi_{\mathbf{n}}\}$ form a complete orthonormal unsymmetrized set, and $\mathbf{n}$ labels entropy microstates.

The partition function is the total number of allowed, distinct states of the total system, which is the total number of reservoir-weighted allowed, distinct states of the sub-system. The emphasis is on the words allowed and distinct, because it would be wrong to count forbidden states, or to count the same state more than once. It is most convenient to work in the grand canonical system (constant chemical potential, variable particle number). With correct counting, the grand partition function is (Attard 2017, 2018)

$$\Xi^{\pm} = \sum_{N=0}^{\infty} z^{N} \sideset{}{'}\sum_{\mathbf{n}} e^{-\beta\mathcal{H}_{\mathbf{n}}}$$

$$= \sum_{N=0}^{\infty} \frac{z^{N}}{N!} \sum_{\mathbf{n}} \chi_{\mathbf{n}}^{\pm} e^{-\beta\mathcal{H}_{\mathbf{n}}}$$

$$= \sum_{N=0}^{\infty} \frac{z^{N}}{N!} \sum_{\mathbf{n}} \sum_{\hat{P}} (\pm 1)^{p} \langle \phi_{\hat{P}\mathbf{n}} \mid \phi_{\mathbf{n}}\rangle e^{-\beta\mathcal{H}_{\mathbf{n}}}$$

$$= \sum_{N=0}^{\infty} \frac{z^{N}}{N!} \sum_{\mathbf{n}} \sum_{\hat{P}} (\pm 1)^{p} \langle \phi_{\hat{P}\mathbf{n}} \mid e^{-\beta\hat{\mathcal{H}}} \mid \phi_{\mathbf{n}}\rangle.$$

(7.71)

Here N is the number of particles, the fugacity is $z \equiv e^{\beta\mu}$, where μ is the chemical potential, and $\beta = 1/k_{\mathrm{B}}T$ is sometimes called the inverse temperature, with k_{B} being

Boltzmann's constant and T the temperature.

The partition function (and statistical average) is usually written as the simple trace of the Maxwell–Boltzmann operator (Neumann 1932, Messiah 1961, Merzbacher 1970). Usually this does not explicitly exclude forbidden states or count duplicate states with correctly reduced weight. The present expression corrects this deficiency. It is dependent upon using entropy microstates and their symmetrization factor $\chi^{\pm}(\phi_{\mathbf{n}})$. If the partition function (and statistical average) is to be written as the trace of an operator, then the definition of trace has to be altered to explicitly count only distinct, allowed states.

The partition function can be usefully cast in an arbitrary basis. Consider another orthonormal, complete, unsymmetrized but otherwise at this stage arbitrary basis $\{\zeta_{\mathbf{p}}\}$. (Here the subscript is used as both an index and a label.) Since this is complete, $\sum_{\mathbf{p}} |\zeta_{\mathbf{p}}\rangle\langle\zeta_{\mathbf{p}}| = \hat{I}$, the identity operator in this form can be inserted to yield

$$\Xi^{\pm} = \sum_{N=0}^{\infty} \frac{z^N}{N!} \sum_{\mathbf{n}} \sum_{\mathbf{p}} \sum_{\hat{P}} (\pm 1)^p \langle \phi_{\hat{P}\mathbf{n}} | e^{-\beta\hat{\mathcal{H}}} | \zeta_{\mathbf{p}} \rangle \langle \zeta_{\mathbf{p}} | \phi_{\mathbf{n}} \rangle$$

$$= \sum_{N=0}^{\infty} \frac{z^N}{N!} \sum_{\mathbf{n}} \sum_{\mathbf{p}} \sum_{\hat{P}} (\pm 1)^p \langle \phi_{\mathbf{n}} | e^{-\beta\hat{\mathcal{H}}} | \zeta_{\hat{P}\mathbf{p}} \rangle \langle \zeta_{\mathbf{p}} | \phi_{\mathbf{n}} \rangle$$

$$= \sum_{N=0}^{\infty} \frac{z^N}{N!} \sum_{\mathbf{p}} \sum_{\hat{P}} (\pm 1)^p \langle \zeta_{\mathbf{p}} | e^{-\beta\hat{\mathcal{H}}} | \zeta_{\hat{P}\mathbf{p}} \rangle.$$

(7.72)

The second equality follows from the replacements $\mathbf{n} \Rightarrow \hat{P}\mathbf{n}$ and $\mathbf{p} \Rightarrow \hat{P}\mathbf{p}$, and the fact that $\langle \zeta_{\hat{P}\mathbf{p}} | \phi_{\hat{P}\mathbf{n}} \rangle = \langle \zeta_{\mathbf{p}} | \phi_{\mathbf{n}} \rangle$. The final equality has the same form as the final form of the sum over entropy states.

Finally, it will prove useful to express the partition function as a sum over a dual basis set. Consider another orthonormal, complete, unsymmetrized, but otherwise at this stage arbitrary basis $\{\zeta_{\mathbf{q}}\}$, with $\sum_{\mathbf{q}} |\zeta_{\mathbf{q}}\rangle\langle\zeta_{\mathbf{q}}| = \hat{I}$. One can write

$$\Xi^{\pm} = \sum_{N=0}^{\infty} \frac{z^N}{N!} \sum_{\mathbf{q},\mathbf{p}} \sum_{\hat{P}} (\pm 1)^p \langle \zeta_{\mathbf{p}} | \zeta_{\mathbf{q}} \rangle \langle \zeta_{\mathbf{q}} | e^{-\beta\hat{\mathcal{H}}} | \zeta_{\hat{P}\mathbf{p}} \rangle$$

$$= \sum_{N=0}^{\infty} \frac{z^N}{N!} \sum_{\mathbf{q},\mathbf{p}} \sum_{\hat{P}} (\pm 1)^p \langle \zeta_{\hat{P}\mathbf{p}} | \zeta_{\mathbf{q}} \rangle \langle \zeta_{\mathbf{q}} | e^{-\beta\hat{\mathcal{H}}} | \zeta_{\mathbf{p}} \rangle.$$

(7.73)

(One could replace here $\mathbf{q} \Rightarrow \mathbf{p}\prime$, in which case the final form is just the non-diagonal sum over a single basis set, analogous to the diagonal sum over the entropy basis set.)

7.4.2 Momentum and position states

7.4.2.1 Basis functions

Now the analysis is restricted to momentum and position basis functions. The eigenfunctions of the momentum operator in the position representation $\mathbf{r}$ are plane waves,

$$\zeta_{\mathbf{p}}(\mathbf{r}) = \frac{1}{V^{N/2}} e^{-\mathbf{p}\cdot\mathbf{r}/i\hbar}$$

$$= \prod_{j=1}^{N} \frac{e^{-\mathbf{p}_j\cdot\mathbf{r}_j/i\hbar}}{V^{1/2}}.$$

(7.74)

Here $\mathbf{p}$ is the configuration momentum, and V is the sub-system volume. These form a complete orthonormal set. Periodic boundary conditions give the width of the momentum state per particle per dimension as $\Delta_p = 2\pi\hbar/V^{1/3}$, assuming a cubic volume.

The position basis functions are Gaussians,

$$\zeta_{\mathbf{q}}(\mathbf{r}) = \frac{e^{-(\mathbf{r}-\mathbf{q})^2/4\xi^2}}{(2\pi\xi^2)^{3N/4}}$$

$$= \prod_{j=1}^{N} \frac{e^{-(\mathbf{r}_j-\mathbf{q}_j)^2/4\xi^2}}{(2\pi\xi^2)^{3/4}}.$$

(7.75)

Here $\mathbf{q}$ is the configuration position. Normalization fixes the spacing of the configuration position states as $\Delta_q = \sqrt{8\pi\xi^2}$, which gives the completeness expression, $\sum_{\mathbf{q}} |\zeta_{\mathbf{q}}(\mathbf{r'})\rangle \langle \zeta_{\mathbf{q}}(\mathbf{r})| = e^{-(\mathbf{r'}-\mathbf{r})^2/8\xi^2}/(8\pi\xi^2)^{3N/2} \equiv \delta_\xi(\mathbf{r'}-\mathbf{r})$. In the limit $\xi \to 0$ the position basis functions form a complete orthonormal set. This limit will be taken in the final results below.

Note the distinction between the representation position $\mathbf{r}$ and the configuration position $\mathbf{q}$.

The transformation coefficient for the two basis sets is readily shown to be

$$\langle \zeta_{\mathbf{p}} | \zeta_{\mathbf{q}} \rangle \equiv \frac{(8\pi\xi^2)^{3N/4}}{V^{N/2}} e^{-\xi^2 p^2/\hbar^2} e^{\mathbf{q}\cdot\mathbf{p}/i\hbar}$$

$$= \prod_{j=1}^{N} \frac{(8\pi\xi^2)^{3/4}}{V^{1/2}} e^{-\xi^2 p_j^2/\hbar^2} e^{\mathbf{q}_j\cdot\mathbf{p}_j/i\hbar}.$$

(7.76)

Again, N is the number of particles in the sub-system. It is important for the treatment of permutation loops below that this is the product of individual particle factors. The product of this and its complex conjugate is obviously

$$\langle \zeta_{\mathbf{p}} | \zeta_{\mathbf{q}} \rangle \langle \zeta_{\mathbf{q}} | \zeta_{\mathbf{p}} \rangle \equiv \frac{(8\pi\xi^2)^{3N/2}}{V^N} e^{-2\xi^2 p^2/\hbar^2},$$

(7.77)

which product will shortly appear in the summand of the partition function.

7.4.2.2 Commutation function

The position and momentum operators do not commute, which fundamental property can be accounted for by recasting the partition function to include what we shall call the commutation function $\omega(\mathbf{p},\mathbf{q}) \equiv e^{W(\mathbf{p},\mathbf{q})}$. We call ω the commutation function and W the commutation function exponent. These are defined by

$$e^{-\beta\mathcal{H}(\mathbf{p},\mathbf{q})}\omega(\mathbf{p},\mathbf{q}) = \frac{\langle\zeta_{\mathbf{q}}\mid e^{-\beta\hat{\mathcal{H}}}\mid\zeta_{\mathbf{p}}\rangle}{\langle\zeta_{\mathbf{q}}\mid\zeta_{\mathbf{p}}\rangle}.$$

<div align="right">(7.78)</div>

In Attard (2021 section 7.1.1) this ω was denoted ω_{p} and its complex conjugate was written ω_{q}.

The commutation function exponent, $W(\mathbf{q},\mathbf{p})$, is extensive. This is one reason that the formulation in terms of it is preferable to those given by Wigner (1932) and by Kirkwood (1933). A recursion relation leading to an expansion for $W(\mathbf{q},\mathbf{p})$ will be given in section 7.5.3. To leading order, when Planck's constant can be regarded as small, or else the temperature is high, the commutation function exponent vanishes, $W(\hbar = 0) = W(\beta = 0) = 0$, as can be seen by inspection.

7.4.2.3 Symmetrization function

The obvious merit of formulating the grand partition function in terms of the asymmetric expectation value of the Maxwell–Boltzmann operator is that the sum over entropy states has become a sum over points in classical phase space. The continuum limit of this is

$$\sum_{\mathbf{q},\mathbf{p}} \Rightarrow \frac{1}{(\Delta_p\Delta_q)^{3N}}\int d\mathbf{\Gamma}.$$

<div align="right">(7.79)</div>

The volume elements are $\Delta_p = 2\pi\hbar/V^{1/3}$ and $\Delta_q = \sqrt{8\pi\xi^2}$.

The present way of representing quantum mechanics in classical phase space is distinctly different to the way advocated by Wigner (1932) and by Kirkwood (1933). The Wigner–Kirkwood method has been followed and modified by many workers (for example, Groenewold 1946, Moyal 1949, Praxmeyer and Wódkiewicz 2002, Barnett and Radmore 2003, Gerry and Knight 2005, Zachos *et al* 2005, Dishlieva 2008, and Barker 2010). In particular, previous authors focus upon a transform of the wave function rather than the states themselves, they do not distinguish the position representation $\mathbf{r}$ from the position configuration $\mathbf{q}$ as here, and they do not account for wave function symmetrization. Finally, the Wigner–Kirkwood phase space representation has principally been applied or used in quantum optics, whereas the present focus is on the generic formulation of quantum statistical mechanics.

In view of the above, the partition function becomes

$$\begin{aligned}
\Xi^{\pm} &= \sum_{N=0}^{\infty}\frac{z^N}{N!}\sum_{\mathbf{q},\mathbf{p}}\sum_{\hat{\mathrm{P}}}(\pm 1)^p\langle\zeta_{\hat{\mathrm{P}}\mathbf{p}}\mid\zeta_{\mathbf{q}}\rangle\,\langle\zeta_{\mathbf{q}}\mid e^{-\beta\hat{\mathcal{H}}}\mid\zeta_{\mathbf{p}}\rangle \\
&= \sum_{N=0}^{\infty}\frac{z^N}{N!(\Delta_p\Delta_q)^{3N}}\int d\mathbf{\Gamma}\sum_{\hat{\mathrm{P}}}(\pm 1)^p\langle\zeta_{\hat{\mathrm{P}}\mathbf{p}}\mid\zeta_{\mathbf{q}}\rangle\langle\zeta_{\mathbf{q}}\mid\zeta_{\mathbf{p}}\rangle e^{-\beta\mathcal{H}(\mathbf{q},\mathbf{p})}\omega(\mathbf{q},\mathbf{p}) \\
&= \sum_{N=0}^{\infty}\frac{z^N}{N!(\Delta_p\Delta_q)^{3N}}\int d\mathbf{\Gamma}\,\eta^{\pm}(\mathbf{q},\mathbf{p})\langle\zeta_{\mathbf{p}}\mid\zeta_{\mathbf{q}}\rangle\langle\zeta_{\mathbf{q}}\mid\zeta_{\mathbf{p}}\rangle e^{-\beta\mathcal{H}(\mathbf{q},\mathbf{p})}\omega(\mathbf{q},\mathbf{p}) \\
&= \sum_{N=0}^{\infty}\frac{z^N}{N!h^{3N}}\int d\mathbf{\Gamma}\,e^{-\beta\mathcal{H}(\mathbf{\Gamma})}\omega(\mathbf{\Gamma})\eta^{\pm}(\mathbf{\Gamma}).
\end{aligned}$$

<div align="right">(7.80)</div>

Notice that the factors of ξ cancel, which allows the limit $\xi \to 0$ to be taken. Here

$$\eta^\pm(\mathbf{\Gamma}) \equiv \frac{1}{\langle \zeta_\mathbf{p} \mid \zeta_\mathbf{q} \rangle} \sum_{\hat{P}} (\pm 1)^p \langle \zeta_{\hat{P}\mathbf{p}} \mid \zeta_\mathbf{q} \rangle$$

(7.81)

is called the symmetrization function.

7.4.2.4 Statistical average
The statistical average of an operator is

$$
\begin{aligned}
\left\langle \hat{A} \right\rangle_{\text{stat}} &= \frac{1}{\Xi^\pm} \text{TR}\prime\, z^N \hat{A} \mathrm{e}^{-\beta\hat{\mathcal{H}}} \\
&= \frac{1}{\Xi^\pm} \sum_{N=0}^{\infty} z^N \sum_{\mathbf{n}}\prime\, A_\mathbf{n} \mathrm{e}^{-\beta\mathcal{H}_\mathbf{n}} \\
&= \frac{1}{\Xi^\pm} \sum_{N=0}^{\infty} \frac{z^N}{N!} \sum_{\mathbf{q},\mathbf{p}} \sum_{\hat{P}} (\pm 1)^p \langle \zeta_{\hat{P}\mathbf{p}} \mid \phi_\mathbf{q} \rangle\, \langle \zeta_\mathbf{q} \mid \hat{A}\mathrm{e}^{-\beta\hat{\mathcal{H}}} \mid \zeta_\mathbf{p} \rangle.
\end{aligned}
$$

(7.82)

One sees that compared to the grand partition function, the Maxwell–Boltzmann operator has simply been replaced by the product of it and the observable operator, $\mathrm{e}^{-\beta\hat{\mathcal{H}}} \Rightarrow \hat{A}\mathrm{e}^{-\beta\hat{\mathcal{H}}}$. Hence as for the commutation function itself, equation (7.78), the statistical average can be obtained as a phase space average of the phase functions

$$\mathrm{e}^{-\beta\mathcal{H}(\mathbf{p},\mathbf{q})} A(\mathbf{p},\mathbf{q}) \omega_\mathrm{A}(\mathbf{p},\mathbf{q}) = \frac{\langle \zeta_\mathbf{q} \mid \mathrm{e}^{-\beta\hat{\mathcal{H}}} \hat{A} \mid \zeta_\mathbf{p} \rangle}{\langle \zeta_\mathbf{q} \mid \zeta_\mathbf{p} \rangle}.$$

(7.83)

In Attard (2021 section 7.1.2) this was denoted $\omega_{\mathrm{A},\mathrm{p}}$ and its complex conjugate was written $\omega_{\mathrm{A},\mathrm{q}}$.

With this, one sees that the expression for the statistical average of an operator observable is essentially the same as that for the grand partition function with $\omega(\mathbf{\Gamma}) \Rightarrow \omega_\mathrm{A}(\mathbf{\Gamma})$. This is

$$\left\langle \hat{A} \right\rangle_{\text{stat}} = \frac{1}{\Xi^\pm} \sum_{N=0}^{\infty} \frac{z^N}{N! h^{3N}} \int \mathrm{d}\mathbf{\Gamma}\, \mathrm{e}^{-\beta\mathcal{H}(\mathbf{\Gamma})} \omega_\mathrm{A}(\mathbf{\Gamma}) \eta^\pm(\mathbf{\Gamma}) A(\mathbf{\Gamma}).$$

(7.84)

If the operator being averaged is 'pure' in that sense that it only depends upon position or it only depends momentum operator, then $\omega_\mathrm{A}(\mathbf{\Gamma}) = \omega(\mathbf{\Gamma})$ (Attard 2021 section 7.1.2). This also holds for any linear combination of pure operators, such as the Hamiltonian operator.

7.5 Expansions for non-commutativity and symmetrization

7.5.1 Monomer and classical grand potential

Shortly it will be shown that the symmetrization function, $\eta^\pm$, breaks up into loops of particles, $l = 1, 2, \ldots$, which may be called monomers, dimers etc. The monomer term is

unity, $\eta^{(1)} = 1$, and the monomer grand partition function is

$$\Xi_1 = \sum_{N=0}^{\infty} \frac{z^N}{N! h^{3N}} \int d\mathbf{\Gamma} \, e^{-\beta \mathcal{H}(\mathbf{\Gamma})} \omega(\mathbf{\Gamma}).$$

(7.85)

The ratio of the full partition function to the monomer partition function is just the monomer average of the symmetrization function,

$$\frac{\Xi^{\pm}}{\Xi_1} = \frac{1}{\Xi_1} \sum_{N=0}^{\infty} \frac{z^N}{N! h^{3N}} \int d\mathbf{\Gamma} \, e^{-\beta \mathcal{H}(\mathbf{\Gamma})} \omega(\mathbf{\Gamma}) \eta^{\pm}(\mathbf{\Gamma})$$
$$= \left\langle \eta^{\pm} \right\rangle_1.$$

(7.86)

Since the leading order of the commutation function exponent is zero, $W(\hbar = 0) = W(\beta = 0) = 0$ (see section 7.5.3 below), the leading order of the monomer grand potential is just the classical grand potential,

$$\Xi_{1,0} \equiv \Xi_1(W = 0) = \sum_{N=0}^{\infty} \frac{z^N}{N! h^{3N}} \int d\mathbf{\Gamma} \, e^{-\beta \mathcal{H}(\mathbf{\Gamma})}.$$

(7.87)

Recall that $\omega(\mathbf{\Gamma}) \equiv e^{W(\mathbf{\Gamma})}$.

The logarithm of this gives the classical equilibrium grand potential,

$$\Omega_{\text{cl}}(\mu, V, T) \equiv \Omega_{1,0} = -k_{\text{B}} T \ln \Xi_{1,0}.$$

(7.88)

With this result, classical statistical mechanics has been derived from quantum statistical mechanics.

The quantum correction from monomers is the classical average

$$e^{-\beta[\Omega_1 - \Omega_{1,0}]} = \frac{\Xi_1}{\Xi_{1,0}} = \left\langle e^W \right\rangle_{1,0},$$

(7.89)

where Ω_1 is the full monomer grand potential. The subscript 1,0 is synonymous with the classical equilibrium average and with the classical equilibrium thermodynamic potential.

7.5.2 Permutation loop expansion of the grand potential

In the treatment of wave function symmetrization, the basis functions can be composed from single-particle functions,
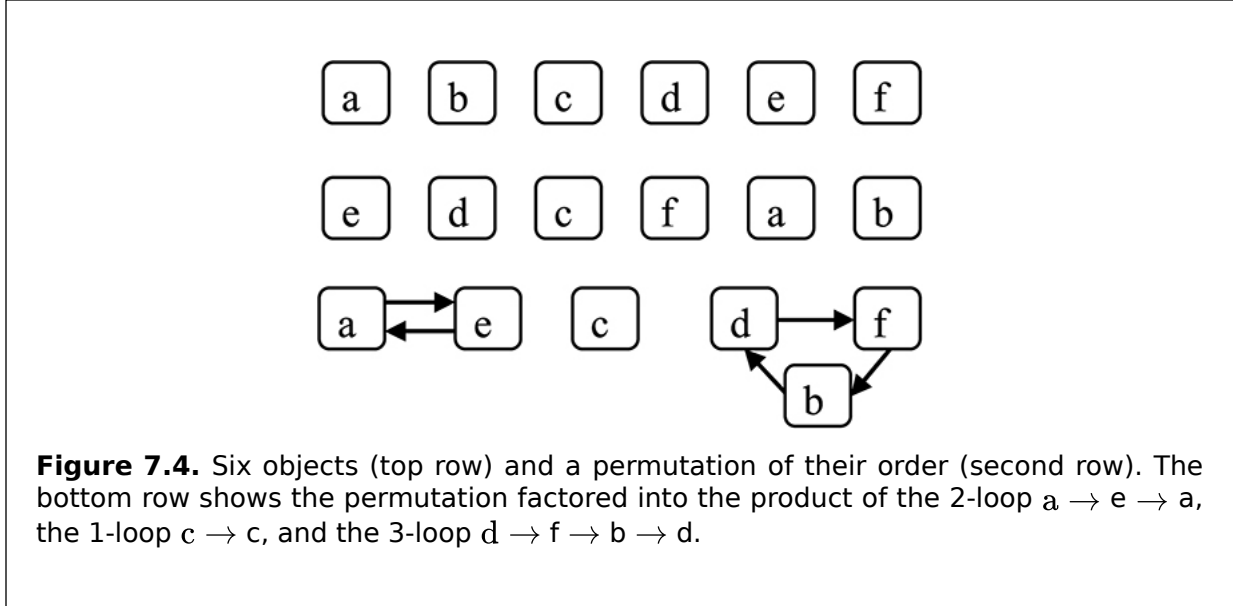
$$\zeta_{\mathbf{n}}(\mathbf{r}) = \zeta_{n_1}(\mathbf{r}_1) \zeta_{n_2}(\mathbf{r}_2) \cdots \zeta_{n_N}(\mathbf{r}_N).$$

(7.90)

This factorized form holds for the position and momentum basis functions, equations (7.74) and (7.75). The permutation operator can be applied to either the arguments, $\zeta_{\mathbf{n}}(\hat{\mathbf{P}}\mathbf{r})$ or else the indices $\zeta_{\hat{\mathbf{P}}\mathbf{n}}(\mathbf{r})$.

Here and in what follows for an N-particle system, bold face is used for an N-dimensional state vector such as $\mathbf{n} = \{n_1, n_2, \ldots, n_N\}$, and plain type is used for a single-particle

state such as *n* or n_j. Such a single-particle state may in fact be a three-dimensional vector, although $\mathbf{q}_j$ and $\mathbf{p}_j$ will be used to denote the position state and the momentum state of particle *j*.

In general, any permutation of objects can be factored as the product of permutation loops. In such a loop, each successive element is the one permuted into the position originally occupied by the preceding element (see figure 7.4).

Figure 7.4. Six objects (top row) and a permutation of their order (second row). The bottom row shows the permutation factored into the product of the 2-loop $a \rightarrow e \rightarrow a$, the 1-loop $c \rightarrow c$, and the 3-loop $d \rightarrow f \rightarrow b \rightarrow d$.

Accordingly, any particular particle permutation operator can be factored into loop permutation operators. Hence the sum over all permutation operators can be written as the sum over all possible factors of loop permutations,

$$\sum_{\hat{P}} (\pm 1)^p \, \hat{P} = \hat{I} \pm {\sum_{i,j}}' \, \hat{P}_{ij} + {\sum_{i,j,k}}' \, \hat{P}_{ij}\hat{P}_{jk} + {\sum_{i,j,k,l}}' \, \hat{P}_{ij}\hat{P}_{kl} \pm \cdots$$

(7.91)

Here $\hat{P}_{jk}$ is the transpose of particles *j* and *k*. The primes on the sums restrict them to unique loops, with each index being different. The first term is just the identity. The second term is a dimer loop, the third term is a trimer loop, and the fourth term shown is the product of two different dimers.

The symmetrization function, $\eta^{\pm}(\mathbf{\Gamma}) = \sum_{\hat{P}} (\pm 1)^p \, \langle \zeta_{\hat{P}\mathbf{p}} \mid \zeta_{\mathbf{q}} \rangle / \langle \zeta_{\mathbf{p}} \mid \zeta_{\mathbf{q}} \rangle$, is the sum of the expectation values of these loops.

The monomer symmetrization function comes from the unpermuted expectation value,

$$\eta^{(1)}(\mathbf{\Gamma}) \equiv \frac{\langle \zeta_{\mathbf{p}} \mid \zeta_{\mathbf{q}} \rangle}{\langle \zeta_{\mathbf{p}} \mid \zeta_{\mathbf{q}} \rangle} = 1.$$

(7.92)

The dimer symmetrization function in the microstate $\mathbf{\Gamma}$ for particles *j* and *k* is

$$\eta_{jk}^{\pm(2)}(\mathbf{\Gamma}) = \frac{\pm\langle\zeta_{\hat{P}_{jk}\mathbf{p}} \mid \zeta_{\mathbf{q}}\rangle}{\langle\zeta_{\mathbf{p}} \mid \zeta_{\mathbf{q}}\rangle}$$

$$= \frac{\pm\langle\zeta_{\mathbf{p}_k} \mid \zeta_{\mathbf{q}_j}\rangle\langle\zeta_{\mathbf{p}_j} \mid \zeta_{\mathbf{q}_k}\rangle}{\langle\zeta_{\mathbf{p}_j} \mid \zeta_{\mathbf{q}_j}\rangle\langle\zeta_{\mathbf{p}_k} \mid \zeta_{\mathbf{q}_k}\rangle}$$

$$= \pm\, e^{(\mathbf{q}_k-\mathbf{q}_j)\cdot\mathbf{p}_j/i\hbar}e^{(\mathbf{q}_j-\mathbf{q}_k)\cdot\mathbf{p}_k/i\hbar}.$$

(7.9
3)

Note that since the basis functions are the product of single-particle functions, equations (7.74) and (7.75), the expectation value factorizes leaving only the permuted particles to contribute.

The symmetrization functions are localized in the sense that they are only nonzero when the separations between consecutive neighbors around the loop are all small. This will be shown explicitly below (cf equation (7.122)), but here it can be noted that the exponents give highly oscillatory behavior and therefore cancel unless the differences in configuration positions are all close to zero.

Similarly, the trimer symmetrization function for particles *j*, *k*, and *l* is

$$\eta_{jkl}^{\pm(3)}(\mathbf{\Gamma}) = \frac{\langle\zeta_{\hat{P}_{jk}\hat{P}_{kl}\mathbf{p}} \mid \zeta_{\mathbf{q}}\rangle}{\langle\zeta_{\mathbf{p}} \mid \zeta_{\mathbf{q}}\rangle}$$

$$= \frac{\langle\zeta_{\mathbf{p}_k} \mid \zeta_{\mathbf{q}_j}\rangle\langle\zeta_{\mathbf{p}_j} \mid \zeta_{\mathbf{q}_l}\rangle\langle\zeta_{\mathbf{p}_l} \mid \zeta_{\mathbf{q}_k}\rangle}{\langle\zeta_{\mathbf{p}_j} \mid \zeta_{\mathbf{q}_j}\rangle\langle\zeta_{\mathbf{p}_k} \mid \zeta_{\mathbf{q}_k}\rangle\langle\zeta_{\mathbf{p}_l} \mid \zeta_{\mathbf{q}_l}\rangle}$$

$$= e^{(\mathbf{q}_j-\mathbf{q}_k)\cdot\mathbf{p}_k/i\hbar}e^{(\mathbf{q}_k-\mathbf{q}_l)\cdot\mathbf{p}_l/i\hbar}e^{(\mathbf{q}_l-\mathbf{q}_j)\cdot\mathbf{p}_j/i\hbar}.$$

(7.9
4)

Continuing in this fashion, the symmetrization function can be written as a series of loop products,

$$\eta^{\pm}(\mathbf{\Gamma}) = 1 + \sum_{ij}{}'\eta_{ij}^{\pm(2)}(\mathbf{\Gamma}) + \sum_{ijk}{}'\eta_{ijk}^{\pm(3)}(\mathbf{\Gamma}) + \sum_{ijkl}{}'\eta_{ij}^{\pm(2)}(\mathbf{\Gamma})\eta_{kl}^{\pm(2)}(\mathbf{\Gamma}) + \cdots$$

(7.9
5)

Here the superscript is the order of the loop, and the subscripts are the atoms involved in the loop.

This gives the ratio of the full to the monomer partition function as

$$\frac{\Xi^\pm}{\Xi_1} = \left\langle \eta^\pm \right\rangle_1$$

$$= 1 + \left\langle \sum_{ij}{}'\eta_{ij}^{\pm(2)} \right\rangle_1 + \left\langle \sum_{ijk}{}'\eta_{ijk}^{\pm(3)} \right\rangle_1 + \left\langle \sum_{ijkl}{}'\eta_{ij}^{\pm(2)}\eta_{kl}^{\pm(2)} \right\rangle_1 + \cdots$$

$$= 1 + \left\langle \frac{N!}{(N-2)!2}\eta^{\pm(2)} \right\rangle_1 + \left\langle \frac{N!}{(N-3)!3}\eta^{\pm(3)} \right\rangle_1$$

$$+ \frac{1}{2}\left\langle \frac{N!}{(N-2)!2}\eta^{\pm(2)} \right\rangle_1^2 + \cdots$$

$$= \sum_{\{m_l\}} \frac{1}{m_l!} \prod_{l=2}^{\infty} \left\langle \frac{N!}{(N-l)!l}\eta^{\pm(l)} \right\rangle_1^{m_l}$$

$$= \prod_{l=2}^{\infty} \sum_{m_l=0}^{\infty} \frac{1}{m_l!} \left\langle \frac{N!}{(N-l)!l}\eta^{\pm(l)} \right\rangle_1^{m_l}$$

$$= \prod_{l=2}^{\infty} \exp \left\langle \frac{N!}{(N-l)!l}\eta^{\pm(l)} \right\rangle_1 .$$

(7.96)

The third and following equalities write the average of the product as the product of the averages. This is valid in the thermodynamic limit, since the product of the average of two loops scales as V^2, whereas the correlated interaction of two loops scales as V. The combinatorial factor accounts for the number of unique loops in each term; $\eta^{\pm(l)}$ without subscripts refers to any one set of l particles, since all sets give the same average. Explicitly, the l-loop symmetrization function is

$$\eta^{\pm(l)}(\mathbf{\Gamma}^l) = (\pm 1)^{l-1} \prod_{j=1}^{l} e^{-(\mathbf{q}_j - \mathbf{q}_{j+1})\cdot\mathbf{p}_j/i\hbar}$$

$$\equiv (\pm 1)^{l-1}e^{-(\mathbf{q}-\mathbf{q}')\cdot\mathbf{p}/i\hbar}.$$

(7.97)

The subscripts for particles in a loop are to be understood $\mathrm{mod}\ l$; in the first equality, $\mathbf{q}_{l+1} \equiv \mathbf{q}_1$. In the second equality, $\mathbf{q} = \{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_l\}$ and $\mathbf{q}' = \{\mathbf{q}_l, \mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{l-1}\}$.

The grand potential is essentially the logarithm of the grand partition function, $\Omega \equiv -k_\mathrm{B}T \ln \Xi$. Hence the difference between the full grand potential and the monomer grand potential is just the series of loop potentials,

$$-\beta[\Omega^\pm - \Omega_1] = \ln \frac{\Xi^\pm}{\Xi_1}$$

$$= \sum_{l=2}^{\infty} \left\langle \frac{N!}{(N-l)!l}\eta^{\pm(l)} \right\rangle_1$$

$$\equiv -\beta \sum_{l=2}^{\infty} \Omega_l^\pm.$$

(7.98)

The monomer grand potential is of course $\Omega_1 \equiv -k_{\mathrm{B}}T \ln \Xi_1$, with the monomer grand partition function being given by equation (7.85).

7.5.3 Expansion of the commutation function

Now several high temperature expansions for the commutation function $w(\mathbf{\Gamma})$ are given. (See Attard (2021) for other expansions of, and approaches to, the commutation function.) The approach is similar to that of Kirkwood (1933). Whereas in essence he expanded the function that is here denoted $\omega = \mathrm{e}^W$, here is expanded the function W itself. This is an extensive function, which is an advantage.

The definition of the commutation function, equation (7.78), may be written as $\mathrm{e}^{-\beta\mathcal{H}(\mathbf{p},\mathbf{q})}\mathrm{e}^{W(\mathbf{p},\mathbf{q})}\langle \mathbf{q} \mid \mathbf{p}\rangle = \langle \mathbf{q} \mid \mathrm{e}^{-\beta\hat{\mathcal{H}}} \mid \mathbf{p}\rangle$. The Hamiltonian operator is the sum of kinetic energy and potential energy operators, $\hat{\mathcal{H}} = (-\hbar^2/2m)\nabla^2 + U(\mathbf{q})$, and the momentum eigenfunction is $\langle \mathbf{q} \mid \mathbf{p}\rangle = \mathrm{e}^{-\mathbf{p}\cdot\mathbf{q}/\mathrm{i}\hbar}/V^{N/2}$. Suppressing the arguments the derivative of the left-hand side is

$$\frac{\partial}{\partial\beta}\left\{\mathrm{e}^{-\beta\mathcal{H}}\mathrm{e}^W\mathrm{e}^{-\mathbf{p}\cdot\mathbf{q}/\mathrm{i}\hbar}\right\} = \left\{\frac{\partial W}{\partial\beta} - \mathcal{H}\right\}\mathrm{e}^{-\beta\mathcal{H}}\mathrm{e}^W\mathrm{e}^{-\mathbf{p}\cdot\mathbf{q}/\mathrm{i}\hbar}.$$

(7.99)

That of the right-hand side is

$$\frac{\partial}{\partial\beta}\langle \mathbf{q} \mid \mathrm{e}^{-\beta\hat{\mathcal{H}}} \mid \mathbf{p}\rangle = -\hat{\mathcal{H}}\left\{\mathrm{e}^{-\beta\mathcal{H}}\mathrm{e}^W\mathrm{e}^{-\mathbf{p}\cdot\mathbf{q}/\mathrm{i}\hbar}\right\}$$

$$= \left\{-\mathcal{H}[\mathrm{e}^{-\beta\mathcal{H}}\mathrm{e}^W] + \frac{\hbar^2}{2m}\nabla^2[\mathrm{e}^{-\beta\mathcal{H}}\mathrm{e}^W] + 2\frac{\mathrm{i}\hbar}{2m}\mathbf{p}\cdot\nabla[\mathrm{e}^{-\beta\mathcal{H}}\mathrm{e}^W]\right\}\mathrm{e}^{-\mathbf{p}\cdot\mathbf{q}/}$$

(7.100)

Equating these and rearranging gives

$$\frac{\partial W}{\partial\beta} = \frac{\mathrm{i}\hbar}{m}\mathrm{e}^{\beta U - W}\mathbf{p}\cdot\nabla\left\{\mathrm{e}^{W-\beta U}\right\} + \frac{\hbar^2}{2m}\mathrm{e}^{\beta U - W}\nabla^2\left\{\mathrm{e}^{W-\beta U}\right\}$$

$$= \frac{\mathrm{i}\hbar}{m}\mathbf{p}\cdot\nabla(W - \beta U) + \frac{\hbar^2}{2m}\left\{\nabla(W - \beta U)\cdot\nabla(W - \beta U) + \nabla^2(W - \beta U)\right\}.$$

(7.101)

The commutation function exponent expanded in powers of Planck's constant is

$$W \equiv \sum_{n=1}^{\infty} W_n\hbar^n.$$

(7.102)

(Actually an expansion in powers of inverse temperature is slightly simpler.) This begins at $n = 1$ because it must reduce to the classical Maxwell–Boltzmann factor when Planck's constant is zero, $W(\hbar = 0) = 0$. This expansion leads to the recursion relation for $n > 2$,

$$\frac{\partial W_n}{\partial \beta} = \frac{\mathrm{i}}{m}\mathbf{p}\cdot\nabla W_{n-1} + \frac{1}{2m}\sum_{j=0}^{n-2}\nabla W_{n-2-j}\cdot\nabla W_j$$

$$- \frac{\beta}{m}\nabla W_{n-2}\cdot\nabla U + \frac{1}{2m}\nabla^2 W_{n-2}.$$

(7.103)

It is straightforward to derive the first several coefficient functions explicitly. One has for $n = 1$,

$$W_1 = \frac{-\mathrm{i}\beta^2}{2m}\mathbf{p}\cdot\nabla U,$$

(7.104)

for $n = 2$,

$$W_2 = \frac{\beta^3}{6m^2}\mathbf{pp}:\nabla\nabla U + \frac{1}{2m}\left\{\frac{\beta^3}{3}\nabla U\cdot\nabla U - \frac{\beta^2}{2}\nabla^2 U\right\},$$

(7.105)

for $n = 3$,

$$W_3 = \frac{\mathrm{i}\beta^4}{24m^3}\mathbf{ppp}\vdots\nabla\nabla\nabla U + \frac{5\mathrm{i}\beta^4}{24m^2}\mathbf{p}(\nabla U):\nabla\nabla U - \frac{\mathrm{i}\beta^3}{6m^2}\mathbf{p}\cdot\nabla\nabla^2 U,$$

(7.106)

and for $n = 4$,

$$W_4 = \frac{-\beta^5}{5!m^4}(\mathbf{p}\cdot\nabla)^4 U - \frac{3\beta^5}{40m^3}(\nabla U)\mathbf{pp}\vdots\nabla\nabla\nabla U$$

$$- \frac{\beta^5}{15m^2}(\nabla U)(\nabla U):\nabla\nabla U + \frac{\beta^4}{16m^2}\nabla U\cdot\nabla\nabla^2 U$$

$$+ \frac{\beta^4}{16m^3}\mathbf{pp}:\nabla\nabla\nabla^2 U + \frac{\beta^4}{48m^2}\nabla^2(\nabla U\cdot\nabla U)$$

$$- \frac{\beta^3}{24m^2}\nabla^2\nabla^2 U - \frac{\beta^5}{15m^3}(\mathbf{p}\cdot\nabla\nabla U)\cdot(\mathbf{p}\cdot\nabla\nabla U).$$

(7.107)

Notice that the odd coefficient functions are pure imaginary and odd in momentum. Because $\mathcal{H}(\mathbf{\Gamma})$ is an even function of momentum, these terms in the quantum weight e^W average out to real oscillatory (cosine) contributions. (The $\eta^{\pm(l)}$ contain terms that are either real and even, or else imaginary and odd in momentum.)

7.5.3.1 Monomer expansion A

The quantum correction to the classical grand potential due to the monomers is just a classical average of the quantum weight due to commutativity, equation (7.89). To fourth order in $\hbar$ this is

$$-\beta[\Omega_1 - \Omega_{1,0}] = \ln \left\langle e^W \right\rangle_{1,0}$$

$$= \ln \left\langle 1 + W + \frac{1}{2!}W^2 + \frac{1}{3!}W^3 + \frac{1}{4!}W^4 \right\rangle_{1,0}$$

$$= \left\langle W + \frac{1}{2}W^2 + \frac{1}{3!}W^3 + \frac{1}{4!}W^4 \right\rangle_{1,0} - \frac{1}{2}\left\langle W + \frac{1}{2}W^2 \right\rangle_{1,0}^2$$

$$= \hbar^2 \langle W_2 \rangle_{1,0} + \frac{\hbar^2}{2}\left\langle W_1^2 \right\rangle_{1,0}$$

$$+ \frac{\hbar^4}{2}\left\langle W_2^2 - \langle W_2 \rangle_{1,0}^2 \right\rangle_{1,0} + \frac{\hbar^4}{4!}\left\langle W_1^4 \right\rangle_{1,0} - \frac{\hbar^4}{8}\left\langle W_1^2 \right\rangle_{1,0}^2$$

$$+ \hbar^4 \langle W_4 \rangle_{1,0} + \hbar^4 \langle W_1 W_3 \rangle_{1,0} + \frac{\hbar^4}{2}\left\langle W_1^2 W_2 \right\rangle_{1,0} - \frac{\hbar^4}{2}\left\langle W_1^2 \right\rangle_{1,0}\langle W_2 \rangle_{1,}$$

All higher order contributions have been set to zero, and the odd imaginary terms have explicitly canceled here. It is $W_j(\mathbf{p}, \mathbf{q})$ whose average is taken. Recall that the subscript 1,0 denotes the classical average over position and momentum configurations. Note that the right-hand side must be extensive, which requires the cancelation of products of extensive terms. This expansion terminated at $\mathcal{O}(\hbar^2)$ may be called A2, and at $\mathcal{O}(\hbar^4)$ it may be called A4.

7.5.3.2 Monomer expansion B

One can define the cumulative weight as

$$W^{(n)} = \hbar W_1 + \hbar^2 W_2 + \cdots + \hbar^n W_n,$$

and the *n*th approximation to the quantum correction to the monomer grand potential as

$$\Delta\Omega_1^{(n)} = -k_B T \ln \left\langle e^{W^{(n)}} \right\rangle_{1,0}.$$

One has $\lim_{n\to\infty} \Delta\Omega_1^{(n)} = \Omega_1 - \Omega_{1,0}$.

The classical average $\langle \cdots \rangle_{1,0}$ includes an average over the momenta $\langle \cdots \rangle_{1,0,p}$ as well as one over the position configurations. The classical monomer probability distribution for momentum is a Gaussian,

$$\wp(\mathbf{p}) = \frac{e^{-\beta p^2/2m}}{[2\pi m k_B T]^{3N/2}}.$$

Hence $\langle \mathbf{p}\mathbf{p} \rangle_{1,0,p} = m k_B T \, \underline{\underline{\mathbf{I}}}$.

Because the probability distribution over momenta is a Gaussian, it is straightforward, if perhaps a little tedious, to perform the momentum integrals analytically. To a particular order one has

$$\left\langle e^{W^{(n)}(\mathbf{q},\mathbf{p})} \right\rangle_{1,0,p} \equiv e^{W^{(n)}(\mathbf{q})}.$$

One can show (details are given in Attard (2017); the same method is used in section 7.5.5) that to fourth order the result is

$$
\begin{aligned}
W^{(4)}(\mathbf{q}) = {} & \frac{\hbar^2 \beta^3}{24m} \nabla U \cdot \nabla U - \frac{\hbar^2 \beta^2}{12m} \nabla^2 U \\
& + \frac{2\hbar^4 \beta^4}{45m^2} (\nabla\nabla U):(\nabla\nabla U) - \frac{\hbar^4 \beta^5}{240m^2} \nabla U \nabla U : \nabla\nabla U \\
& + \frac{\hbar^4 \beta^4}{20m^2} (\nabla U) \cdot \nabla\nabla^2 U - \frac{11\hbar^4 \beta^3}{240m^2} \nabla^2\nabla^2 U + \mathcal{O}(\hbar^6) \\
\equiv {} & \hbar^2 \tilde{W}_2(\mathbf{q}) + \hbar^4 \tilde{W}_4(\mathbf{q}).
\end{aligned}
$$

(7.113)

This is the exponent that one has to use to weight the position configurations (in addition to the Boltzmann factor of the potential energy). This expansion terminated at $\mathcal{O}(\hbar^2)$ may be called B2, and at $\mathcal{O}(\hbar^4)$ it may be called B4.

7.5.3.3 Monomer expansion C

Expansion A has the merit of requiring less analysis, and of requiring the classical average of extensive terms and their products. It has the additional advantage of being explicitly dependent upon the momenta, which means it can be used for higher order loops (dimer, trimer, etc). It has the disadvantage of requiring the average over the momentum configurations to be taken numerically. Also, as a fluctuation expression, it requires relatively high accuracy in the individual averages of products to get the necessary cancelation between these super-extensive terms to end up with the final extensive result for $\Omega_1 - \Omega_{1,0}$. The larger the system size, the greater the number of configurations that need to be generated to get acceptable statistical accuracy.

Expansion B does not require numerical momentum averaging, which leads to higher accuracy. In a computer simulation, this can make it 1–2 orders of magnitude more efficient than expansion A. Also, exponentiating the expansion, as in B, should give faster convergence than expanding the final expression, as in A. A disadvantage expansion B has is that it requires more explicit algebra. Also, more problematic, since W is an extensive variable, taking the average of $\mathrm{e}^{W^{(n)}}$ can lead to computational overflow problems for large systems.

One way to avoid numerical overflow in evaluating the exponent in expansion B, but to preserve the advantage of analytical momentum average, is to use the momentum averaged $W^{(4)}$ in the expansion A. That is, writing equation (7.113) as explicit powers of Planck's constant, $W^{(4)} \equiv \hbar^2 \tilde{W}_2 + \hbar^4 \tilde{W}_4$, this can be inserted into expansion A, equation (7.108), with $W_1 = W_3 = 0$,
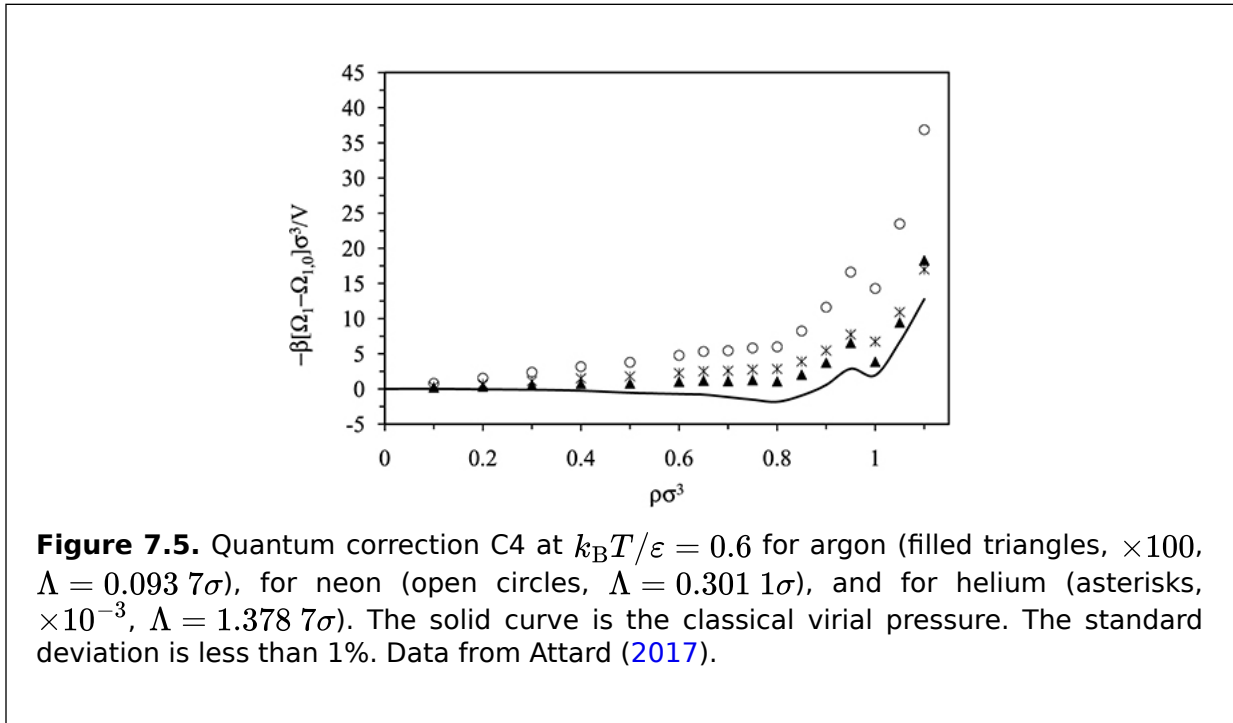
$$
\begin{aligned}
-\beta[\Omega_1 - \Omega_{1,0}] &= \ln \left\langle \mathrm{e}^{W^{(4)}} \right\rangle_{1,0} \\
&= \hbar^2 \left\langle \tilde{W}_2 \right\rangle_{1,0} + \frac{\hbar^4}{2} \left\langle \tilde{W}_2^2 - \left\langle \tilde{W}_2 \right\rangle_{1,0}^2 \right\rangle_{1,0} + \hbar^4 \left\langle \tilde{W}_4 \right\rangle_{1,0}.
\end{aligned}
$$

(7.114)

This may be called expansion C4. The averages here are classical averages over configuration positions. The neglected terms are $\mathcal{O}(\hbar^6)$. The difference from expansion A4 is that the momentum averages have been performed analytically to leading order before expanding the exponential.

7.5.3.4 Results for Lennard-Jones Ar, Ne, and He

Classical Monte Carlo simulations for Lennard-Jones models of argon, neon, and helium have been carried to test the three expansions given above for the change in the monomer grand potential due to non-commutativity, $\Omega_1 - \Omega_{1,0}$ (see Attard (2017) for full details). The Lennard-Jones pair potential is $u(r) = 4\varepsilon[(\sigma/r)^{12} - (\sigma/r)^6]$, and the thermal wave length is $\Lambda = [2\pi\hbar^2/mk_{\mathrm{B}}T]^{1/2}$. Comparing this difference with the classical virial pressure, $p_{\mathrm{cl}} = -\Omega_{1,0}/V$, measures the quantum correction due to non-commutativity, since symmetrization does not affect the monomer term. Basically, the Metropolis algorithm was used to generate the position configurations for the averages, and, for expansion A, momentum configurations (typically 32 for each position configuration used for the average) were drawn directly from the Gaussian distribution. There was generally quite good agreement between the three expansions, although A had the worst statistical error. There was good agreement between A2 and the Wigner–Kirkwood second order expansion (Wigner 1932, Kirkwood 1933).

Results on a typical sub-critical isotherm are shown in figure 7.5. Regions where the pressure has negative slope are unstable in the thermodynamic limit. From the classical virial pressure, there is a gas phase for $\rho\sigma^3 \lesssim 0.1$, a liquid phase $0.8 \lesssim \rho\sigma^3 \lesssim 0.95$, and a solid phase $\rho\sigma^3 \gtrsim 1$. Inclusion of the quantum correction may shift these phase boundaries. The C4 quantum correction is qualitatively similar for all three noble elements, with it being positive and mainly increasing with increasing density. Compared to the classical virial pressure, the quantum correction is substantially less for argon, somewhat larger for neon, and very much larger for helium. Judging by the change between the second and the fourth order expansions, the fourth order quantum correction calculated here as an estimate of the total quantum correction appears reliable for argon, marginal for neon, and unreliable for helium.

Figure 7.5. Quantum correction C4 at $k_{\mathrm{B}}T/\varepsilon = 0.6$ for argon (filled triangles, $\times 100$, $\Lambda = 0.093\ 7\sigma$), for neon (open circles, $\Lambda = 0.301\ 1\sigma$), and for helium (asterisks, $\times 10^{-3}$, $\Lambda = 1.378\ 7\sigma$). The solid curve is the classical virial pressure. The standard deviation is less than 1%. Data from Attard (2017).

7.5.4 Expansion of the symmetrization function

The difference between the full grand potential and the monomer grand potential may be expanded in terms of loop potentials

$$\Omega^{\pm} - \Omega_1 = \sum_{l=2}^{\infty} \Omega_l^{\pm}.$$

(7.115)

The loop potentials arise from factoring the permutations involved in symmetrizing the wave function, with the *l*-loop potential being given by equation (7.98),

$$
\begin{aligned}
-\beta \Omega_l^{\pm} &\equiv \left\langle \frac{N!}{(N-l)!\,l} \eta^{\pm(l)} \right\rangle_1 \\
&= \frac{1}{\Xi_1} \sum_{N=0}^{\infty} \frac{z^N}{N!\,h^{3N}} \int d\mathbf{\Gamma} \; e^{-\beta \mathcal{H}(\mathbf{\Gamma})} e^{W(\mathbf{\Gamma})} \frac{N!}{(N-l)!\,l} \eta^{\pm(l)}(\mathbf{\Gamma}).
\end{aligned}
$$

(7.116)

The monomer grand partition function is given by equation (7.85),

$$\Xi_1 = \sum_{N=0}^{\infty} \frac{z^N}{N!\,h^{3N}} \int d\mathbf{\Gamma} \; e^{-\beta \mathcal{H}(\mathbf{\Gamma})} e^{W(\mathbf{\Gamma})}.$$

(7.117)

The monomer grand potential is of course $\Omega_1 \equiv -k_B T \ln \Xi_1$.

7.5.4.1 Commuting part of the expansion

In the preceding sub-section, three expansions for the weight $W(\mathbf{\Gamma})$ were discussed. Of particular interest is what might be called the commuting part of the expansion, in which case one sets $W(\mathbf{\Gamma}) = 0$. This can be denoted by an additional subscript 0, so that the commuting part of the loop potential is

$$
\begin{aligned}
-\beta \Omega_{l,0}^{\pm} &\equiv \left\langle \frac{N!}{(N-l)!\,l} \eta^{\pm(l)} \right\rangle_{1,0} \\
&= \frac{1}{\Xi_{1,0}} \sum_{N=0}^{\infty} \frac{z^N}{N!\,h^{3N}} \int d\mathbf{\Gamma} \; e^{-\beta \mathcal{H}(\mathbf{\Gamma})} \frac{N!}{(N-l)!\,l} \eta^{\pm(l)}(\mathbf{\Gamma}).
\end{aligned}
$$

(7.118)

Since this *l*-loop symmetrization factor depends only on the first *l* particles, its classical equilibrium average can be written as an integral over their configuration momenta and positions. The standard definition in classical equilibrium statistical mechanics of the *l*-particle density is (Attard 2002)

$$
\begin{aligned}
\rho^{(l)}(\mathbf{q}^l) &= \frac{1}{\Xi_{1,0}} \sum_{N=l}^{\infty} \frac{\Lambda^{-3N} z^N}{(N-l)!} \int d\mathbf{q}_{l+1} \cdots d\mathbf{q}_N \; e^{-\beta U(\mathbf{q}^N)} \\
&= \frac{1}{\Xi_{1,0}} \sum_{N=l}^{\infty} \frac{z^N \Lambda^{-3l} e^{\beta \mathcal{K}(\mathbf{p}^l)}}{(N-l)!\,h^{3(N-l)}} \int d\mathbf{\Gamma}_{l+1} \cdots d\mathbf{\Gamma}_N \; e^{-\beta \mathcal{H}(\mathbf{\Gamma}^N)}.
\end{aligned}
$$

(7.119)

Recall that the thermal wave length is $\Lambda = [2\pi\hbar^2/mk_{\mathrm{B}}T]^{1/2}$. With this the commuting part of the l-loop grand potential can be written

$$
\begin{aligned}
-\beta\Omega_{l,0}^{\pm} &= \frac{1}{\Xi_{1,0}} \sum_{N=l}^{\infty} \frac{z^N}{l(N-l)!h^{3N}} \int \mathrm{d}\mathbf{\Gamma}\ \mathrm{e}^{-\beta\mathcal{H}(\mathbf{\Gamma})} \eta^{\pm(l)}(\mathbf{\Gamma}) \\
&= \frac{\Lambda^{3l}}{lh^{3l}} \int \mathrm{d}\mathbf{\Gamma}^l\ \mathrm{e}^{-\beta\mathcal{K}(\mathbf{p}^l)} \rho^{(l)}(\mathbf{q}^l) \eta^{\pm(l)}(\mathbf{\Gamma}) \\
&= \frac{(\pm1)^{l-1}\Lambda^{3l}}{lh^{3l}} \int \mathrm{d}\mathbf{\Gamma}^l\ \mathrm{e}^{-\beta\mathcal{K}(\mathbf{p}^l)} \rho^{(l)}(\mathbf{q}^l) \mathrm{e}^{\mathbf{p}\cdot(\mathbf{q}-\mathbf{q}\prime)/\mathrm{i}\hbar}.
\end{aligned}
$$
(7.120)

The final equality uses the l-loop symmetrization function, equation (7.97). In this $\mathbf{q} = \{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_l\}$ and $\mathbf{q}\prime = \{\mathbf{q}_l, \mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{l-1}\}$.

The part of the exponent that depends upon the momenta is

$$
\frac{-\Lambda^2 p^2}{4\pi\hbar^2} + \frac{\mathbf{p}\cdot(\mathbf{q}-\mathbf{q}\prime)}{\mathrm{i}\hbar} = \frac{-\Lambda^2}{4\pi\hbar^2}\left[\mathbf{p} - \frac{2\pi\hbar^2}{\Lambda^2\mathrm{i}\hbar}(\mathbf{q}-\mathbf{q}\prime)\right]^2 - \frac{\pi}{\Lambda^2}(\mathbf{q}-\mathbf{q}\prime)^2.
$$
(7.121)

Hence the momentum contributes just a Gaussian integral, which is readily evaluated, reducing the loop potential to

$$
\begin{aligned}
-\beta\Omega_{l,0}^{\pm} &= \frac{(\pm1)^{l-1}\Lambda^{3l}}{lh^{3l}} (4\pi^2\hbar^2/\Lambda^2)^{3l/2} \int \mathrm{d}\mathbf{q}^l\ \mathrm{e}^{-\pi(\mathbf{q}-\mathbf{q}\prime)^2/\Lambda^2} \rho^{(l)}(\mathbf{q}) \\
&= \frac{(\pm1)^{l-1}}{l} \int \mathrm{d}\mathbf{q}^l\ \mathrm{e}^{-\pi(\mathbf{q}-\mathbf{q}\prime)^2/\Lambda^2} \rho^{(l)}(\mathbf{q}) \\
&= \frac{(\pm1)^{l-1}V}{l} \int \mathrm{d}\mathbf{q}^{l-1}\ \mathrm{e}^{-\pi(\mathbf{q}-\mathbf{q}\prime)^2/\Lambda^2} \rho^{(l)}(\mathbf{q}), \quad \mathbf{q}_l = 0.
\end{aligned}
$$
(7.122)

Since this is homogeneous in space, in the final equality particle l has been fixed at the origin, and a factor of V has replaced the integration over this coordinate.

The thermal wavelength $\Lambda = [2\pi\hbar^2/mk_{\mathrm{B}}T]^{1/2}$ provides the length scale for the symmetrization Gaussian. If any nearest neighbors around the loop are separated by much more than this, then the Gaussian is zero and the configuration does not contribute to the loop potential. Therefore, one needs $q_{j,j+1} \lesssim \Lambda$. Conversely, it is generally the case that particles have finite size, say σ, and at low and moderate densities and the l-particle density is zero if any two particles are separated by less than this. Therefore, at low and moderate densities one needs

$$
\Lambda \gtrsim q_{j,j+1} \gtrsim \sigma
$$
(7.123)

for symmetrization effects to be measurable. For the noble elements on the isotherm $k_{\mathrm{B}}T/\varepsilon = 0.6$, in figure 7.5, this suggests that symmetrization is negligible for argon and neon.

In fact of course there is never such a thing as a perfectly impenetrable and rigid hard core; σ is really a measure of the onset of a steep inter-particle repulsion. Hence if the density is high, then the separation between nearest neighbors is approximately the cube

root of the particle volume, $\rho^{-1/3}$. In this case, for symmetrization effects to be measurable, one requires

$$\rho\Lambda^3 \geq 1.$$

By this criterion, only helium is affected by wave function symmetrization over the density range of figure 7.5. But even in this case one suspects that the effects of non-commutativity, which are $\mathcal{O}(10^4)$ times the classical pressure, quantitatively dominate those of particle symmetrization, since $\rho\Lambda^3_{\mathrm{He}} \lesssim 3$ at this temperature ($T^*_{\mathrm{LJ}} = 0.6$ corresponds to $T \approx 6.1$ K for He). However, the qualitative effects of symmetrization in liquid helium-4, such as superfluidity, can be remarkable (see chapters 8 and 9).

It is worth pointing out that the Gaussian exponent can be written as a quadratic form. With $\mathbf{q}_l = 0$, and $\alpha = x, y, z$, one simply has

$$\begin{aligned}
(\mathbf{q} - \mathbf{q}\prime)^2_\alpha &= q^2_{1;\alpha} + q^2_{12;\alpha} + q^2_{23;\alpha} + \cdots + q^2_{l-2,l-1;\alpha} + q^2_{l-1;\alpha} \\
&= \underline{\underline{A}}^{(l-1)} : \underline{q}^{l-1}_\alpha \underline{q}^{l-1}_\alpha,
\end{aligned}$$

where $q^2_{jk,\alpha} = (q_{j,\alpha} - q_{k,\alpha})^2$. Here $\underline{\underline{A}}^{(l-1)}$ is an $(l-1) \times (l-1)$ tridiagonal matrix with 2 on the main diagonal and -1 immediately above and below the main diagonal, and all other entries 0. It is readily shown that this has determinant

$$\begin{aligned}
|\underline{\underline{A}}^{(l-1)}| &= 2|\underline{\underline{A}}^{(l-2)}| - |\underline{\underline{A}}^{(l-3)}| \\
&= l.
\end{aligned}$$

This result will be used in section 7.6.

7.5.5 Second order non-commuting dimer contribution

For the case of a dimer loop, $l = 2$, the symmetrization function, equation (7.93), is

$$\begin{aligned}
\frac{N!}{(N-2)!2}\eta^{\pm(2)}(\mathbf{\Gamma}^N) &= \pm \sum_{j<k}^{N} e^{-\mathbf{q}_{jk} \cdot \mathbf{p}_{jk}/i\hbar} \\
&= \frac{\pm N(N-1)}{2} e^{-\mathbf{q}_{12} \cdot \mathbf{p}_{12}/i\hbar},
\end{aligned}$$

since the averaging makes all particle pairs equivalent. From the general expression, equation (7.98), the dimer loop potential is

$$\begin{aligned}
-\beta\Omega^\pm_2 &\equiv \left\langle \frac{N!}{(N-2)!2}\eta^{\pm(2)} \right\rangle_1 \\
&= \frac{1}{\Xi^\pm_1} \sum_{N=0}^{\infty} \frac{\pm z^N}{2h^{3N}(N-2)!} \int d\mathbf{\Gamma}^N \, e^{-\beta\mathcal{H}(\mathbf{\Gamma})} e^{W(\mathbf{\Gamma})} e^{-\mathbf{q}_{12} \cdot \mathbf{p}_{12}/i\hbar}.
\end{aligned}$$

Note that the integrand is highly oscillatory unless $\mathbf{q}_1 \approx \mathbf{q}_2$. Hence the integral vanishes unless the two particles are close together, which means that the integral has effectively

lost a factor of volume compared to $\Xi_1^{\pm}$. Conversely, the denominator of $(N-2)!$ is a factor of N^2 smaller than the denominator $N!$ of $\Xi_1^{\pm}$. Hence in total $\Omega_2^{\pm} \sim \mathcal{O}(V^{-1}N^2)$, which is extensive, as it must be.

The second order approximation for non-commutativity, using equations (7.104) and (7.105), is

$$
\begin{aligned}
W^{(2)} &= \hbar W_1 + \hbar^2 W_2 \\
&= \frac{-i\hbar\beta^2}{2m}\mathbf{p}\cdot\nabla U + \frac{\hbar^2\beta^3}{6m^2}\mathbf{pp}:\nabla\nabla U + \frac{\hbar^2}{2m}\left\{\frac{\beta^3}{3}\nabla U\cdot\nabla U - \frac{\beta^2}{2}\nabla^2 U\right\} \\
&\equiv W_p^{(2)} + W_q^{(2)}.
\end{aligned}
$$

(7.129)

The two momenta terms have here been separated out.

In the integrand of the loop potential, it is simplest to deal with the Gaussian exponent for the momenta directly by completing the squares. Write $\mathbf{U}\prime \equiv \nabla U$, $\mathbf{U}\prime\prime \equiv \nabla\nabla U$, and

$$
\{\tilde{\mathbf{U}}\prime\}_{j\alpha} \equiv \nabla_{j\alpha}U - \frac{2m}{\hbar^2\beta^2}q_{12;\alpha}\{\delta_{j,1}-\delta_{j,2}\}.
$$

(7.130)

The final term here comes from the exponent of the dimer loop symmetrization function. With these, that part of the exponent in the integrand that depends upon the momenta is

$$
\begin{aligned}
e_p &\equiv \frac{-\beta}{2m}\mathbf{p}\cdot\mathbf{p} - \frac{i\hbar\beta^2}{2m}\mathbf{p}\cdot\nabla U + \frac{\hbar^2\beta^3}{6m^2}\mathbf{pp}:\nabla\nabla U - \frac{\mathbf{q}_{12}\cdot\mathbf{p}_{12}}{i\hbar} \\
&= \frac{-\beta}{2m}\mathbf{p}\cdot\mathbf{p} - \frac{i\hbar\beta^2}{2m}\mathbf{p}\cdot\tilde{\mathbf{U}}\prime + \frac{\hbar^2\beta^3}{6m^2}\mathbf{pp}:\mathbf{U}\prime\prime \\
&= \frac{-\beta}{2m}\mathbf{A}:\left[\mathbf{p}+\frac{2m}{2\beta}\frac{i\hbar\beta^2}{2m}\mathbf{A}^{-1}\tilde{\mathbf{U}}\prime\right]^2 - \frac{\hbar^2\beta^3}{8m}\mathbf{A}^{-1}:\tilde{\mathbf{U}}\prime\tilde{\mathbf{U}}\prime,
\end{aligned}
$$

(7.131)

where $\mathbf{A} \equiv \mathbf{I} - \hbar^2\beta^2\mathbf{U}\prime\prime/3m$. The final equality is correct to $\mathcal{O}(\hbar^2)$. Clearly,

$$
\mathbf{A}^{-1} = \mathbf{I} + \frac{\hbar^2\beta^2}{3m}\mathbf{U}\prime\prime + \mathcal{O}(\hbar^4).
$$

(7.132)

Since the sum of the squares of the eigenvalues is the trace of the square of the matrix, one has

$$
\begin{aligned}
\ln|\mathbf{A}|^{1/2} &= \frac{1}{2}\sum_j \ln[1+\lambda_j] \\
&= \frac{1}{2}\sum_j[\lambda_j - \frac{1}{2}\lambda_j^2 + \cdots] \\
&= \frac{-\hbar^2\beta^2}{6m}\nabla^2 U - \frac{1}{4}\frac{\hbar^4\beta^4}{9m^2}(\nabla\nabla U):(\nabla\nabla U).
\end{aligned}
$$

(7.133)

With these the integral over the momenta for the dimer loop grand potential with second order expansion for the commutation factor yields

$$
\begin{aligned}
-\beta\Omega_{2,2}^{\pm} &= \frac{1}{\Xi_{1,2}^{\pm}} \sum_{N=0}^{\infty} \frac{\pm z^N/2h^{3N}}{(N-2)!} \int \mathrm{d}\mathbf{\Gamma}^N \; \mathrm{e}^{-\beta\mathcal{H}(\mathbf{\Gamma})}\mathrm{e}^{W^{(2)}(\mathbf{\Gamma})}\mathrm{e}^{-\mathbf{q}_{12}\cdot\mathbf{p}_{12}/\mathrm{i}\hbar} \\
&= \frac{1}{\Xi_{1,2}^{\pm}} \sum_{N=0}^{\infty} \frac{\pm z^N/2h^{3N}}{(N-2)!} \int \mathrm{d}\mathbf{q}^N \; \mathrm{e}^{-\beta U(\mathbf{q})}\mathrm{e}^{W_q^{(2)}(\mathbf{q})} \\
&\quad \times \left|2\pi m k_{\mathrm{B}} T \mathbf{A}(\mathbf{q})^{-1}\right|^{1/2}\mathrm{e}^{(-\hbar^2\beta^3/8m)\mathbf{A}(\mathbf{q})^{-1}:\tilde{\mathbf{U}}\prime(\mathbf{q})\tilde{\mathbf{U}}\prime(\mathbf{q})} \\
&= \frac{1}{\Xi_{1,2}^{\pm}} \sum_{N=0}^{\infty} \frac{\pm z^N \Lambda^{-3N}}{2(N-2)!} \int \mathrm{d}\mathbf{q}^N \; \mathrm{e}^{-\beta U(\mathbf{q})}\mathrm{e}^{W_q^{(2)}(\mathbf{q})} \\
&\quad \times \mathrm{e}^{(\hbar^2\beta^2/6m)\nabla^2 U(\mathbf{q})+(\hbar^4\beta^4/36m^2)(\nabla\nabla U(\mathbf{q})):(\nabla\nabla U(\mathbf{q}))} \\
&\quad \times \mathrm{e}^{(-\hbar^2\beta^3/8m)\mathbf{A}(\mathbf{q})^{-1}:\tilde{\mathbf{U}}\prime(\mathbf{q})\tilde{\mathbf{U}}\prime(\mathbf{q})}.
\end{aligned}
$$

(7.134)

Arguably, only terms to $\mathcal{O}(\hbar^2)$ should be retained in the exponent. The second order approximation to the monomer grand partition function, the denominator $\Xi_{1,2}^{\pm}$, is identical to this with the replacements $\tilde{\mathbf{U}}\prime(\mathbf{q}) \Rightarrow \mathbf{U}\prime(\mathbf{q})$ and $2(N-2)! \Rightarrow N!$. The full result may be called 'the' $\Omega_{2,2}^{\pm}$ term, although of course there is more than one way of expanding the commutation weight. As mentioned above, $\Omega_{2,2}^{\pm}$ is extensive.

7.6 Quantum ideal gas

For an ideal gas the potential energy is zero, $U(\mathbf{r}) = 0$, and the classical Hamiltonian is equal to the classical kinetic energy, $\mathcal{H}^{\mathrm{id}}(\mathbf{r}, \mathbf{p}) = \mathcal{K}(\mathbf{p}) = p^2/2m$. In this case the defining equation for the commutation function, equation (7.78), becomes

$$
\mathrm{e}^{-\beta\mathcal{K}(\mathbf{p})}\omega^{\mathrm{id}}(\mathbf{q}, \mathbf{p}) = \frac{\langle \zeta_{\mathbf{q}} \mid \mathrm{e}^{-\beta\hat{\mathcal{K}}} \mid \zeta_{\mathbf{p}} \rangle}{\langle \zeta_{\mathbf{q}} \mid \zeta_{\mathbf{p}} \rangle} = \mathrm{e}^{-\beta\mathcal{K}(\mathbf{p})}.
$$

(7.135)

Obviously, for the ideal gas the commutation function is unity, and the commutation function exponent vanishes,

$$
\omega^{\mathrm{id}}(\mathbf{q}, \mathbf{p}) = 1, \quad \text{and} \quad W^{\mathrm{id}}(\mathbf{q}, \mathbf{p}) = 0.
$$

(7.136)

This is to be expected since for the ideal gas the momentum eigenfunctions, $\zeta_{\mathbf{p}}(\mathbf{r})$, are also energy eigenfunctions. In view of this result, one need only retain the commuting part of the expansion, section 7.5.4.1, $\Omega_l^{\pm,\mathrm{id}} \equiv \Omega_{l,0}^{\pm,\mathrm{id}}$.

The commuting part of the l-loop grand potential, equation (7.122), requires the classical l-particle density. For the case of the ideal gas this is (Attard 2002)

$$
\rho^{(l),\mathrm{id}}(\mathbf{q}) = \Lambda^{-3l}z^l.
$$

(7.137)

This assumes a homogeneous system.

With this and using equation (7.126), the l-loop grand potential is

$$
\begin{aligned}
-\beta\Omega_l^{\pm,\mathrm{id}} &= \frac{(\pm1)^{l-1}V\Lambda^{-3l}z^l}{l}\int \mathrm{d}\mathbf{q}^{l-1}\; \mathrm{e}^{-\pi(\mathbf{q}-\mathbf{q}')^2/\Lambda^2}\\
&= \frac{(\pm1)^{l-1}V\Lambda^{-3l}z^l}{l}\left(\frac{2\pi\Lambda^2}{2\pi}\right)^{3(l-1)/2}\left|\underline{\underline{A}}^{(l-1)}\right|^{-3/2}\\
&= \frac{(\pm1)^{l-1}\Lambda^{-3}Vz^l}{l^{5/2}}.
\end{aligned}
$$

(7.1
38)

The upper sign is for bosons, and the lower for fermions. This holds for $l \geqslant 2$. The monomer case, $l = 1$, is the classical case, and direct calculation shows that $-\beta\Omega_1^{\mathrm{id}} = zV/\Lambda^3$. This is just this expression for $-\beta\Omega_l^{\pm,\mathrm{id}}$ with $l = 1$.

The thermodynamic relation between the pressure and the grand potential is given by equation (2.66), $p = -\Omega/V$. The classical ideal gas pressure is $p^{\mathrm{cl,id}} = zk_\mathrm{B}T/\Lambda^3 = -\Omega_1^{\mathrm{id}}/V$. With these and the above result, the pressure of the quantum ideal gas is given by

$$
\begin{aligned}
\beta p^{\pm,\mathrm{id}}(z,T)\Lambda^3 &= \frac{-\beta\Lambda^3}{V}\sum_{l=1}^{\infty}\Omega_l^{\pm,\mathrm{id}}\\
&= \sum_{l=1}^{\infty}(\pm1)^{l-1}z^l l^{-5/2}.
\end{aligned}
$$

(7.1
39)

This is the known result (Pathria 1972).

The relationship between density and fugacity for the quantum ideal gas is readily obtained from the thermodynamic result, equation (2.61),

$$
\bar{N}(z,V,T) = \frac{-\partial\Omega(z,V,T)}{\partial\mu}.
$$

(7.1
40)

Applying this to the loop expansion gives

$$
\begin{aligned}
\rho^{\pm,\mathrm{id}}(z,T) &= \frac{1}{\beta}\frac{\partial\beta p^{\pm,\mathrm{id}}(z,T)}{\partial z}\frac{\partial z}{\partial\mu}\\
&= \Lambda^{-3}\sum_{l=1}^{\infty}(\pm1)^{l-1}z^l l^{-3/2}.
\end{aligned}
$$

(7.1
41)

7.7 Permutation loop expansion

The following treatment of the loop expansion closely follows section 3.4.3 of Attard (2021). It is an example of the exact permutation loop expansion for general single-particle states. The loop expansion is applied to a system of independent harmonic oscillators in section 7.7.4, which follows section 4.2 of Attard (2021). The permutation loop expansion plays an essential role in the treatment of Bose–Einstein condensation, superfluidity, and superconductivity in subsequent chapters.

7.7.1 Loop expansion for single-particle states

In section 7.5, a permutation loop expansion of the symmetrization function contribution to the grand potential was given. This used momentum single-particle states. An analogous expansion holds for non-interacting particles in terms of single-particle energy states (see section 3.4.3 of Attard (2021)).

The sum over states for the grand partition function, whose logarithm is the grand potential, can be made unrestricted, equation (7.71), by invoking the symmetrization factor (Attard 2021 section 3.2), which requires the sum over all permutations of the quantum states, equation (7.66). We now obtain a systematic series expansion for the symmetrization factor and hence the grand potential using single-particle energy states.

For N particles, the unsymmetrized normalized sub-system basis functions for single-particle states are

$$\phi_{\mathbf{n}}(\mathbf{r}) = \prod_{j=1}^{N} \phi_{n_j}^{(1)}(\mathbf{r}_j), \quad \left\langle n_j \mid n_j' \right\rangle = \delta_{n_j, n_j'}.$$

(7.142)

Recall that for the quantum state, a bold face label indicates the system state, $\mathbf{n} = \{n_1, n_2, \ldots, n_N\}$, and a plain type indicates a single-particle state, n or n_j. Such single-particle state labels may be multi-dimensional.

The symmetrization factor, equation (7.66), is

$$\chi_{\mathbf{n}}^{\pm} = \sum_{\hat{P}} (\pm 1)^p \langle \phi_{\mathbf{n}}(\hat{P}\mathbf{r}) \mid \phi_{\mathbf{n}}(\mathbf{r}) \rangle.$$

(7.143)

The sum over all particle permutators is systematically decomposed into sums over permutation loops and their products, equation (7.91), and the symmetrization factor is the sum of the expectation values of these loops. The monomer symmetrization factor is unity,

$$\chi_{\mathbf{n}}^{\pm,(1)} = \langle \phi_{\mathbf{n}}(\mathbf{r}) \mid \phi_{\mathbf{n}}(\mathbf{r}) \rangle = 1.$$

(7.144)

The dimer symmetrization factor for particles j and k in the state $\mathbf{n}$ is

$$\begin{aligned}
\chi_{\mathbf{n};jk}^{\pm,(2)} &= \pm \langle \phi_{\mathbf{n}}(\hat{P}_{jk}\mathbf{r}) \mid \phi_{\mathbf{n}}(\mathbf{r}) \rangle \\
&= \pm \langle \phi_{n_j}^{(1)}(\mathbf{r}_k) \mid \phi_{n_k}^{(1)}(\mathbf{r}_k) \rangle \langle \phi_{n_k}^{(1)}(\mathbf{r}_j) \mid \phi_{n_j}^{(1)}(\mathbf{r}_j) \rangle \\
&= \pm \langle n_j \mid n_k \rangle \langle n_k \mid n_j \rangle.
\end{aligned}$$

(7.145)

This dimer symmetrization factor for the permutation of particles j and k only involves their respective single-particle wave functions.

Similarly the trimer symmetrization factor is

$$\begin{aligned}
\chi_{\mathbf{n};jkl}^{\pm,(3)} &= \langle \phi_{\mathbf{n}}(\hat{P}_{jk}\hat{P}_{kl}\mathbf{r}) \mid \phi_{\mathbf{n}}(\mathbf{r}) \rangle \\
&= \langle \phi_{n_j,n_k,n_l}^{(3)}(\mathbf{r}_l, \mathbf{r}_j, \mathbf{r}_k) \mid \phi_{n_j,n_k,n_l}^{(3)}(\mathbf{r}_j, \mathbf{r}_k, \mathbf{r}_l) \rangle \\
&= \langle \phi_{n_k}^{(1)}(\mathbf{r}_j) \mid \phi_{n_j}^{(1)}(\mathbf{r}_j) \rangle \langle \phi_{n_j}^{(1)}(\mathbf{r}_l) \mid \phi_{n_l}^{(1)}(\mathbf{r}_l) \rangle \langle \phi_{n_l}^{(1)}(\mathbf{r}_k) \mid \phi_{n_k}^{(1)}(\mathbf{r}_k) \rangle \\
&= \langle n_k \mid n_j \rangle \langle n_j \mid n_l \rangle \langle n_l \mid n_k \rangle.
\end{aligned}$$

(7.146)

Again since the expectation values of the unpermuted monomers each contribute a factor of unity, only the permuted particles and their states contribute to this.

In general for single-particle states the *l*-loop symmetrization factor is

$$\chi^{\pm,(l)}_{\mathbf{n};j_1 j_2 \ldots j_l} = (\pm 1)^{l-1} \langle n_{j_1} \mid n_{j_2} \rangle \langle n_{j_2} \mid n_{j_3} \rangle \cdots \langle n_{j_l} \mid n_{j_1} \rangle$$
$$= (\pm 1)^{l-1} \delta_{n_{j_1}, n_{j_2}} \, \delta_{n_{j_2}, n_{j_3}} \cdots \delta_{n_{j_l}, n_{j_1}}.$$

(7.147)

The product of Kronecker-deltas means that this is only nonzero if all the particles in the permutation loop occupy the same single-particle state. This result is exact, correct, and necessary, even for fermions. It is only the sum over all permutations (i.e. over all permutation loops and products thereof) that must come to zero if more than one fermion is in the same single-particle state. In section 7.7.4 it will be shown by explicit calculation that this expression gives the known textbook result for independent harmonic oscillators for both bosons and fermions.

Since the particle states are independent, disjoint permutation loops factorize into the product of their corresponding expectation values. Hence the factor for the product of dimer loops reduces to the product of dimer symmetrization factors,

$$\chi^{\pm,(2,2)}_{\mathbf{n};ij,kl} = \langle \phi_{\mathbf{n}}(\hat{P}_{ij}\hat{P}_{kl}\mathbf{r}) \mid \phi_{\mathbf{n}}(\mathbf{r}) \rangle$$
$$= \langle \phi_{\mathbf{n}}(\hat{P}_{ij}\mathbf{r}) \mid \phi_{\mathbf{n}}(\mathbf{r}) \rangle \langle \phi_{\mathbf{n}}(\hat{P}_{kl}\mathbf{r}) \mid \phi_{\mathbf{n}}(\mathbf{r}) \rangle$$
$$= \chi^{\pm,(2)}_{\mathbf{n};ij} \, \chi^{\pm,(2)}_{\mathbf{n};kl}.$$

(7.148)

By definition of distinct permutations, the $i, j, k,$ and l must all be different, in which case this factorization is exact. By this same argument, a similar factorization holds for the symmetrization factors of all products of permutation loops.

Because the symmetrization factor $\chi_{\mathbf{n}}$ is the sum over all permutations, it can be rewritten as the sum over all possible monomers and loops. This gives the loop expansion for the symmetrization factor for use in the partition function as

$$\chi^{\pm}_{\mathbf{n}} = 1 + \sum_{ij}{}' \chi^{\pm,(2)}_{\mathbf{n};ij} + \sum_{ijk}{}' \chi^{\pm,(3)}_{\mathbf{n};ijk} + \sum_{ijkl}{}' \chi^{\pm,(2)}_{\mathbf{n};ij} \chi^{\pm,(2)}_{\mathbf{n};kl} + \cdots$$

(7.149)

Note that the parity factor for fermions and bosons, $(\pm 1)^{l-1}$, has been incorporated into the definition of the $\chi^{\pm,(l)}_{\mathbf{n}}$. This is analogous to the result for the symmetrization function for classical phase space, equation (7.95). Accordingly, the factorization of the grand partition function when written in terms of the present single-particle energy states is identical.

The grand partition function is

$$\Xi^{\pm}(\mu, V, T) = \sum_{N=0}^{\infty} z^N Z^{\pm}(N, V, T) = \sum_{N=0}^{\infty} \frac{z^N}{N!} \sum_{\mathbf{n}} \chi^{\pm}_{\mathbf{n}} e^{-\beta \mathcal{H}_{\mathbf{n}}},$$

(7.150)

where the fugacity is $z = e^{\beta\mu}$, μ being the chemical potential, and the total energy eigenvalue is $\mathcal{H}_{\mathbf{n}} = \sum_{j=1}^{N} \epsilon_{n_j}$, with the single-particle energy eigenvalue being written $\epsilon_{n_j} \equiv \mathcal{H}^{(1)}_{n_j}$.

Analogous to the derivation for momentum states in section 7.5, the loop expansion and resummation gives the grand partition function using single-particle energy states as

$$\Xi^{\pm}(\mu, V, T) = \Xi_1 + (-\beta\Omega_2^{\pm})\Xi_1 + (-\beta\Omega_3^{\pm})\Xi_1 + \frac{1}{2}(-\beta\Omega_2^{\pm})^2\Xi_1 + \cdots$$

$$= \Xi_1 \sum_{\{m_l\}} \prod_{l=2}^{\infty} \frac{1}{m_l!}(-\beta\Omega_l^{\pm})^{m_l}$$

$$= \Xi_1 \prod_{l=2}^{\infty} \sum_{m_l=0}^{\infty} \frac{1}{m_l!}(-\beta\Omega_l^{\pm})^{m_l}$$

$$= \Xi_1 \prod_{l=2}^{\infty} e^{-\beta\Omega_l^{\pm}},$$

(7.1
51)

where m_l is the number of loops of l particles, and the loop grand potential for $l \geqslant 2$ is given by

$$-\beta\Omega_l^{\pm} = \frac{z^l}{l} \sum_{n_1,\ldots,n_l} e^{-\beta\mathcal{H}_{n_1\ldots n_l}^{(l)}} \chi_{n_1\ldots n_l}^{\pm,(l)}$$

$$= \frac{(\pm 1)^{l-1}z^l}{l} \sum_{n_1,\ldots,n_l} e^{-\beta\mathcal{H}_{n_l}^{(1)}} \delta_{n_l,n_1} \prod_{j=1}^{l-1} \left[e^{-\beta\mathcal{H}_{n_j}^{(1)}} \delta_{n_j,n_{j+1}} \right]$$

$$= \frac{(\pm 1)^{l-1}z^l}{l} \sum_{\epsilon} e^{-l\beta\epsilon}$$

(7.1
52)

This formula also holds for the monomer grand potential, $l = 1$.

Notice how equation (7.147) for the loop symmetrization factor collapses the l sums over all loop particles' single-particle states to a single sum over the single-particle states, since the loop symmetrization factor is only nonzero when all the loop particles are in the same single-particle state. This makes the loop expansion very efficient. If one includes M single-particle states and L loop potentials, then there are just $L \times M$ terms in the series for the grand potential. The contribution of successive loops decays exponentially with l, so L need not be large. Contrast this with the direct calculation of the partition function for $\bar{N}$ particles, which would require on the order of $\bar{N}^M$ terms.

The grand potential is given by the logarithm of the grand partition function, $\Omega^{\pm}(\mu, V, T) = -k_{\mathrm{B}}T \ln \Xi^{\pm}(\mu, V, T)$, which is just the sum of the loop potentials,

$$-\beta\Omega^{\pm}(\mu, V, T) = -\beta \sum_{l=1}^{\infty} \Omega_l^{\pm}.$$

$$= \sum_{l=1}^{\infty} \frac{(\pm 1)^{l-1}z^l}{l} \sum_{\epsilon} e^{-l\beta\epsilon}$$

$$= \mp \sum_{\epsilon} \ln\left[1 \mp ze^{-\beta\epsilon}\right].$$

(7.1
53)

Even with permutation symmetrization, this reduces to a sum over single-particle states, as one might expect for independent, non-interacting particles.

In the case that the particles can move throughout the volume of the sub-system V, the loop potentials are extensive. That is, they scale with the volume of the sub-system (assuming no externally applied potential). Hence one can define the loop grand potential

density, $\Omega_l^\pm(\mu, V, T)/V$. The reason that the loop potentials are extensive is that the center of mass of the *l*-mer is free to roam throughout the homogeneous volume *V*. But the particles of the loop must remain in the vicinity of the center of mass because of the connectivity of the loop. It is essential that the loop potentials are extensive because the grand potential of the quantum system has to be extensive.

7.7.2 Occupancy of single-particle energy states

The fugacity derivative of the grand potential gives the average or most likely number, equation (2.61)

$$\bar{N} = \frac{z\partial\bigl(-\beta\Omega^\pm(\mu, V, T)\bigr)}{\partial z} = \sum_\epsilon \frac{ze^{-\beta\epsilon}}{1 \mp ze^{-\beta\epsilon}}.$$

(7.154)

From this one sees that the average occupancy of a single-particle state is

$$\langle N_\epsilon \rangle^\pm = \frac{ze^{-\beta\epsilon}}{1 \mp ze^{-\beta\epsilon}}.$$

(7.155)

For fermions, the lower sign, this is less than unity, and for bosons, the upper sign, it diverges for the ground energy state, $\epsilon = 0$, as $z \to 1^-$.

Since the grand partition function for single-particle energy states can be written as a sum over occupancies,

$$\Xi^\pm = \sum_{N=0}^\infty z^N \sum_{\mathbf{N}} {}'\prod_\epsilon e^{-\beta N_\epsilon \epsilon},$$

(7.156)

one sees that the average occupancy can also be obtained from the energy derivative,

$$\langle N_\epsilon \rangle^\pm = -\beta^{-1}\frac{\partial\Xi^\pm}{\Xi^\pm\partial\epsilon} = -\beta^{-1}\frac{\partial(-\beta\Omega^\pm)}{\partial\epsilon}.$$

(7.157)

Similarly, one sees that the fluctuation squared is

$$\langle N_\epsilon^2 \rangle^\pm - \bigl(\langle N_\epsilon \rangle^\pm\bigr)^2 = -\beta^{-1}\frac{\partial}{\partial\epsilon}\langle N_\epsilon \rangle^\pm$$

$$= \frac{ze^{-\beta\epsilon}}{\bigl[1 \mp ze^{-\beta\epsilon}\bigr]^2}.$$

(7.158)

Hence the relative fluctuation in occupancy is given by

$$\frac{\langle N_\epsilon^2 \rangle^\pm - \bigl(\langle N_\epsilon \rangle^\pm\bigr)^2}{\bigl(\langle N_\epsilon \rangle^\pm\bigr)^2} = z^{-1}e^{\beta\epsilon}$$

$$= \frac{1}{\langle N_\epsilon \rangle^\pm} \pm 1.$$

(7.159)

For classical statistics the second term on the right-hand side of the final equality would be missing, in which case one sees that the relative fluctuations decrease with increasing average occupancy. For fermions, the lower sign, the fluctuations are less than the classical case. For the fully occupied fermion case at low temperatures, $\langle N_\epsilon \rangle^- \to 1$, the fluctuations vanish. For bosons, the upper sign, a state with high average occupancy, $\langle N_\epsilon \rangle^+ \gg 1$, has relative fluctuation of order unity. This is quite different to the classical case, and it means that the state might sometimes be occupied by many times the average number of bosons. Equally, a state highly-occupied on average has non-negligible probability of being few-occupied or even empty. This latter observation will prove of some significance in the discussion of the Bose–Einstein condensation in chapter 8, and of superfluidity in chapter 9.

7.7.3 Effective pair potential for ideal particles

Following section 3.4.1.3 of Attard (2021), one can analyze the effect of symmetrization on the position density for an ideal sub-system. The complete orthogonal set of position eigenfunctions are Dirac-δ functions,

$$| \mathbf{q} \rangle = \delta(\mathbf{r} - \mathbf{q}),$$

(7.160)

and the symmetrized position basis functions are

$$| \mathbf{q}^\pm \rangle = \frac{1}{\sqrt{N!}} \sum_{\hat{\mathrm{P}}} (\pm 1)^p \delta(\hat{\mathrm{P}}\mathbf{r} - \mathbf{q}).$$

(7.161)

This sets the symmetrization factor to unity, $\chi_{\mathbf{q}}^\pm = 1$, which gives the correct orthonormalization,

$$
\begin{aligned}
\langle \mathbf{q'}^\pm \mid \mathbf{q''}^\pm \rangle &= \sum_{\hat{\mathrm{P}}', \hat{\mathrm{P}}''} (\pm 1)^{p'+p''} \frac{1}{N!} \int d\mathbf{r} \ \delta(\hat{\mathrm{P}}'\mathbf{r} - \mathbf{q'})\delta(\hat{\mathrm{P}}''\mathbf{r} - \mathbf{q''}) \\
&= \frac{1}{N!} \sum_{\hat{\mathrm{P}}', \hat{\mathrm{P}}''} (\pm 1)^{p'+p''} \delta(\hat{\mathrm{P}}'\mathbf{q'} - \hat{\mathrm{P}}''\mathbf{q''}) \\
&= \sum_{\hat{\mathrm{P}}} (\pm 1)^p \delta(\hat{\mathrm{P}}\mathbf{q'} - \mathbf{q''}).
\end{aligned}
$$

(7.162)

The right-hand side is zero unless $\mathbf{q''}$ is a permutation of $\mathbf{q'}$. In the latter case, if, for example, $\hat{\mathrm{P}}_0 \mathbf{q'} = \mathbf{q''}$, then one term on the right-hand side is nonzero, namely $(\pm 1)^{p_0} \delta(\mathbf{q''} - \mathbf{q''})$. (This assumes that at any instant not more than one particle can occupy a given position.) This is the correct normalization for the continuum.

The quantum probability density for the superposition of position, neglecting proportionality constants, is

$$W_N(\mathbf{q}'', \mathbf{q}') \propto \langle \mathbf{q}''^{\pm}| \, \mathrm{e}^{-\beta \hat{\mathcal{H}}} \, | \, \mathbf{q}'^{\pm} \rangle$$

$$= \sum_{\mathbf{p}} \chi_{\mathbf{p}}^{\pm} \langle \mathbf{q}''^{\pm}| \, \mathrm{e}^{-\beta \hat{\mathcal{H}}} \, | \, \phi_{\mathbf{p}}^{\pm} \rangle \, \langle \phi_{\mathbf{p}}^{\pm} \, | \, \mathbf{q}'^{\pm} \rangle$$

$$= \sum_{\mathbf{p}} \chi_{\mathbf{p}}^{\pm} \mathrm{e}^{-\beta p^2/2m} \langle \mathbf{q}''^{\pm}| \, \phi_{\mathbf{p}}^{\pm} \rangle \, \langle \phi_{\mathbf{p}}^{\pm} \, | \, \mathbf{q}'^{\pm} \rangle.$$

(7.1
63)

Here $\phi_{\mathbf{p}}^{\pm}(\mathbf{r})$ is a symmetrized momentum eigenfunction, and $\chi_{\mathbf{p}}^{\pm}$ is the associated symmetrization factor that ensures the correct normalization and completeness condition. The momentum eigenfunctions are also energy eigenfunctions of the ideal gas Hamiltonian operator. Now

$$\langle \mathbf{q}''^{\pm}| \, \phi_{\mathbf{p}}^{\pm} \rangle = \frac{1}{N!\sqrt{\chi_{\mathbf{p}}^{\pm}}} \sum_{\hat{\mathrm{P}}', \hat{\mathrm{P}}''} (\pm 1)^{p'+p''} \langle \delta(\hat{\mathrm{P}}''\mathbf{r} - \mathbf{q}'') \, | \, \phi_{\mathbf{p}}(\hat{\mathrm{P}}'\mathbf{r}) \rangle$$

$$= \frac{1}{\sqrt{\chi_{\mathbf{p}}^{\pm}}} \sum_{\hat{\mathrm{P}}'} (\pm 1)^{p'} \langle \delta(\mathbf{r} - \mathbf{q}'') \, | \, \phi_{\mathbf{p}}(\hat{\mathrm{P}}'\mathbf{r}) \rangle$$

$$= \frac{1}{\sqrt{\chi_{\mathbf{p}}^{\pm}}} \sum_{\hat{\mathrm{P}}'} (\pm 1)^{p'} \phi_{\mathbf{p}}(\hat{\mathrm{P}}'\mathbf{q}'').$$

(7.1
64)

With this and transforming to the momentum continuum, one sees that the position probability density is proportional to

$$W_N(\mathbf{q}'', \mathbf{q}') \propto \sum_{\hat{\mathrm{P}}', \hat{\mathrm{P}}''} (\pm 1)^{p'+p''} \int \mathrm{d}\mathbf{p} \; \mathrm{e}^{-\beta p^2/2m} \phi_{\mathbf{p}}(\hat{\mathrm{P}}''\mathbf{q}'')\phi_{\mathbf{p}}(\hat{\mathrm{P}}'\mathbf{q}')^{*}$$

$$\propto \sum_{\hat{\mathrm{P}}', \hat{\mathrm{P}}''} (\pm 1)^{p'+p''} \int \mathrm{d}\mathbf{p} \; \mathrm{e}^{-\beta p^2/2m} \mathrm{e}^{-\mathbf{p}\cdot(\hat{\mathrm{P}}''\mathbf{q}'')/\mathrm{i}\hbar} \mathrm{e}^{\mathbf{p}\cdot(\hat{\mathrm{P}}'\mathbf{q}')/\mathrm{i}\hbar}$$

$$\propto \sum_{\hat{\mathrm{P}}} (\pm 1)^{p} \mathrm{e}^{-\pi\left(\mathbf{q}'-\hat{\mathrm{P}}\mathbf{q}''\right)^2/\Lambda^2}.$$

(7.1
65)

Here and throughout, the thermal wave length is $\Lambda \equiv \sqrt{2\pi\hbar^2/mk_{\mathrm{B}}T}$. The diagonal element of the particle position density is therefore

$$W_N(\mathbf{q}, \mathbf{q}) = \sum_{\hat{\mathrm{P}}} (\pm 1)^{p} \mathrm{e}^{-\pi\left(\mathbf{q}-\hat{\mathrm{P}}\mathbf{q}\right)^2/\Lambda^2}$$

$$= \sum_{\hat{\mathrm{P}}} (\pm 1)^{p} \prod_{j=1}^{N} \mathrm{e}^{-\pi(\mathbf{q}_j-\mathbf{q}_{j'})^2/\Lambda^2},$$

(7.1
66)

where j' is the label of particle j after the permutation $\hat{\mathrm{P}}$.

The nature of the Gaussians that appear here means that the particle position density is dominated by permutations amongst particles that are close together (since in order for

$\left(\mathbf{q}_j - \mathbf{q}_{j\prime}\right)^2$ to be small, particle j must be close to particle $j\prime$). To leading order this is the identity permutation, with $j\prime = j$, followed by a single transposition, say $\hat{\mathrm{P}}_{jk}$, with $j\prime = k$ and $k\prime = j$, followed by the permutation of three particles, $\hat{\mathrm{P}}_{jk\ell} = \hat{\mathrm{P}}_{\ell j}\hat{\mathrm{P}}_{jk}$, with $j\prime = \ell$, $k\prime = j$, and $\ell\prime = k$, etc. Hence one can expand the density as

$$W_N(\mathbf{q}, \mathbf{q}) = 1 \pm \sum_{j<k} \mathrm{e}^{-2\pi q_{jk}^2/\Lambda^2} + \sum_{j<k<\ell} \mathrm{e}^{-\pi q_{jk}^2/\Lambda^2}\mathrm{e}^{-\pi q_{k\ell}^2/\Lambda^2}\mathrm{e}^{-\pi q_{\ell j}^2/\Lambda^2} + \cdots$$

(7.1
67)

where $\mathbf{q}_{jk} = \mathbf{q}_j - \mathbf{q}_k$ and $q_{jk}^2 = \mathbf{q}_{jk} \cdot \mathbf{q}_{jk}$. The Gaussian in q_{jk} vanishes when $q_{jk} \gg \Lambda$. In the low density $N/V \to 0$ and/or high temperature $\Lambda \to 0$ limits, one has $N\Lambda^3/V \ll 1$. Since the spacing between particles is on the order of the inverse of the cube root of the number density, one sees that in the high temperature, low density limit the corrections due to quantum symmetrization are negligible. This decay of symmetrization effects with distance illustrates how it is localized to clusters of neighboring particles.

In the low density, high temperature limit the single-pair transpositions dominate the permutation sum. In this case one can write the position probability density in Maxwell–Boltzmann form with an effective pair potential (Pathria 1972, section 5.5, Attard 2021, equation (3.43)),

$$v(q_{jk}) = -k_{\mathrm{B}}T \ln\left[1 \pm \mathrm{e}^{-2\pi q_{jk}^2/\Lambda^2}\right].$$

(7.1
68)

This potential is for ideal particles and it arises from wave function symmetrization. The upper sign is for bosons and the lower sign is for fermions. This is graphed in figure 7.6. One can see that wave function symmetrization induces an effective attraction between bosons and an effective repulsion between fermions. This is how particle occupancy statistics manifest themselves in the continuum. Although this effective potential was derived for ideal non-interacting particles, one will not go too far wrong in thinking of real interacting particles as having a similar effective potential.

Figure 7.6. Effective pair potential due to wave function symmetrization for ideal bosons (solid curve) and ideal fermions (dashed curve).
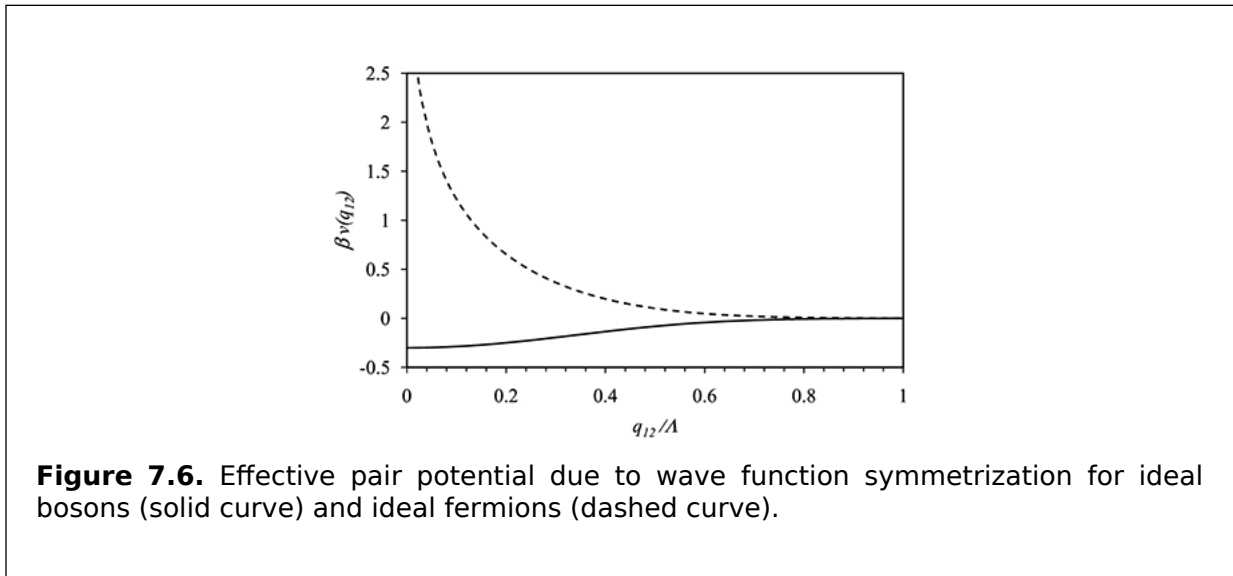
7.7.4 Quantum harmonic oscillators

For the case of the simple harmonic oscillator of frequency ω in d-dimensions, the well-known single-particle energy is (Messiah 1961, Merzbacher 1970)

$$\epsilon_n = \left[\frac{d}{2} + n_x + n_y + \cdots + n_d\right]\hbar\omega,$$

(7.169)

with the quantum state label being $n_\alpha = 0, 1, 2, \ldots$. Below for brevity we shall often sum over ϵ rather than n. Using the above loop expansion, the grand potential for a system of non-interacting simple harmonic oscillators is therefore

$$
\begin{aligned}
-\beta\Omega^\pm &= \sum_{l=1}^{\infty} \frac{(\pm 1)^{l-1} z^l}{l} e^{-dl\beta\hbar\omega/2} \prod_{\alpha=x}^{d} \sum_{n_\alpha=0}^{\infty} e^{-l\beta\hbar\omega n_\alpha} \\
&= \sum_{l=1}^{\infty} \frac{(\pm 1)^{l-1} z^l}{l} \left[\frac{e^{-l\beta\hbar\omega/2}}{1 - e^{-l\beta\hbar\omega}}\right]^d \\
&= \sum_{l=1}^{\infty} \frac{(\pm 1)^{l-1} z^l}{l} \left[e^{l\beta\hbar\omega/2} - e^{-l\beta\hbar\omega/2}\right]^{-d}.
\end{aligned}
$$

(7.170)

Note that this diverges for $z > e^{d\beta\hbar\omega/2}$.

The average energy of this system of independent harmonic oscillators is

$$
\begin{aligned}
\left\langle \hat{\mathcal{H}} \right\rangle_{z,T}^{\pm} &= \left(\frac{\partial(\beta\Omega^\pm)}{\partial\beta}\right)_z \\
&= \frac{d\hbar\omega}{2} \sum_{l=1}^{\infty} (\pm 1)^{l-1} z^l \frac{e^{l\beta\hbar\omega/2} + e^{-l\beta\hbar\omega/2}}{\left[e^{l\beta\hbar\omega/2} - e^{-l\beta\hbar\omega/2}\right]^{d+1}}.
\end{aligned}
$$

(7.171)

This is plotted in figure 7.7, where the monomer $l = 1$ term may also be called the result for classical or distinguishable particles. It can be seen that at fixed temperature the average energy increases with increasing fugacity. In general, the average energy for bosons is greater than that for fermions, and the monomer energy lies between the two. At low values of the fugacity all three energies converge. As $z \to e^{d\beta\hbar\omega/2}$, the boson energy diverges.
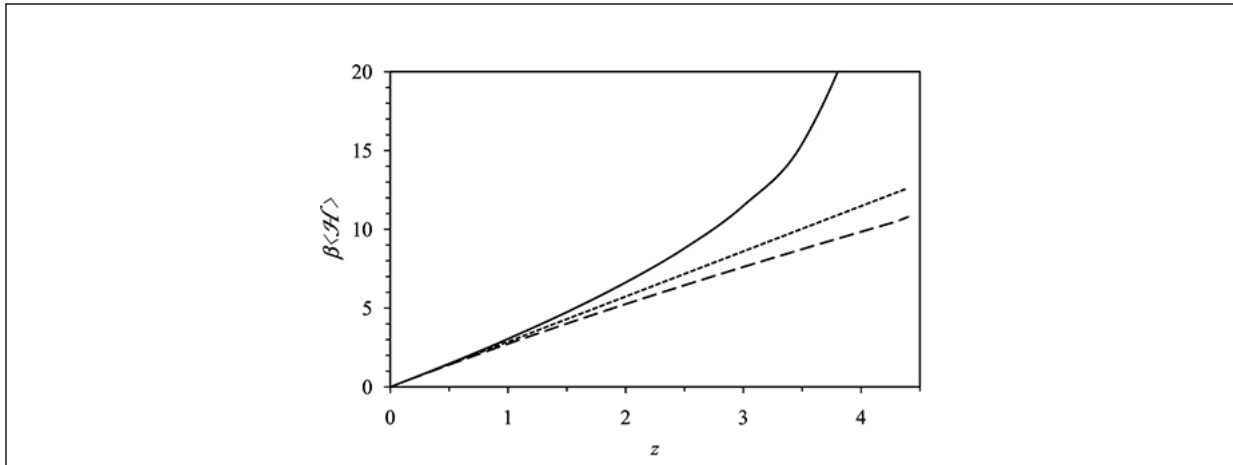
Figure 7.7. Average energy of a system of independent harmonic oscillators at $\beta\hbar\omega = 1$ and $d = 3$ for bosons (full curve), for fermions (long dashed curve), and for monomers (short dashed curve). From Attard (2021).

The average number of oscillators in the open system is given by the fugacity derivative of the grand potential. Hence the average number of oscillators in a permutation loop of size l is

$$
\begin{aligned}
\langle N \rangle_{z,T}^{\pm,(l)} &= -\beta z \left(\frac{\partial \Omega^{\pm,(l)}}{\partial z} \right)_T \\
&= (\pm 1)^{l-1} z^l \sum_n e^{-\beta l \varepsilon_n} \\
&= (\pm 1)^{l-1} z^l e^{-dl\beta\hbar\omega/2} \prod_{\alpha=x}^d \sum_{n_\alpha=0}^{\infty} e^{-l\beta\hbar\omega n_\alpha} \\
&= (\pm 1)^{l-1} z^l \left[e^{l\beta\hbar\omega/2} - e^{-l\beta\hbar\omega/2} \right]^{-d}.
\end{aligned}
$$

(7.172)

The average number of oscillators in total is

$$
\langle N \rangle_{z,T}^{\pm} = -\beta z \left(\frac{\partial \Omega^{\pm}}{\partial z} \right)_T = \sum_{l=1}^{\infty} (\pm 1)^{l-1} z^l \left[e^{l\beta\hbar\omega/2} - e^{-l\beta\hbar\omega/2} \right]^{-d}.
$$

(7.173)

It can be seen in figure 7.8 that the average number increases monotonically with increasing fugacity. Fractional numbers of course result from the averaging process. At low values of the fugacity particle statistics again have no influence, and all three cases—bosons, fermions, and monomers—coincide. In general, there are more bosons for a given fugacity than fermions, and the number of monomers lies between the two. The average number of bosons diverges in the limit $z \to e^{d\beta\hbar\omega/2}$. The average energy per particle, $\langle \hat{\mathcal{H}} \rangle_{z,T}^{\pm} / \langle N \rangle_{z,T}^{\pm}$, at constant temperature with increasing fugacity decreases monotonically for bosons, and increases monotonically for fermions. For monomers it is a constant (equal to the ratio of the first terms in the respective series) that lies between the two.
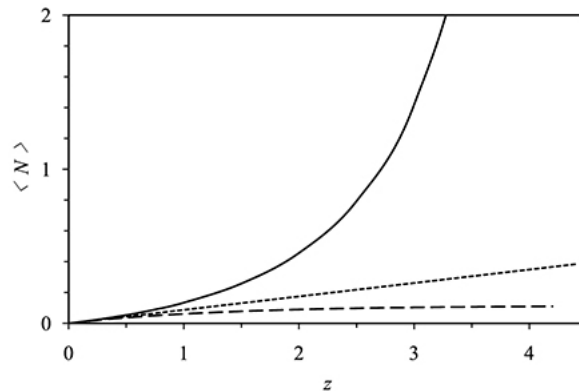
Figure 7.8. Average number of independent oscillators in the system at $\beta\hbar\omega = 1$ and $d = 3$ for bosons (full curve), for fermions (long dashed curve), and for monomers (short dashed curve). From Attard (2021).

7.7.5 Conventional quantum oscillators and interpretation

Although the present derivation of the general result for non-interacting particles via the loop expansion is different to the approach usually taken in textbooks, it may be shown to agree with those standard results. The loop series provides a novel interpretation of how symmetrization effects contribute to the grand potential.

Following Pathria (1972 section 6.2), single-particle states labeled by ε can be occupied by $N_\epsilon = 0, 1, \ldots, N_\pm$ particles, with $N_+ = \infty$ for bosons, and $N_- = 1$ for fermions. The grand partition function is the weighted sum over all possible occupancies of each state,

$$
\begin{aligned}
\Xi^\pm(z) &= \sum_{N_0=0}^{N_\pm} \sum_{N_1=0}^{N_\pm} \cdots \sum_{N_\infty=0}^{N_\pm} z^{\sum_\epsilon' N_\epsilon} \, \mathrm{e}^{-\beta \sum_\epsilon' \epsilon N_\epsilon} \\
&= \prod_{\epsilon=0}^{\infty}{}' \left\{ \sum_{N_\epsilon=0}^{N_\pm} z^{N_\epsilon} \mathrm{e}^{-\beta\epsilon N_\epsilon} \right\} \\
&= \prod_{\epsilon=0}^{\infty}{}' \left[1 \mp z\mathrm{e}^{-\beta\epsilon} \right]^{\mp 1} \\
&= \prod_{n_x=0}^{\infty} \prod_{n_y=0}^{\infty} \cdots \prod_{n_d=0}^{\infty} \left[1 \mp z\mathrm{e}^{-\beta\epsilon_n} \right]^{\mp 1} .
\end{aligned}
$$

(7.174)

The sums and products over energy is an abbreviated notation that visits each state once, as is explicit in the final equality. This is the expression given by Pathria (1972). The grand potential is given by the logarithm of this, a subsequent expansion of which yields

$$
\begin{aligned}
-\beta\Omega^\pm &= \mp \sum_n \ln\left[1 \mp z\mathrm{e}^{-\beta\epsilon_n} \right] \\
&= \pm \sum_n \sum_{l=1}^{\infty} \frac{(\pm z)^l}{l} \mathrm{e}^{-l\beta\epsilon_n} \\
&= \sum_{l=1}^{\infty} \frac{(\pm 1)^{l-1} z^l}{l} \sum_n \mathrm{e}^{-l\beta\epsilon_n} .
\end{aligned}
$$

(7.175)

This agrees with the above expression based on symmetrization loops, equation (7.153).

Again it must be emphasized that in the loop expansion all particles in the loop must be in the same single-particle state in order for the symmetrization factor to be nonzero. This holds even for fermions. From this perspective it is remarkable that the loop expansion approach gives the same result as the textbook particle occupancy approach, which explicitly excludes system states with more than one fermion in the same single-particle state.

For the case of a system of independent simple harmonic oscillators, the first equality in equation (7.175) is

$$-\beta\Omega^{\pm} = \mp \sum_{n_x=0}^{\infty} \cdots \sum_{n_d=0}^{\infty} \ln\left[1 \mp z e^{-\beta\hbar\omega d/2} e^{-\beta\hbar\omega n_x} \cdots e^{-\beta\hbar\omega n_d}\right].$$

(7.176)

This multi-dimensional sum over possible energy states is not as neat as the one-dimensional sum over symmetrization loops, equation (7.170). That latter appears to be rapidly converging and computationally efficient.

Although the present approach and the textbook approach arrive at the same result for ideal states, the present approach proceeds from a rather different view-point, namely that the symmetrization of the wave function is the fundamental axiom, and that the occupancy of states (multiple for bosons, single for fermions) is a quantity derived from it. The present approach is extremely useful when the results are transformed to the continuum that is classical phases space. In the case of the continuum it is impossible to define unambiguously discrete states and the occupancy thereof, whereas the symmetrization of the wave function itself remains a valid concept.

7.8 The classical world

In view of the above formulation of quantum statistical mechanics as an expansion with classical statistical mechanics as the leading term, it is of interest to discuss the origin of the classical world around us. The puzzle is that the underlying laws of the Universe must be quantum mechanical, but the behavior that we actually observe, and the equations of motion that have historically been developed to describe the world quantitatively, are classical in nature. How in actual fact does quantum mechanics give rise to classical behavior in general, and to classical mechanics in particular, on the terrestrial sphere?

The usual answer to this question, with which teachers have fobbed off students for generations, is that quantum mechanics applies to the very small, and that classical mechanics applies to macroscopic objects. This answer is as unsatisfactory as it is uninformative. It is nothing more than a restatement of the original question and it provides no mechanistic explanation as to why the everyday world is classical.

It seems to me that a genuine attempt to address this issue should proceed in two stages. First qualitatively, namely to explain why quantum phenomena are absent from everyday experience. And second quantitatively, namely to demonstrate how Newton's (equivalently, Hamilton's) equations of motion account for the movement that we observe and measure.

The following discussion invokes the quantum statistical mechanical results of sections 7.3 and 7.4 for a sub-system interacting with a reservoir. We shall apply those results to part of a macroscopic object, taking it as axiomatic that if any part behaves classically, then the object as a whole must behave classically. The results also imply the classical behavior of individual microscopic particles and molecules that interact with each other and with an environment.

The non-classical world is defined by three quantum phenomena: lack of simultaneity for certain variables (non-commutativity), the superposition of states (coherence, or interference), and the complete symmetrization or anti-symmetrization of the wave function with respect to particle interchange. These are absent in the classical Universe, where in the first place there is no impediment in principle to measuring any two properties at the same time, which is to say that the system can be simultaneously in macrostates of different collectives. Also, a classical measurement of a property yields only one value at a time, which is to say that it is in one and only one state of a collective at a time, and that these states cannot interfere with each other. Finally, classically identical particles can be interchanged without measurable consequences.

The results in section 7.3 explain the absence of the superposition of states in the classical world around us. There it was shown that the conservation law for the exchange of

energy and material between a sub-system and a reservoir (equivalently, an open system and its environment) caused the total wave function to become entangled, which caused the suppression of the superposition of the principle quantum states, and the cancelation of the superposition of the degenerate quantum states. The result of this collapse of the total wave function is that the sub-system is manifest as a mixture of pure quantum states, which do not interfere with each other and which defines a classical statistical system.

This result explains the corresponding absence of superposition states in our classical world. Any macroscopic particle, can be considered as the sum total of sub-systems that can exchange energy and material with each other. Hence each such sub-system acts as a mixture of pure states, and the total wave function of such an object has collapsed. Measurements of the mass, energy, position, or momentum of the object yield results compatible with classical experience.

The formulation of quantum statistical mechanics in section 7.4 showed that the effects of non-commutativity of the position and momentum operators could be dealt with by a particular expansion, the leading order term of which corresponded to the classical case in which both could be measured simultaneously. Higher order terms, which are weighted by powers of Planck's constant and of inverse temperature, are negligible at terrestrial temperatures and pressures. As above, any macroscopic object must therefore behave classically in so far as it simultaneously possesses a well-defined position and momentum.

Also, in section 7.4 wave function symmetrization was expressed as an expansion, with the higher order term depending on the spacing between particles relative to their thermal wave length. Again, for terrestrial temperatures and pressures, these are practically negligible. Hence any macroscopic object must behave classically in so far as insensitivity to identical particle interchange is concerned.

The analysis of sections 7.3 and 7.4 explains qualitatively the origins of the classical nature of the world around us given the underlying quantum properties of the atoms and molecules. The three unique features of a quantum system—the superposition of states, the lack of simultaneity, and particle interchange total symmetry or asymmetry—are absent in the observed Universe because of the collapse of the total wave function due to entanglement, and because the effects of non-commutativity and of interchange symmetry are quantitatively negligible at terrestrial temperatures and pressures. This argument, which was made explicitly for macroscopic particles, also holds for microscopic particles and molecules that interact with each other and their environment.

It remains to explain how the quantitative characterization of classical motion, namely Hamilton's equations of motion, arise from Schrödinger's equation. An elementary result in quantum mechanics is Ehrenfest's theorem (Messiah 1961, Merzbacher 1970). For a system in the wave state ψ, the particle positions and momenta are given by the expectation values

$$\mathbf{q}(\psi) = \langle \psi \mid \hat{\mathbf{q}} \mid \psi \rangle, \quad \text{and} \quad \mathbf{p}(\psi) = \langle \psi \mid \hat{\mathbf{p}} \mid \psi \rangle.$$

(7.177)

Ehrenfest's theorem says that the (adiabatic) rate of change of these is given by the expected derivatives of the Hamiltonian operator,

$$\dot{\mathbf{q}}^0(\psi) = \langle \psi \mid \frac{\partial \hat{\mathcal{H}}}{\partial \hat{\mathbf{p}}} \mid \psi \rangle, \quad \text{and} \quad \dot{\mathbf{p}}^0(\psi) = \langle \psi \mid \frac{-\partial \hat{\mathcal{H}}}{\partial \hat{\mathbf{q}}} \mid \psi \rangle.$$

(7.178)

The Hamiltonian operator is just the classical Hamiltonian function of the position and momentum operators, $\hat{\mathcal{H}} \equiv \mathcal{H}_{\mathrm{cl}}(\hat{\mathbf{p}}, \hat{\mathbf{q}})$. Ehrenfest's theorem derives directly from Schrödinger's equation (Messiah 1961, Merzbacher 1970).

The derivation of the first of these from Schrödinger's equation is

$$\dot{\mathbf{q}}^0(\psi) = \langle \psi \mid \hat{\mathbf{q}} \mid \dot{\psi}^0 \rangle + \langle \dot{\psi}^0 \mid \hat{\mathbf{q}} \mid \psi \rangle$$

$$= \frac{1}{i\hbar} \langle \psi \mid \{\hat{\mathbf{q}}\hat{\mathcal{H}} - \hat{\mathcal{H}}\hat{\mathbf{q}}\} \mid \psi \rangle$$

$$= \frac{-\hbar^2/2m}{i\hbar} \int d\mathbf{r} \, \psi(\mathbf{r})^* \{\mathbf{r}\nabla^2 - \nabla^2 \mathbf{r}\}\psi(\mathbf{r})$$

$$= \frac{-i\hbar}{m} \int d\mathbf{r} \, \psi(\mathbf{r})^* \nabla \psi(\mathbf{r})$$

$$= \frac{1}{m} \langle \psi \mid \hat{\mathbf{p}} \mid \psi \rangle$$

$$= \langle \psi \mid \frac{\partial \hat{\mathcal{H}}}{\partial \hat{\mathbf{p}}} \mid \psi \rangle.$$

(7.179)

The Hermitian nature of the Hamiltonian operator gives the second equality, and integration by parts gives the fourth equality after some cancelation. The derivation of the second of these from Schrödinger's equation is

$$\dot{\mathbf{p}}^0(\psi) = \langle \psi \mid \hat{\mathbf{p}} \mid \dot{\psi}^0 \rangle + \langle \dot{\psi}^0 \mid \hat{\mathbf{p}} \mid \psi \rangle$$

$$= \frac{1}{i\hbar} \langle \psi \mid \{\hat{\mathbf{p}}\hat{\mathcal{H}} - \hat{\mathcal{H}}\hat{\mathbf{p}}\} \mid \psi \rangle$$

$$= -\int d\mathbf{r} \, \psi(\mathbf{r})^* \{\nabla U(\mathbf{r}) - U(\mathbf{r})\nabla\}\psi(\mathbf{r})$$

$$= -\int d\mathbf{r} \, \psi(\mathbf{r})^* \psi(\mathbf{r})\nabla U(\mathbf{r})$$

$$= \langle \psi \mid \frac{-\partial \hat{\mathcal{H}}}{\partial \hat{\mathbf{q}}} \mid \psi \rangle.$$

(7.180)

The fourth equality follows from the product rule for differentiation, since the gradient operator acts on everything to its right.

For the case of momentum eigenfunctions, $\hat{\mathbf{p}} \mid \zeta_{\mathbf{p}} \rangle = \mathbf{p} \mid \zeta_{\mathbf{p}} \rangle$, the first of these is just

$$\dot{\mathbf{q}}^0 = \langle \zeta_{\mathbf{p}} \mid \frac{\partial \hat{\mathcal{H}}}{\partial \hat{\mathbf{p}}} \mid \zeta_{\mathbf{p}} \rangle = \frac{\partial \mathcal{H}_{\rm cl}(\mathbf{p}, \mathbf{q})}{\partial \mathbf{p}},$$

(7.181)

since in an eigenstate, the expectation of a function of an operator is just that function of the eigenvalue. For the case of position eigenfunctions, $\hat{\mathbf{q}} \mid \zeta_{\mathbf{q}} \rangle = \mathbf{q} \mid \zeta_{\mathbf{q}} \rangle$, the second of these is just

$$\dot{\mathbf{p}}^0 = \langle \zeta_{\mathbf{q}} \mid \frac{-\partial \hat{\mathcal{H}}}{\partial \hat{\mathbf{q}}} \mid \zeta_{\mathbf{q}} \rangle = \frac{-\partial \mathcal{H}_{\rm cl}(\mathbf{p}, \mathbf{q})}{\partial \mathbf{q}}.$$

(7.182)

These are just Hamilton's equations of motion. Since we have just seen that any macroscopic object or microscopic particle that interacts with its environment is

simultaneously in a well-defined position and momentum state, this shows how in the terrestrial sphere Hamilton's (equivalently, Newton's) classical equations of motion apply.

Summary

- The wave space of an isolated quantum system has uniform weight. This implies that the microstates of an isolated system have equal weight.
- In an open quantum system, the conservation laws entangle the wave functions of the sub-system and reservoir, which causes them to collapse into principle entropy states. Statistical averaging over the reservoir causes the degenerate entropy states to also collapse. These mean that an open sub-system is a mixture of pure entropy microstates.
- The probability operator is the exponential of the entropy operator, which, for the canonical equilibrium system, is the negative of the sub-system Hamiltonian operator divided by the reservoir temperature.
- The partition function and statistical average are the weighted sum over distinct, allowed entropy microstates. Accounting for symmetrization transforms this into a sum over all entropy microstates, and thence into a double sum over simultaneous position and momentum states.
- Written as an integral over classical phase space, the partition function and the statistical average include the classical Maxwell–Boltzmann weight, a commutation factor for the position and momentum states, and a permutation loop factor from wave function symmetrization.
- Expansion of these factors in powers of Planck's constant shows that classical statistical mechanics is the leading order contribution to quantum statistical mechanics. Classical mechanics is just quantum statistical mechanics for an open system.

References

Attard P 2002 *Thermodynamics and Statistical Mechanics: Equilibrium by Entropy Maximisation* (London: Academic)

Attard P 2015 *Quantum Statistical Mechanics: Equilibrium and Non-Equilibrium Theory from First Principles* (Bristol: IOP Publishing)

Attard P 2017 Quantum statistical mechanics results for argon, neon, and helium using classical Monte Carlo arXiv:1702.00096v1 [quant-ph]

Attard P 2018 Quantum statistical mechanics in classical phase space. Expressions for the multi-particle density, the average energy, and the virial pressure arXiv:1811.00730

Attard P 2021 *Quantum Statistical Mechanics in Classical Phase Space* (Bristol: IOP Publishing)

Barker J R 2010 Quantum phase-space distributions with compact support *Physica* E **42** 491

Barnett S M and Radmore P M 2003 *Methods in Theoretical Quantum Optics* (Oxford: Oxford University Press)

Breuer H-P and Petruccione F 2002 *The Theory of Open Quantum Systems* (Oxford: Oxford University Press)

Davies E B 1976 *Quantum Theory of Open Systems* (London: Academic)

Dishlieva K G 2008 Kirkwood and Wigner distribution functions: graphical imaging *Int. J. Pure Appl. Math.* **42** 583

Dubos R 1950 *Louis Pasteur: Free Lance of Science* (Boston, MA: Little Brown)

Feynman R P and Hibbs A R 1965 *Quantum Mechanics and Path Integrals* (New-York: McGraw-Hill)

Feynman R P 1998 *Statistical Mechanics* 2nd edn (Perseus: Westview Press)

Gerry C and Knight P 2005 *Introductory Quantum Optics* (Cambridge: Cambridge University Press)

Groenewold H J 1946 On the principles of elementary quantum mechanics *Physica* **12** 405

Kirkwood J 1933 Quantum statistics of almost classical particles *Phys. Rev.* **44** 31

Kleinert H 2009 *Path Integrals in Quantum Mechanics, Statistics Polymer Physics, and Financial Markets* (Singapore: World Scientific)

Messiah A 1961 *Quantum Mechanics* **vols I and II** (Amsterdam: North-Holland)

Merzbacher E 1970 *Quantum Mechanics* 2nd edn (New York: Wiley)

Moyal J E 1949 Quantum mechanics as a statistical theory *Proc. Cambridge Phil. Soc.* **45** 99

Pathria R K 1972 *Statistical Mechanics* (Oxford: Pergamon)

Praxmeyer L and Wódkiewicz K 2002 Quantum interference in the Kirkwood-Rihaczek representation arXiv:quant-ph/0207127v1

Schlosshauer M 2005 Decoherence, the measurement problem, and interpretations of quantum mechanics arXiv:quant-ph/0312059v4

Schulman L S 1981 *Techniques and Applications of Path Integration* (New York: Wiley)

von Neumann J 1932 *Mathematische Grundlagen der Quantenmechanik* (Berlin: Springer)

Weinberg S 2015 *To Explain the World. The Discovery of Modern Science* (New York: Harper Collins)

Weiss U 2008 *Quantum Dissipative Systems* 3rd edn (Singapore: World Scientific)

Wigner E 1932 On the quantum correction for thermodynamic equilibrium *Phys. Rev.* **40** 749

Zeh H D 2001 *The Physical Basis of the Direction of Time* 4th edn (Berlin: Springer)

Zachos C, Fairlie D and Curtright T 2005 *Quantum Mechanics in Phase Space* (Singapore: World Scientific)

Zurek W H 2003 Decoherence, einselection, and the quantum origins of the classical arXiv:quant-ph/0105127v3

Entropy Beyond the Second Law (Second Edition)

Thermodynamics and statistical mechanics for equilibrium, non-equilibrium, classical, and quantum systems

Phil Attard

Chapter 8

Entropy in Bose–Einstein condensation and the λ-transition

'McLennan noticed that when pumping to cool it down, liquid helium stopped boiling below 2.2 K'

S Balibar (2017)

'In 1924... Einstein remarked that this removal of the last vestige of individuality ... would imply a statistical preference of the molecules for having the same velocity, and this preference would lead, at a well-defined temperature, to a kind of change of state of aggregation; the molecules would 'condense' into the lowest quantum state, the state of momentum zero.'

F London (1950)

8.1 Introduction

It is generally believed that Bose–Einstein condensation signifies the occupancy of the ground energy state. Here evidence is presented that condensation is instead into multiple multiply-occupied momentum states.

The λ-transition in liquid helium-4 is signified by a sharp spike in the heat capacity at 2.17 K and the onset of superfluidity. This chapter applies the classical phase space formulation of quantum statistical mechanics to the phenomenon. Compared to existing treatments, what is new is the formulation in terms of momentum states, the emphasis on the permutation entropy, and the inclusion of interactions between the ^{4}He bosons. Initially at least, the analysis follows previous presentations (Attard 2021, chapter 5, Attard 2022).

The λ-transition in liquid helium is one of the earliest quantum phenomenon discovered in a condensed matter system (Keesom *et al* 1927, 1932). Since the heat capacity at the spike appeared to be continuous, Keesom *et al* (1927) argued that the transition was a second order liquid–liquid phase transition, which they dubbed helium I at high temperatures and helium II at low. F London (1938) made the connection with Bose–Einstein condensation, which had been theorized to occur at low temperatures when the majority of bosons transitioned from occupying the excited energy states to occupying the ground energy state (Landsberg 1954). F London (1938) analyzed ideal (non-interacting) bosons, and showed that they had a peak in the heat capacity that resembled the λ-transition. For ideal bosons with the mass of ^{4}He this occurred at $T_\lambda^{\text{id}} = 3.13$ K, which is remarkably close to the measured value of $T_\lambda = 2.17$ K (Donnelly and Barenghi 1998). Tisza (1938) suggested that superconductivity, which had been observed at and below the λ-transition, could be explained by a mixture of the two fluids, with He I comprising bosons in the excited energy states with normal viscosity, and He II being composed of bosons in the ground energy state and having zero viscosity.

Landau (1941) proposed a different model of the λ-transition and superfluidity to that of F London (1938) and Tisza (1938). He eschewed the ideas of Bose–Einstein condensation, the two-fluid model, and the non-interacting boson picture and instead attempted to quantize hydrodynamics by including collective behavior for interacting bosons as quantum excitations (phonons). He postulated the existence of rotons, whose energy spectrum he

fitted to measured heat capacity data. Later theoretical work focussed on interpreting rotons as quantized vortices (Feynman 1954, Kawatra and Pathria 1966, Pathria 1972), although direct evidence for this is lacking. Landau (1941) never accepted Bose–Einstein condensation as playing a role in the λ-transition, nor that Tisza's two-fluid model of superfluidity would not display friction and viscosity from collisions. Accounts of the fascinating history of the λ-transition and superfluidity have been given by Donnelly (1995, 2009) and by Balibar (2014, 2017). Griffin (1999) has reviewed the theoretical developments in the understanding of Bose–Einstein condensation.

The analysis of Bose–Einstein condensation and the λ-transition in this chapter begins with ideal bosons, sections 8.2 and 8.3. The formal analysis for interacting bosons and Monte Carlo computer simulation results are then given for pure permutation loops, section 8.4, and for mixed loops, section 8.5. This is followed by a discussion of the physical interpretation of the analysis and results, section 8.6.

The analysis in this chapter is based upon the binary division of bosons into either ground momentum state or else the set of excited momentum states, which form a continuum. This is a first approximation that makes direct contact with F London's (1938) original ideal boson analysis, and it also enables a relatively simple Monte Carlo simulation algorithm for interacting bosons. Undoubtedly it contains many elements of reality and it reveals the major physical principles that underly the λ-transition. However, some of the predictions that arise from this simple binary division at and below the λ-transition are difficult to reconcile with experimental measurement. At the end of this chapter, in section 8.6.6, and in the following chapter 9, a more nuanced model is discussed that replaces condensation into the ground momentum state by condensation into low-lying momentum states. This appears to be more realistic below the λ-transition, particularly in understanding the mechanism that underpins superfluid flow.

8.2 Ideal boson approach to the λ-transition

We now analyze Bose–Einstein condensation for ideal bosons. In this case the energy is purely kinetic, and energy states are effectively equivalent to momentum states. This is a significant point, which is useful in that it allows a deliberate ambiguity in referring to the ground state: one does not have to decide whether it is energy or momentum condensation that is occurring. On the other hand, it can lead to confusion, because for interacting bosons energy and momentum states are no longer equivalent. Because the ideal boson calculations appear to give Bose–Einstein condensation into the energy ground state, it has often been assumed that the same applies for interacting bosons. Logically, however, one should consider the possibility that the correct generalization of the ideal boson results to interacting bosons is that condensation occurs in the momentum ground state, which in such a case is not equivalent to the energy ground state. From the formulation of quantum statistical mechanics in chapter 7, the reader may anticipate that in this and the following chapters the view is that condensation is into momentum states, which in the binary division approximation is the momentum ground state.

8.2.1 Loop forms of the grand potential and number

The grand partition function, equation (7.71), for bosons, in terms of energy states, is

$$\Xi^+ = \sum_{N=0}^{\infty} \frac{z^N}{N!} \sum_{\mathbf{n}} \chi_{\mathbf{n}}^+ \mathrm{e}^{-\beta \mathcal{H}_{\mathbf{n}}}.$$

(8.1)

Here N is the number of bosons, the fugacity is $z \equiv \mathrm{e}^{\beta \mu}$, where μ is the chemical potential, and $\beta = 1/k_{\mathrm{B}} T$ is sometimes called the inverse temperature, with k_{B} being Boltzmann's constant and T the temperature. In the context of the following discussion,

small fugacity means $z \to 0$, and large fugacity means $z \to 1^-$. For ideal non-interacting bosons, the energy states are the sum of single-particle kinetic energy states,

$$\mathcal{H}_{\mathbf{n}} = \mathcal{K}_{\mathbf{n}} = \sum_{j=1}^{N} \mathcal{K}_{\mathbf{n}_j}^{(1)} = \frac{1}{2m} \sum_{j=1}^{N} \mathbf{p}_{\mathbf{n}_j}^2,$$

(8.2)

where m is the boson mass and $p_{\alpha,n_{\alpha j}} = n_{\alpha j} \Delta_p$ is the α-component of momentum of boson j, with the quantized momentum states being $n_{\alpha j} = 0, \pm 1, \pm 2, \ldots$. We shall assume ordinary three-dimensional space, with the bosons being confined to a cube of volume $V = L^3$, giving the spacing between momentum states as $\Delta_p = 2\pi\hbar/L$ (Merzbacher 1970, Messiah 1961). Because these are single-particle energy states, the analysis of section 7.7 holds, and the grand potential is given by equation (7.150) applied to bosons

$$-\beta \Omega^+(\mu, V, T) = -\beta \sum_{l=1}^{\infty} \Omega^{+,(l)}.$$

$$= \sum_{l=1}^{\infty} \frac{z^l}{l} \sum_{\mathbf{n}_1} e^{-l\beta \mathcal{K}_{\mathbf{n}_1}^{(1)}}.$$

(8.3)

Here $\mathbf{n}_1$ is a single-particle momentum state label.

Now transform the sum over states to a continuum integral. We need to retain the discrete ground momentum state explicitly in order to identify condensation. Therefore, to correct for double counting, we need to subtract the integral over the ground momentum state interval. This gives

$$-\beta \Omega^{+,(l)} = \frac{z^l}{l} \left[1 + \Delta_p^{-3} \int d\mathbf{p} \, e^{-l\beta p^2/2\,m} - \Delta_p^{-3} \int_{\Delta_p^3} d\mathbf{p} \, e^{-l\beta p^2/2\,m} \right]$$

$$= \frac{z^l}{l} \left[1 + \left(\frac{L^2}{l\Lambda^2} \right)^{3/2} - \left(\frac{L^2}{l\Lambda^2} \right)^{3/2} \mathrm{erf} \left(\sqrt{\frac{2\pi l\Lambda^2}{8L^2}} \right)^3 \right].$$

(8.4)

The thermal wavelength is $\Lambda = \sqrt{2\pi\hbar^2\beta/m}$.

The asymptotic forms of the error function are $\mathrm{erf}(x) \sim (2/\sqrt{\pi})x$, $x \to 0$, and $\mathrm{erf}(x) \sim 1 - e^{-x^2}/\sqrt{\pi}\,x$, $x \to \infty$ (Abramowitz and Stegun 1970). It follows that the limiting results for the loop grand potential are

$$-\beta \Omega^{+,(l)} \sim \begin{cases} \dfrac{z^l}{l} \left(\dfrac{L^2}{l\Lambda^2} \right)^{3/2}, & l\Lambda^2/L^2 \ll 1, \\[2ex] \dfrac{z^l}{l}, & l\Lambda^2/L^2 \gg 1. \end{cases}$$

(8.5)

In the limit of small argument, $l\Lambda^2/L^2 \ll 1$ (high temperature, small loops), the two ground state contributions cancel exactly and only the integral over excited states survives. In the limit of large argument, $l\Lambda^2/L^2 \gg 1$ (low temperature, large loops), the

integral over excited states cancels with the integral about the ground state, and only the discrete ground state contribution survives.

Each of these two asymptotic limits dominates the other in its respective regime. Hence if we simply add them together the resultant function is guaranteed correct in both asymptotic limits,

$$
-\beta\Omega^{+,(l)} \approx \frac{z^l}{l} + \frac{z^l}{l}\left(\frac{L^2}{l\Lambda^2}\right)^{3/2}
$$

$$
\equiv -\beta\Omega_0^{+,(l)} - \beta\Omega_*^{+,(l)}.
$$

(8.6)

When summed over l this gives

$$
-\beta\Omega^+ = -\ln[1-z] + \Lambda^{-3}V\sum_{l=1}^{\infty} z^l l^{-5/2}
$$

$$
\equiv -\beta\Omega_0^+ - \beta\Omega_*^+.
$$

(8.7)

This is the same as the expression given by F London (1938) (see also Pathria (1972 chapter 7)). The first term is the discrete ground state contribution and the second term is the continuum excited state contribution. The contribution from the ground state correction integral about the origin has been neglected. One sees that the approximate expression gives the correct result in the two respective asymptotic limits. It might therefore be expected to be a reasonable approximation throughout, as is confirmed in section 8.3.

The excited state contribution here is extensive, being proportional to volume *V*. From this one sees that for small values of the fugacity, $z \ll 1$, the excited states dominate the grand potential.

It will be shown next that $z = \bar{N}_0^+/(1 + \bar{N}_0^+) \to 1 - 1/\bar{N}_0^+$ in the large fugacity limit. Hence the ground state contribution grows as the logarithm of system size, $-\beta\Omega_0^+ \sim \ln \bar{N}$, which is relatively negligible. Hence for all values of the fugacity only the excited state contribution to the boson grand potential itself needs to be retained (Pathria 1972 section 7.1). This does not hold for the derivatives of the grand potential.

8.2.1.1 Number and energy of ideal bosons
The average number of bosons contributed by *l*-loops is the fugacity derivative of the loop grand potential, $\bar{N}^{+,(l)} = z(\partial(-\beta\Omega^{+,(l)})/\partial z)_{T,V}$. This is evidently just *l* times the loop grand potential itself,

$$
\bar{N}^{+,(l)} = z^l + \Lambda^{-3}Vz^l l^{-3/2}, \quad l \geqslant 1.
$$

(8.8)

One can identify z^l as the average number of bosons contributed by *l*-loops to the ground state. The average number of bosons in the ground state in total is

$$
\bar{N}_0^+ = \sum_{l=1}^{\infty} z^l = \frac{z}{1-z}.
$$

(8.9)

The number in the ground state evidently diverges as $z \to 1^-$. One can invert this to express the fugacity as $z = \bar{N}_0^+/(1 + \bar{N}_0^+)$, which is strictly less than one.

The average total number of bosons is the sum of that in the ground state and that in the excited states, $\bar{N}^+ = \bar{N}_0^+ + \bar{N}_*^+$. For the average number of bosons in a loop of given size, the ratio of those in the excited states to those in the ground state is independent of the fugacity and decreases with increasing size, $\bar{N}_*^{+,(l)}/\bar{N}_0^{+,(l)} \sim l^{-3/2} \to 0$, $l \to \infty$. As $z \to 1^-$, larger loops contribute increasingly to the total number, and hence the total proportion of contributions from the ground state also increases, $\bar{N}_0^+/\bar{N}^+ \to 1$, $z \to 1^-$. One can conclude that as the fugacity approaches one from below, large symmetrization loops become filled in the ground state but not in the excited states. Hence the extra bosons entering the system do so in the ground state. This is the phenomenon of Bose–Einstein condensation.

The loop energy for these ideal bosons is

$$\bar{\mathrm{E}}^{+,(l)} = \left(\frac{\partial(\beta\Omega^{+,(l)})}{\partial\beta}\right)_{V,z} = \frac{3}{2\beta}\Lambda^{-3}Vz^l l^{-5/2}.$$

(8.10)

Obviously the ground state does not contribute to this. Since the total pressure is related to the total grand potential by $\beta\Omega^+ = -\beta p^+ V$, we may identify the contribution to the pressure from each permutation loop as $p^{+,(l)} = -\Omega^{+,(l)}/V$. Hence one sees that the pressure is proportional to the energy,

$$p^{+,(l)} = \frac{2\mathrm{E}^{+,(l)}}{3V}, \quad \text{and} \quad p^+ = \frac{2\mathrm{E}^+}{3V}.$$

(8.11)

This is the usual result for the ideal gas.

8.2.2 Total grand potential and number

We now analyze the total grand potential and number, which are the sum of the loop contributions. To this end it is useful to define the Bose–Einstein integral (Pathria 1972 appendix D)

$$g_n(z) = \frac{1}{\Gamma(n)}\int_0^\infty \mathrm{d}x\ x^{n-1}\frac{z\mathrm{e}^{-x}}{1-z\mathrm{e}^{-x}}$$

$$= \sum_{l=1}^\infty z^l l^{-n}.$$

(8.12)

This is related to the Riemann zeta-function, $\zeta(n) = g_n(1)$. The symmetrization loop analysis provides a physical interpretation for the otherwise mathematical fugacity expansion of the Bose–Einstein integral. With this one sees that the total grand potential, neglecting the ground state contribution, is

$$-\beta\Omega^+ = -\sum_{l=1}^\infty \beta\Omega^{+,(l)} = \Lambda^{-3}Vg_{5/2}(z),$$

(8.13)

and, keeping the ground state, the total number is

$$\bar{N}^+ = \sum_{l=1}^{\infty} \bar{N}^{+,(l)} = \bar{N}_0^+ + \bar{N}_*^+ = \bar{N}_0^+ + \Lambda^{-3}V g_{3/2}(z).$$

<div align="right">(8.1 4)</div>

In the limit $z \to 0$, where $g_{3/2}(z) \sim z$, then $\bar{N}_0^+ \sim z$ and $\bar{N}_*^+ \sim \Lambda^{-3}Vz$. The second is a factor of volume larger than the first, and so for a macroscopic system one concludes that for small values of the fugacity most bosons are in excited states, and that $\bar{N}^+ \approx \bar{N}_*^+ \sim \Lambda^{-3}Vz$, which is the classical result.

The Bose–Einstein integral $g_{3/2}(z)$ increases monotonically with z and has largest value at the terminus of its domain, $g_{3/2}(1) = \zeta(3/2) = 2.612$. It follows that the total number of excited bosons at a given temperature is bounded, $\bar{N}_*^+(z,V,T) \leqslant \Lambda(T)^{-3}V\zeta(3/2)$.

For large values of the fugacity, $z \lesssim 1$, the Bose–Einstein integral is a relatively slowly varying function of z, $g_{3/2}(z) \approx g_{3/2}(1) = \zeta(3/2)$. Hence as the fugacity is increased the actual number of bosons in the excited states is close to the upper bound at the specified temperature.

When the total number of bosons exceeds the maximum number that can be accommodated in the excited states at the specified temperature, $N^+ > \Lambda^{-3}V\zeta(3/2)$, then the excess must go into the ground state. This is Bose–Einstein condensation. Rewriting this criterion for the temperature one can say that condensation occurs when

$$T < T_{\mathrm{c}} \equiv \frac{h^2}{2\pi m k_{\mathrm{B}}} \left(\frac{N^+}{V\zeta(3/2)} \right)^{2/3}.$$

<div align="right">(8.1 5)</div>

The thermal wavelength at this condensation temperature is about 40% larger than the average spacing of neighbors in the liquid.

For low temperatures, $T \leqslant T_{\mathrm{c}}$, the system can be regarded as a mixture of two phases:

- the gaseous phase, which consists of $\bar{N}_*^+ = (T/T_{\mathrm{c}})^{3/2}N$ bosons in excited states;
- the condensed phase, which consists of $\bar{N}_0^+ = N^+ - \bar{N}_*^+$ bosons in the ground state.

For $T > T_{\mathrm{c}}$, there is only a gaseous phase of bosons in excited states $\bar{N}_*^+ = N^+$ and $\bar{N}_0^+ = 0$. This corresponds to Keesom's (1927) helium I. Below the λ-transition these two coexisting phases that emerge from F London's (1938) ideal gas model of Bose–Einstein condensation correspond to Tisza's (1938) two-fluid model of superconductivity, with bosons in the ground state forming the zero-viscosity liquid. Figure 8.1 plots the fraction of the system in the ground and excited states according to this model.
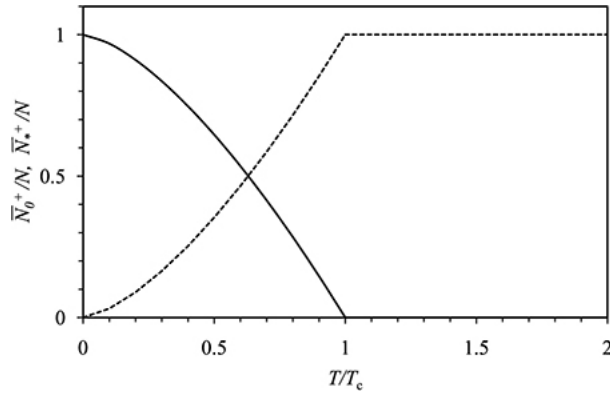
Figure 8.1. Ideal gas model of Bose–Einstein condensation. The graph shows the fraction of ideal bosons in the ground state (full curve) and in excited states (dashed curve) as a function of temperature.

The boson number density is $\rho(z, V, T) = \bar{N}^+(z, V, T)/V$. For temperatures greater than the condensation temperature, $T > T_c$, $\rho(z, V, T) = \Lambda(T)^{-3} g_{3/2}(z)$. In the high temperature limit this is $\rho(z, V, T) = \Lambda(T)^{-3} z$, $T \to \infty$, which is the classical result. For $T < T_c$, the ground state ground state plus continuum excited states ideal gas picture gives $\rho(z, V, T) = z/(1 - z)V + \Lambda(T)^{-3} \zeta(3/2)$. One expects that for fixed N, $z \sim 1 - \mathcal{O}(V^{-1})$.

The pressure is proportional to the grand potential,

$$\beta p^+(z, V, T) = \frac{-\beta \Omega^+(z, V, T)}{V} = \Lambda(T)^{-3} g_{5/2}(z),$$

(8.16)

as also derived above, equation (8.11). The present result derives from classical thermodynamics and is based on the facts that the pressure is the volume derivative of the grand potential, and that the grand potential is proportional to the volume. In the present case we have shown above that the ground state contribution to the grand potential, which is not proportional to the volume, can be neglected. The excited state contribution is proportional to the volume, and so the above result holds in the present case. One sees that the pressure depends upon only two intensive variables, even in the condensed regime.

8.2.3 Heat capacity

Now for the heat capacity at constant number and volume. One has at constant volume $\mathrm{d}\bar{N} = (\partial \bar{N}/\partial z)_{T,V} \, \mathrm{d}z + (\partial \bar{N}/\partial T)_{z,V} \, \mathrm{d}T$. Hence $(\partial z/\partial T)_{N,V} = -(\partial \bar{N}/\partial T)_{z,V} / (\partial \bar{N}/\partial z)_{T,V}$. Now

$$\left(\frac{\partial \bar{N}^+}{\partial z} \right)_{T,V} = \frac{1}{(1 - z)^2} + \Lambda^{-3} V z^{-1} g_{1/2}(z),$$

since $g'_n(z) = z^{-1}g_{n-1}(z)$, and

$$\left(\frac{\partial \bar{N}^+}{\partial T}\right)_{z,V} = \frac{3}{2}T^{-1}\Lambda^{-3}V g_{3/2}(z).$$

From the average loop energy found above, the average energy is

$$\bar{\mathrm{E}}^+(z,V,T) = \sum_{l=1}^{\infty}\bar{\mathrm{E}}^{+,(l)}(z,V,T) = \frac{3}{2\beta}\Lambda^{-3}V g_{5/2}(z).$$

Putting these together the heat capacity per particle at constant number and volume is

$$\frac{C_V^+}{k_{\mathrm{B}}N^+} = \frac{1}{k_{\mathrm{B}}N^+}\left(\frac{\partial \bar{\mathrm{E}}^+}{\partial T}\right)_{N,V}$$

$$= \frac{1}{k_{\mathrm{B}}N^+}\left\{\left(\frac{\partial \bar{\mathrm{E}}^+}{\partial T}\right)_{z,V} - \left(\frac{\partial \bar{\mathrm{E}}^+}{\partial z}\right)_{T,V}\left(\frac{\partial \bar{N}^+}{\partial T}\right)_{z,V}\left(\frac{\partial \bar{N}^+}{\partial z}\right)_{T,V}^{-1}\right\}$$

$$= \frac{1}{k_{\mathrm{B}}N^+}\left\{\frac{15k_{\mathrm{B}}}{4}\Lambda^{-3}V g_{5/2}(z)\right.$$

$$\left. - \frac{3k_{\mathrm{B}}T}{2}\Lambda^{-3}Vz^{-1}g_{3/2}(z)\frac{(3/2)T^{-1}\Lambda^{-3}V g_{3/2}(z)}{(1-z)^{-2} + \Lambda^{-3}Vz^{-1}g_{1/2}(z)}\right\}.$$

We now evaluate this for the ideal gas model above and below the transition temperature.

For $T \leqslant T_{\mathrm{c}}$, $\bar{N}_*^+ = (T/T_{\mathrm{c}})^{3/2}N^+$ and $\bar{N}_0^+ = N^+(1-(T/T_{\mathrm{c}})^{3/2})$. In this regime we can take $g_n(z) = g_n(1) = \zeta(n)$, and $(\partial \bar{N}^+/\partial z)_{T,V} = (1-z)^{-2}$. This gives

$$\frac{C_V^+}{k_{\mathrm{B}}N^+} = \frac{1}{k_{\mathrm{B}}N^+}\left\{\frac{15k_{\mathrm{B}}}{4}\Lambda^{-3}V\zeta(5/2) - \frac{9k_{\mathrm{B}}}{4}\Lambda^{-6}V^2\zeta(3/2)^2z^{-1}(1-z)^2\right\}$$

$$\approx \frac{15}{4\rho\Lambda^3}\zeta(5/2), \quad T \leqslant T_{\mathrm{c}}.$$

The second term in the first equality is proportional to the fugacity times the square of the ratio of the number of bosons in excited states to the number in the ground state, which may be taken to be negligible for temperature less than the condensation temperature. Since in the ideal gas model of Bose–Einstein condensation $\rho\Lambda(T)^3 = (T/T_{\mathrm{c}})^{-3/2}\zeta(3/2)$, this says that at low temperatures the heat capacity is proportional to $T^{3/2}$.

At the condensation temperature, where $\rho\Lambda(T_{\mathrm{c}})^3 = \zeta(3/2)$ this is

$$\frac{C_V^+(T_{\mathrm{c}})}{k_{\mathrm{B}}N^+} = \frac{15\zeta(5/2)}{4\zeta(3/2)} = 1.925.$$

(Experimentally, the measured specific heat capacity at the λ-transition is integrably divergent.)

For $T > T_{\mathrm{c}}$, we can neglect the ground state contributions and take $N^+ = \bar{N}_*^+ = \Lambda^{-3} V g_{3/2}(z)$, or $\rho \Lambda^3 = g_{3/2}(z)$, which implicitly gives $z(\rho, T)$. This gives

$$\frac{C_V^+}{k_{\mathrm{B}} N^+} = \frac{1}{k_{\mathrm{B}} \Lambda^{-3} V g_{3/2}(z)} \left\{ \frac{15 k_{\mathrm{B}}}{4} \Lambda^{-3} V g_{5/2}(z) - \frac{9 k_{\mathrm{B}}}{4} \frac{\Lambda^{-3} V g_{3/2}(z)^2}{g_{1/2}(z)} \right\}$$

$$= \frac{15 g_{5/2}(z)}{4 g_{3/2}(z)} - \frac{9 g_{3/2}(z)}{4 g_{1/2}(z)}, \quad T > T_{\mathrm{c}}.$$

As $z \to 0$, this gives $C_V^+ / k_{\mathrm{B}} N^+ \to (15/4) - (9/4) = 3/2$, which is the classical result.

As $z \to 1$, $g_{1/2}(z) \to \infty$, and the second term in can be neglected. The formula then gives $C_V^+ / k_{\mathrm{B}} N^+ = 15\zeta(5/2)/4\zeta(3/2)$, which equals the result found above at the condensation temperature. Hence the model predicts that the heat capacity is continuous at the condensation temperature. However, it does give a second order discontinuity at the transition point.

For plotting purposes use the facts that $\rho \Lambda(T)^3 = g_{3/2}(z)$, $T \geqslant T_{\mathrm{c}}$, which implicitly gives $z(\rho, T)$, $\rho \Lambda(T)^3 / \rho \Lambda(T_{\mathrm{c}})^3 = (T_{\mathrm{c}}/T)^{3/2}$ at constant density, and $\rho \Lambda(T_{\mathrm{c}})^3 = \zeta(3/2)$. Hence

$$\frac{C_V^+}{k_{\mathrm{B}} N^+} = \frac{15}{4 \rho \Lambda^3} \zeta(5/2) = \frac{15 \zeta(5/2)}{4 \zeta(3/2)} (T/T_{\mathrm{c}})^{3/2}, \quad T \leqslant T_{\mathrm{c}},$$

and

$$\frac{C_V^+}{k_{\mathrm{B}} N^+} = \frac{15 g_{5/2}(z)}{4 g_{3/2}(z)} - \frac{9 g_{3/2}(z)}{4 g_{1/2}(z)}, \quad T > T_{\mathrm{c}},$$

with $T(z)/T_{\mathrm{c}} = [\zeta(3/2)/g_{3/2}(z)]^{2/3}$.

These are plotted in figure 8.2, where the first order discontinuity in the heat capacity at $T = T_{\mathrm{c}}$ is evident. It can be seen that the heat capacity asymptotes to the classical ideal gas result at high temperatures. Also, it falls to zero as $T \to 0$, where almost all the bosons are in the ground state of zero energy, and those that are excited by increasing the temperature from absolute zero are a negligible fraction of the total. The famous λ shape of the curve is reasonably clear in the figure. The similarity in shape to the measured heat capacity of He^4 allows the laboratory data to be interpreted as signifying ground state condensation, which is ultimately due to the fully symmetrized wave function of bosons. Undoubtedly, particle interactions play some role in the quantitative values of the heat capacity. For example, the transition in He^4 is measured at 2.17 K, and the ideal gas model gives $T_{\mathrm{c}} = 3.13$ K (Pathria 1972, section 7.1). The finite value of the peak of the ideal boson specific heat capacity contrasts with the experimentally measured integrable divergence at the λ-transition.
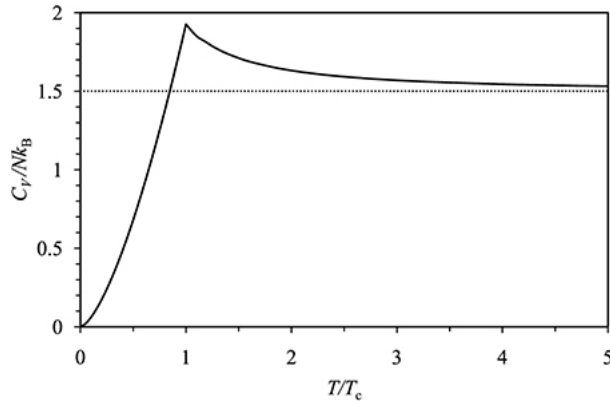
Figure 8.2. Specific heat capacity of ideal bosons as a function of temperature. The dotted line is a guide to the eye.

8.2.4 Discussion of ground state condensation

There is an issue with the F London (1938) ideal boson theory for the occupancy of the ground momentum state following condensation. As shown in equation (8.9), this gives the number of ideal bosons in the ground momentum state as $\bar{N}_0^{+,\mathrm{id}} = z/[1-z]$. Here z is the fugacity, which is an intensive variable.

In general the thermodynamic state of a single component system can be specified by two intensive variables and one extensive variable (chapter 2). Any intensive variable can be expressed as a function of two other independent intensive variables. For two fixed intensive variables, any extensive variable is linearly proportional to any other extensive variable. In the present case of the λ-transition, the prediction that the number of condensed bosons, $\bar{N}_0^{+,\mathrm{id}}$, is a function only of the fugacity, which is a function of the chemical potential and the temperature, means that it is an intensive variable. Hence, this predicts that two systems with the same pressure and temperature below the λ-transition have the same number of condensed bosons even if they have different sizes. Since the total number of bosons N is an extensive variable, this says that the fraction of condensed bosons $\bar{N}_0^{+,\mathrm{id}}/N$ must go to zero in the thermodynamic limit (fixed intensive variables, $V \to \infty$).

On the other hand, equation (8.14) gives the number of excited momentum state bosons as $\bar{N}_*^{+,\mathrm{id}} = \Lambda^{-3} V g_{3/2}(z) \sim N\zeta(3/2)/\rho\Lambda^3$, which is extensive. This means that the number of ground momentum state bosons must also be extensive $\bar{N}_0^{+,\mathrm{id}} = N - \bar{N}_*^{+,\mathrm{id}} \sim N[\rho\Lambda^3 - \zeta(3/2)]/\rho\Lambda^3$.

It is not possible to make the fugacity depend on number in an extensive fashion without fundamentally changing the usual thermodynamic relationships. Hence one has to conclude that this contradiction is an artefact of the procedures used by F London (1938) in his analysis of condensation of ideal bosons. In section 8.6.6, it is argued that the binary division into ground and excited momentum state bosons is responsible for the problem. Although a reasonable first approximation, a more sophisticated picture of condensation and superfluidity is required. In section 8.6.6 it is shown that in fact the occupancy of a momentum state, including the ground momentum state, is an intensive variable. The more sophisticated picture of condensation beyond the binary division is given in sections

, , and . The new interpretation reconciles the lack of a measured latent heat at the λ-transition (Donnelly and Barenghi 1998), with the notion that superfluidity is carried by condensed bosons.

8.3 Ideal bosons: exact enumeration

One can further explore the ideal gas model of Bose–Einstein condensation and the λ-transition in liquid helium by the exact enumeration of the quantum states. This avoids the transformation to the continuum, and the binary division approximation invoked by F London (1938) that was just discussed.

   For non-interacting particles the energy states are single-particle states. For the present ideal gas the single-particle energy eigenvalues are

$$\mathcal{H}_{\mathbf{n}_1}^{(1)} = \frac{\Delta_p^2}{2m}[n_x^2 + n_y^2 + n_z^2], \quad n_\alpha = 0, \pm 1, \pm 2, \ldots, \quad \Delta_p = 2\pi\hbar/L.$$

(8.26)

Here the volume of the system is $V = L^3$.

   The loop grand potential for ideal bosons, equation (8.3), is

$$-\beta\Omega^{+,(l)} = \frac{z^l}{l} \sum_{\mathbf{n}_1} \mathrm{e}^{-l\beta\mathcal{H}_{\mathbf{n}_1}^{(1)}}$$

$$= \frac{z^l}{l} \left[\sum_{n_x=-\infty}^{\infty} \mathrm{e}^{-\pi l(\Lambda/L)^2 n_x^2} \right]^3, \quad l \geqslant 1,$$

(8.27)

where the thermal wave-length is $\Lambda = \sqrt{h^2\beta/2\pi m}$. The total grand potential is

$$-\beta\Omega^+ = -\sum_{n_x,n_y,n_z} \ln\left[1 - z\mathrm{e}^{-\pi(\Lambda/L)^2 n^2} \right], \quad n^2 = n_x^2 + n_y^2 + n_z^2.$$

(8.28)

This is exact. Differentiating this one can obtain expressions for the most likely number, energy, and heat capacity as sums over states (Attard 2021 section 5.3).

8.3.1 Numerical results

Figure 8.3 shows the specific heat capacity at constant density for ideal bosons at a function of temperature near the λ-transition. The ground state plus continuum excited states model, equations (8.21) and (8.23) (full curve) is the same as in figure 8.2. The exact enumeration results use the derivatives of the sum over states, which were terminated when the Maxwell–Boltzmann exponent exceeded 20.
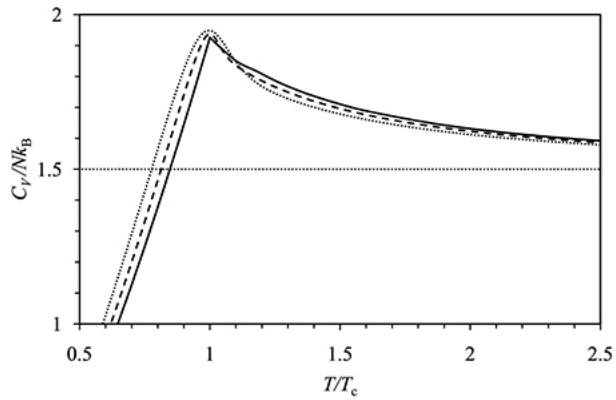
Figure 8.3. Specific heat capacity of ideal bosons as a function of temperature. The solid curve is the ground state plus continuum excited states model, section 8.2, and the remaining curves are exact results for fixed density $\rho = 0.3$ and $N = 500$ (dotted) and $N = 5000$ (short dashed). The condensation temperature T_c was defined by the location of the maximum for each curve. The dotted line is a guide to the eye.

It can be seen that there is relatively good agreement between the ground state plus continuum excited states model and the exact results for finite N. Indeed it appears that in the thermodynamic limit, the present exact results converge upon the earlier results. As an example of this, the present peak occurs at: $\{N, \rho\Lambda_c^3, C_{V,c}/Nk_B\} = \{500, 2.25, 1.948\}$, $\{1000, 2.34, 1.944\}$, $\{2000, 2.40, 1.940\}$ $\{5000, 2.45, 1.937\}$ and $\{10\,000, 2.49, 1.934\}$. The ground state plus continuum excited states model gives $\rho\Lambda_c^3 = 2.612$ and $C_{V,c}/Nk_B = 1.925$.

Figure 8.4 shows the fractional occupancy of the ground and one of each of the first several excited momentum states for ideal bosons around the λ-transition. The number in the ground state increases monotonically with decreasing temperature, whereas the number in each excited momentum state first increases and then decreases. The change in behavior coincides with the peak in the heat capacity. The number in each excited momentum state decreases with increasing quantum number. Note that the energy is degenerate in momentum states, and so at higher temperatures there is not monotonic order in the occupancy of the energy states. Obviously most bosons are distributed amongst even higher excited energy states in the high temperature regime. The number occupancy for the energy continuum would be peaked at a nonzero value that increased with increasing temperature.
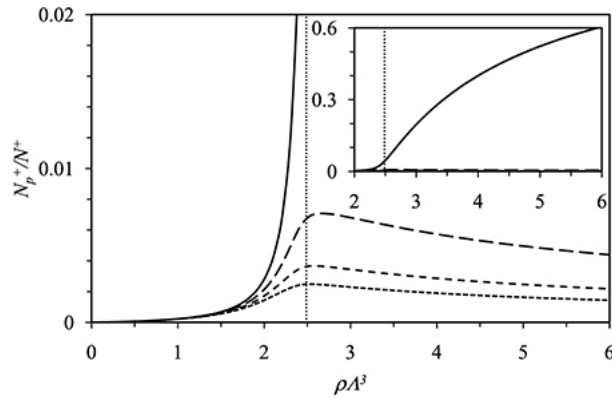
Figure 8.4. Average fraction of ideal bosons in the ground momentum state (solid curve), and in one of the six first excited momentum states (long dashed curve), one of the 12 second excited momentum states (dashed curve), and one of the eight third excited momentum states (short dashed curve). The total number is $N^+ = 10\,000$ and the density is $\rho = 0.3$, both fixed while the temperature and fugacity are varied. The dotted line locates the heat capacity maximum. **Inset.** Long-range view.

More or less at the transition, where the heat capacity peaks, the fraction of bosons in the excited states also reaches a peak before decreasing to zero at absolute zero. The fraction of bosons in the ground momentum state grows after the transition and appears to dominate the occupancy (inset to figure 8.4). This, however, is not the whole story as the results depend upon the system size (see section 9.2). For $N = 500$ about 20% of the bosons are in the ground momentum state at the λ-transition. For $N = 1000$ it is about 10%, for $N = 5000$ it is about 5% and for $N = 10\,000$ it is about 4%. In the thermodynamic limit the exact treatment of ideal bosons apparently shows a continuous condensation transition, beginning with zero ground momentum state occupation at the transition itself. The F London (1938) treatment of ideal bosons also begins at zero and shows a continuous increase in ground momentum state occupation below the transition, which is consistent with the exact behavior shown in the inset to figure 8.4. But as discussed in section 8.2.4, the F London (1938) theory inconsistently gives the ground momentum state occupancy as both an intensive and an extensive variable. If the former, then the fractional occupancy would be dependent on the system size, and it must vanish in the thermodynamic limit. The present exact calculations do indeed show a system-size dependence for the occupancy, and they confirm the intensive nature of the ground momentum state occupancy (see section 9.2).

Figure 8.5 shows the number of bosons as monomers and in the first three loops as a function of temperature. At high temperatures most bosons are monomers and permutation loops are practically non-existent. The number of monomers decreases monotonically with decreasing temperature. Obviously, this means that the number of bosons in permutation loops increases monotonically with decreasing temperature. The number of bosons in each of the loops shown increases with decreasing temperature and reaches a peak on the high temperature side of the transition. For dimers this peak occurs at $T/T_c = 1.44$ for $N^+ = 1000$, and 1.48 for $N^+ = 5000$, which is almost within the plotting point digitization error. The number of bosons in dimers exceeds those as trimers, which exceeds those as tetramers, for all temperatures shown. The gap between these decreases with decreasing temperature.
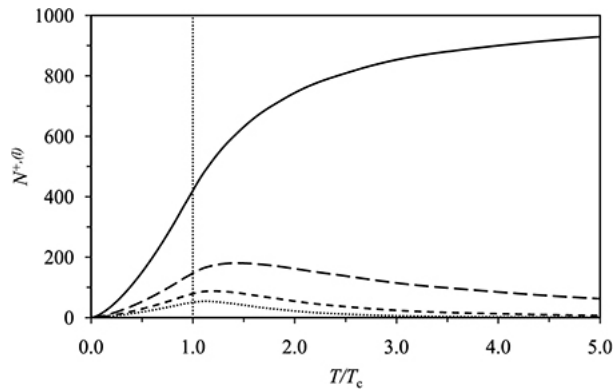
Figure 8.5. Average number of ideal bosons as monomers (solid curve), as dimers (long dashed curve), as trimers (short dashed curve), and as tetramers (dotted curve). The total number is $N^+ = 1000$ and the density is $\rho = 0.3$, both fixed while the temperature and fugacity are varied. The dotted line is a guide to the eye.

The decrease with decreasing temperature in the number of bosons in the first several loops on the low temperature side of the transition, together with the decrease in the number of monomers, suggests that larger loops are being populated at the expense of monomers and smaller loops (because the total number of bosons is fixed). But bosons in these larger loops must increasingly be in the ground state, because of the l in the Maxwell–Boltzmann exponent in equation (8.27). (It was pointed out in section 8.2.1 that large loops are mainly in the ground state, and that large loops increasingly dominate as $T \to 0$.) This is consistent with the increasing occupancy of the ground state shown in figure 8.4.

The total heat capacity is the sum of the heat capacity of the monomers and loops. The loop heat capacity varies with the excited state of the loop. Ground state bosons contribute nothing to the heat capacity. Hence one can see the reason why the heat capacity first increases and then decreases with decreasing temperature at the λ-transition: On the high temperature side the number of bosons in small loops is increasing, and a significant number of these bosons are in excited states. On the low temperature side the number of bosons in small loops decreases as they are cannibalized by large loops of bosons in the ground state.

8.4 The λ-transition for interacting bosons

This section goes beyond ideal bosons to present mathematical analysis and numerical results for Bose–Einstein condensation and the λ-transition for interacting spin-zero bosons that experience a pair potential. The analysis is applied to helium-4, whose atom is a spin-zero boson because it comprises three pairs of spin-half fermions each with opposite spin and occupying the same single-particle state. The analysis utilizes the classical phase space formulation of quantum statistical mechanics, chapter 7.

The analysis accounts for the permutation symmetrization of bosons via the symmetrization function. This is ultimately the cause of Bose–Einstein condensation, and, as will be seen, is a non-local property that can be dominated by long-range effects. Accordingly, the commutation function, which accounts for the non-commutativity of the position and momentum operators in the classical phase space formulation of quantum statistical mechanics, is neglected because in general it is short-ranged compared to the

pair potential. The neglect of the commutation function certainly simplifies the analysis and allows the exploration of extensive numerical results. The fact that these results are qualitatively and even semi-quantitatively in agreement with experimental measurements for the λ-transition in ^{4}He tends to confirm the validity of the approximation.

A counter argument is provided by figure 7.5, which shows the effect of the commutation function evaluated to fourth order on the pressure of Lennard-Jones ^{4}He. Such a large increase in the quantum contribution to the liquid-state pressure would require a liquid-state density smaller than classical to equal the required saturation vapor pressure. This would reduce the position permutation loop symmetrization contribution compared to the results given below.

As is shown in this and the following sections, the results that include the symmetrization function but not the commutation function do depend upon the pair potential and other short-range effects. The discrepancies with laboratory results might indicate the need to take into account the commutation function to improve the accuracy and the reliability of the numerical results.

The numerical results given here for the Lennard-Jones model of ^{4}He were obtained by Monte Carlo simulation. Details of the algorithm are given in section 5.4 of Attard (2021).

8.4.1 Formal analysis

Consider a system of N identical bosons interacting with potential energy $U(\mathbf{q}^N)$, or more simply $U(\mathbf{q})$, which does not depend on the momentum state of the bosons. At any instant there are N_0 bosons in the ground momentum state and N_* in excited momentum states, with $N = N_0 + N_*$. Ground momentum state bosons are labeled $j \in N_0$, and excited momentum state bosons are labeled $j \in N_*$. The momentum eigenvalue of boson j is $\mathbf{p}_j$. The classical kinetic energy is $\mathcal{K}(\mathbf{p}^N) = \mathcal{K}(\mathbf{p}^{N_*}) = \sum_{j \in N_*} p_j^2/2\,m = p^2/2m$. The normalized momentum eigenfunctions for the discrete momentum case are $|\,\mathbf{p}\rangle = V^{-N/2}\mathrm{e}^{-\mathbf{q}\cdot\mathbf{p}/\mathrm{i}\hbar}$. The system has volume $V = L^3$ and the spacing between momentum states is $\Delta_p = 2\pi\hbar/L$ (Messiah 1961, Merzbacher 1970).

The grand partition function for bosons is (cf equation (7.80), and also Attard (2018, 2021))

$$
\begin{aligned}
\Xi &= \mathrm{TR}\prime\,\mathrm{e}^{-\beta\hat{\mathcal{H}}} \\
&= \sum_{N=0}^{\infty} \frac{z^N}{N!} \sum_{\hat{\mathrm{P}}} \sum_{\mathbf{p}} \left\langle \hat{\mathrm{P}}\mathbf{p}|\mathrm{e}^{-\beta\hat{\mathcal{H}}}|\mathbf{p} \right\rangle \\
&= \sum_{N=0}^{\infty} \frac{z^N}{N!} \sum_{\hat{\mathrm{P}}} \sum_{\mathbf{p}} \int \mathrm{d}\mathbf{q}\, \left\langle \hat{\mathrm{P}}\mathbf{p} \mid \mathbf{q} \right\rangle \left\langle \mathbf{q}|\mathrm{e}^{-\beta\hat{\mathcal{H}}}|\mathbf{p} \right\rangle \\
&\approx \sum_{N=0}^{\infty} \frac{z^N}{N!V^N} \sum_{\hat{\mathrm{P}}} \sum_{\mathbf{p}} \int \mathrm{d}\mathbf{q}\, \mathrm{e}^{-\beta\mathcal{H}(\mathbf{q},\mathbf{p})} \frac{\left\langle \hat{\mathrm{P}}\mathbf{p} \mid \mathbf{q} \right\rangle}{\langle \mathbf{p} \mid \mathbf{q} \rangle} \\
&= \sum_{N=0}^{\infty} \frac{z^N}{N!V^N} \sum_{\mathbf{p}} \int \mathrm{d}\mathbf{q}\, \mathrm{e}^{-\beta\mathcal{H}(\mathbf{q},\mathbf{p})} \eta(\mathbf{q},\mathbf{p}).
\end{aligned}
$$

(8.29)

The superscript + has been dropped because this chapter deals only with bosons. The permutation operator is $\hat{\mathrm{P}}$. In the penultimate equality the commutation function has been neglected; the error introduced by this approximation is negligible in systems that are

dominated by long-range effects (Attard 2018, 2021). The classical Hamiltonian phase space function is $\mathcal{H}(\mathbf{q}, \mathbf{p}) = \mathcal{K}(\mathbf{p}) + U(\mathbf{q})$. The symmetrization function $\eta(\mathbf{q}, \mathbf{p})$ is the sum of Fourier factors over all boson permutations.

As shown above in the derivation of equation (8.6), the F London (1938) analysis of the λ-transition for ideal bosons depended upon making the transformation to the momentum continuum while including separately the ground momentum state, and neglecting the continuum integral correction around the ground momentum state. We make an identical approximation for the present interacting bosons, which gives

$$
\begin{aligned}
\Xi &= \sum_{N=0}^{\infty} \frac{z^N}{N! V^N} \prod_{j=1}^{N} \left\{ \delta_{\mathbf{p}_j, 0} + \Delta_p^{-3} \int d\mathbf{p}_j \right\} \\
&\quad \times \int d\mathbf{q}^N \ e^{-\beta \mathcal{H}(\mathbf{q}^N, \mathbf{p}^N)} \eta(\mathbf{q}^N, \mathbf{p}^N) \\
&= \sum_{N=0}^{\infty} \frac{z^N}{N! V^N} \sum_{N_0=0}^{N} \frac{N! \, \Delta_p^{-3N_*}}{N_0! N_*!} \int d\mathbf{p}^{N_*} \ e^{-\beta \mathcal{K}(\mathbf{p}^{N_*})} \\
&\quad \times \int d\mathbf{q}^N \ e^{-\beta U(\mathbf{q}^N)} \eta(\mathbf{q}^N, \mathbf{p}^N).
\end{aligned}
$$

(8.30)

In the final equality there are N_0 ground momentum state bosons and $N_* = N - N_0$ excited momentum state bosons; below we write $\sum_{N, N_0} \Rightarrow \sum_{N_0, N_*}$. All N bosons contribute to the potential energy and to the symmetrization function. Also $\mathbf{p}_j = 0$ if $j \in N_0$.

8.4.2 Pure loops

The symmetrization function is the sum of all permutations of the bosons in the Fourier factors. Each permutation can be expressed as a product of disjoint permutation loops, section 7.5 (Attard 2018, 2021). In the present context permutation loops can be classified as pure or mixed. A pure permutation loop consists only of ground momentum state bosons, or only of excited momentum state bosons. A mixed permutation loop contains both.

In this section we shall pursue the leading order approximation in which all mixed permutation loops are neglected, so that the symmetrization function is the product of the ground momentum state symmetrization function and the excited momentum state symmetrization function,

$$
\eta(\mathbf{q}^N, \mathbf{p}^N) \approx \eta_0(\mathbf{q}^{N_0}, \mathbf{p}^{N_0}) \, \eta_*(\mathbf{q}^{N_*}, \mathbf{p}^{N_*}).
$$

(8.31)

Here η_0 is the sum total of weighted permutations of ground momentum state bosons. Since their momentum is zero, $\mathbf{p}_j = 0$, $j \in N_0$, one has that

$$
\frac{\left\langle \hat{P} \mathbf{p}^{N_0} \mid \mathbf{q}^{N_0} \right\rangle}{\left\langle \mathbf{p}^{N_0} \mid \mathbf{q}^{N_0} \right\rangle} = 1, \text{all } \hat{P}.
$$

(8.32)

Hence

$$
\eta_0(\mathbf{q}^{N_0}, \mathbf{p}^{N_0}) = N_0!.
$$

Of relevance to the later discussion of superfluidity, this result is a manifestation of quantum non-locality for ground momentum state bosons. All the ground momentum state bosons in the system contribute equally regardless of spatial location.

The remaining factor η_* is the sum total of weighted permutations of excited momentum state bosons. In previous work (Attard 2021 section 5.4) the number of discrete ground momentum state bosons was taken to be negligible, $N_0 = 0$ and the continuum integral over momentum space was applied to all N bosons. Hence all of those previous formal results carry over directly to the present case of pure excited momentum state permutation loops with the replacement $N \Rightarrow N_*$. (An adjustment has to be made to the statistical averages when the ground momentum state bosons are not negligible, as is discussed below.) In particular, η_*, is a sum of products of loops. One can define $\mathring{\eta}_* = \sum_{l \geqslant 2} \eta_*^{(l)}$ as the sum of single loops, and one can write either

$$\eta_*(\mathbf{q}^{N_*}, \mathbf{p}^{N_*}) = \exp \mathring{\eta}_*(\mathbf{q}^{N_*}, \mathbf{p}^{N_*}),$$

or else

$$\langle \eta_* \rangle_{\mathrm{cl}} = \mathrm{e}^{\left\langle \mathring{\eta}_* \right\rangle_{\mathrm{cl}}}.$$

One can drop the subscript $*$ on η when the bosons that contribute are obvious from its arguments.

A formal derivation of the symmetrization function as the exponential of a loop series that gives the grand partition function as the series of loop grand potentials is given in section 7.5.2. Here it may be sufficient to illustrate these two results by writing out the first few terms of the permutation series explicitly,

$$\eta(\mathbf{q}^{N_*}, \mathbf{p}^{N_*}) = 1 + \sum_{j,k}^{N_*} {}'' \, \mathrm{e}^{-\mathbf{p}_j \cdot \mathbf{q}_{jk}/i\hbar} \mathrm{e}^{-\mathbf{p}_k \cdot \mathbf{q}_{kj}/i\hbar}$$

$$+ \sum_{j,k,n}^{N_*} {}'' \, \mathrm{e}^{-\mathbf{p}_j \cdot \mathbf{q}_{jk}/i\hbar} \mathrm{e}^{-\mathbf{p}_k \cdot \mathbf{q}_{kn}/i\hbar} \mathrm{e}^{-\mathbf{p}_n \cdot \mathbf{q}_{nj}/i\hbar}$$

$$+ \sum_{j,k,n,m}^{N_*} {}'' \, \mathrm{e}^{-\mathbf{p}_j \cdot \mathbf{q}_{jk}/i\hbar} \mathrm{e}^{-\mathbf{p}_k \cdot \mathbf{q}_{kj}/i\hbar} \mathrm{e}^{-\mathbf{p}_n \cdot \mathbf{q}_{nm}/i\hbar} \mathrm{e}^{-\mathbf{p}_m \cdot \mathbf{q}_{mn}/i\hbar} + \cdots$$

$$\approx 1 + \eta^{(2)}(\mathbf{q}^{N_*}, \mathbf{p}^{N_*}) + \eta^{(3)}(\mathbf{q}^{N_*}, \mathbf{p}^{N_*}) + \frac{1}{2} \eta^{(2)}(\mathbf{q}^{N_*}, \mathbf{p}^{N_*})^2 + \cdots$$

The double prime indicates that the sums are over unique permutations, and also that in any product term no boson may belong to more than one permutation loop. The first term is the monomer or unpermuted one, and it gives rise to classical statistics. The second term is the dimer, the third is the trimer, and the fourth is the double dimer. The second equality no longer forbids permutation loop intersections; the error from this approximation ought to be negligible in the thermodynamic limit.

In view of these results, in particular the one that writes the classical average of the full symmetrization function as the exponential of the classical average of the single loop

symmetrization function, it is straightforward to write the grand potential as the sum of loop grand potentials,

$$\Omega(z, V, T) = -k_{\mathrm{B}}T \, \ln \, \Xi(z, V, T)$$

$$= \sum_{l=1}^{\infty} \Omega_{*}^{(l)}(z, V, T).$$

The monomer grand potential is just the classical grand potential, with a trivial adjustment for N_0 and N_* as independent. It is given by

$$\mathrm{e}^{-\beta\Omega^{(1)}(z,V,T)} = \Xi_{\mathrm{cl}}(z, V, T)$$

$$= \sum_{N_0,N_*} \frac{z^N \Delta_p^{-3N_*}}{N_*! V^N} \int \mathrm{d}\mathbf{p}^{N_*} \, \mathrm{e}^{-\beta\mathcal{K}(\mathbf{p}^{N_*})} \int \mathrm{d}\mathbf{q}^N \, \mathrm{e}^{-\beta U(\mathbf{q}^N)}$$

$$= \sum_{N_0,N_*} \frac{z^N \Lambda^{-3N_*}}{N_*! V^{N_0}} \int \mathrm{d}\mathbf{q}^N \, \mathrm{e}^{-\beta U(\mathbf{q}^N)}$$

$$= \sum_{N_0,N_*} \frac{z^N \Lambda^{-3N_*}}{N_*! V^{N_0}} Q(N, V, T).$$

Again $N = N_0 + N_*$. The $N_0!$ in the denominator has canceled with $\eta_0 = N_0!$ in the integrand, since this multiplies each of the terms in η_*. The classical configurational integral, $Q(N, V, T)$, does not distinguish between ground and excited momentum state bosons. The thermal wavelength is $\Lambda = \sqrt{2\pi\beta\hbar^2/m}$.

The loop grand potentials $l \geqslant 2$ are classical averages (Attard 2021 section 5.3), which can be taken in a canonical system

$$-\beta\Omega_*^{(l)} = \left\langle \eta^{(l)}(\mathbf{p}^{N_*}, \mathbf{q}^{N_*}) \right\rangle_{N_0,N_*,\mathrm{cl}}$$

$$= \left\langle G^{(l)}(\mathbf{q}^{N_*}) \right\rangle_{N_0,N_*,\mathrm{cl}}$$

$$= \left(\frac{N_*}{N}\right)^l \left\langle G^{(l)}(\mathbf{q}^N) \right\rangle_{N,\mathrm{cl}}$$

$$\equiv N_* \left(\frac{N_*}{N}\right)^{l-1} g^{(l)}.$$

We have transformed the average from the mixed $\{N_0, N_*\}$ system to the classical configurational system of N bosons that does not distinguish their state. This transformation invokes a factor of $(N_*/N)^l$, which is the uncorrelated probability that l bosons chosen at random in the original mixed system are all excited. The Gaussian position loop function is

$$G^{(l)}(\mathbf{q}^N) = \sum_{j_1,\dots,j_l}{}' \mathrm{e}^{-\pi q_{j_l,j_1}^2/\Lambda^2} \prod_{k=1}^{l-1} \mathrm{e}^{-\pi q_{j_k,j_{k+1}}^2/\Lambda^2}.$$

The prime indicates that no two indeces may be equal and that distinct loops must be counted once only. There are $N!/(N-l)!l$ distinct l-loops here, the overwhelming number of which are negligible upon averaging. Since the pure excited momentum state permutation loops are compact in configuration space, one can define an intensive form of the average loop Gaussian, $g^{(l)} \equiv \left\langle G^{(l)}(\mathbf{q}^N) \right\rangle_{N,\mathrm{cl}}/N$. This is convenient because it does not depend upon N_*.

(8.40)

The mix of ground and excited momentum state bosons can be determined by minimizing the grand potential with respect to N_* at constant N. This is equivalent to maximizing the total entropy. The derivative is

$$\left(\frac{\partial(-\beta\Omega)}{\partial N_*} \right)_N = -\ln \frac{N_*\Lambda^3}{V} + \sum_{l=2}^{\infty} l \left(\frac{N_*}{N} \right)^{l-1} g^{(l)}.$$

(8.41)

If this is positive, then N_* should be increased, and vice versa. In terms of the number density $\rho_* = N_*/V$ and $\rho = N/V$, if $\rho_*\Lambda^3 \leqslant \rho\Lambda^3 < 1$, then both terms are positive (since $g^{(l)} \geqslant 0$). In this case the only stable solution is $\bar{\rho}_* = \rho$, which is to say that there are no ground momentum state bosons. The value $\rho\Lambda^3 = 1$ provides a lower bound on the λ-transition.

The structure of this equation indicates that $\bar{N}_*$ is extensive, and hence so is $\bar{N}_0 = N - \bar{N}_*$. This should be considered in the context of the discussion in section 8.2.4, and also section 8.6.6. The present analysis, like that for ideal bosons in section 8.2, is based on the binary division into ground and excited momentum state bosons.

8.4.3 Pure loop numerical results

Figure 8.6 shows the simulated heat capacity for interacting bosons, namely helium-4 with Lennard-Jones pair interactions. Details of the Monte Carlo algorithm are given by Attard (2021 section 5.3.2). The difference between the present results and those reported by Attard (2021 section 5.3.3) is that the earlier results were for an inhomogeneous canonical system consisting of a liquid drop in equilibrium with its own vapor. The present results are for a homogenous canonical system at the saturated liquid density previously established at each temperature by the inhomogeneous simulation. Artifacts due to the fluctuations in the liquid drop that contribute significantly to the heat capacity for the total system reported by Attard (2021) are absent in figure 8.6. It can be seen that the magnitude of the heat capacity for the Lennard-Jones liquid at the lowest temperature shown is much larger than that of the ideal gas, figure 8.2, and that measured experimentally for ^{4}He in the vicinity of the λ-transition (Donnelly and Barenghi 1998).
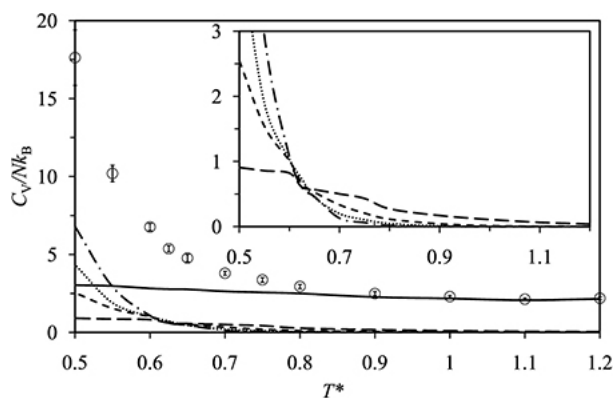
Figure 8.6. The heat capacity for Lennard-Jones ^{4}He along the saturation curve (canonical Monte Carlo, homogeneous, $N_* = N = 5000$). The symbols are the total of the monomer and loop contributions. The full curve is the monomer contribution. The loop contribution curves are dimer (long dashed), trimer (short dashed), tetramer (dotted), and pentamer (dash-dotted). Each arm of an error bar is twice the standard error. **Inset.** Magnification of the loop contributions.

The individual loop contributions come from expressing the second temperature derivative of the loop grand potentials as classical averages (Attard 2021 section 5.3.1). It can be seen that for $T \lesssim 0.9$ the quantum contributions are negligible. For $0.6 \lesssim T^* \lesssim 0.8$ the loop contributions increase with decreasing temperature, with the lower order loops having largest effect. At about $T^* = 0.6$ the order of the loops flips, and the loop series appears divergent. This suggests that for the Lennard-Jones fluid the λ-transition occurs around about $T_\lambda^* \approx 0.6$, which for the usual Lennard-Jones parameters of ^{4}He (van Sciver 2012) corresponds to $T_\lambda \approx 6$ K. (It will be shown in section 8.5 that including mixed loops gives the temperature $0.625 \leqslant T^* < 0.65$ as the onset of ground momentum state condensation.) At $T^* = 0.65$, the pure loop heat capacity is $C_V^{LJ}/Nk_B = 4.77$, which is comparable to the maximum heat capacity measured at the λ-transition, $C_s/Nk_B = 5.49$ (Donnelly and Barenghi 1998).

One should not be unduly concerned that this predicted λ-transition temperature differs from the measured one, $T_c = 2.17$ K (Donnelly and Barenghi 1998). The present Lennard-Jones parameters, $\varepsilon_{He} = 10.22k_B$ J and $\sigma_{He} = 0.255\,6$ nm (van Sciver 2012) could be simply rescaled, $\varepsilon = 0.43\varepsilon_{He}$ and $\sigma = 1.4\sigma_{He}$, to bring the transition temperature and the saturation density into consonance with the measured values. The functional form of the Lennard-Jones pair potential, and the absence of many-body potentials, appear inadequate for a quantitatively accurate description of liquid helium at these low temperatures. The real message from figure 8.6 is that particle interactions play an essential role in the λ-transition and that this can be explored with computer simulations.

A second limitation of the simulation results in figure 8.6 is that they used classical canonical averages with N bosons and the momentum continuum, $\langle \cdots \rangle_{N,cl}$. This is valid if the number of bosons in the ground momentum state is negligible compared to the number of excited momentum state bosons, $N_0 \ll N_*$. Such an assumption obviously breaks down after the λ-transition. In the unmixed case the primary correction for this effect is to multiply the energy due to an l-loop, $l \geqslant 2$, by $(\bar{N}_*/N)^l$, using an estimate of

the optimum fraction of excited momentum state bosons at each temperature. This is discussed in section 8.5.4. If condensation increases continuously with decreasing temperature, as will now be shown in the unmixed case, such a correction reduces the heat capacity in figure 8.6 following the λ-transition.

Using the mean field results of the previous section the number of ground momentum state bosons can be estimated. Table 8.1 gives the intensive Gaussian loop weights, $g^{(l)} \equiv \langle G^{(l)}(\mathbf{q}^N)\rangle_{N,\mathrm{cl}}/N$. These were obtained from the Monte Carlo simulations of Lennard-Jones ^{4}He (Attard 2021). The statistical error was less than the displayed digits. It can be seen that the terms increase with decreasing temperature. At each temperature the terms decrease with increasing loop size, except possibly at the lowest temperature shown.

Table 8.1. Intensive Gaussian loop $g^{(l)} = \langle G^{(l)}\rangle_{N,\mathrm{cl}}/N$ from liquid drop Monte Carlo simulations ($N = 500$, $\rho\sigma^3 = 0.3$) for saturated Lennard-Jones He^4 at various temperatures, $T^* \equiv k_\mathrm{B}T/\varepsilon$.

l	$T^* = 1$	0.9	0.8	0.7
2	2.44×10^{-3}	5.59×10^{-3}	1.24×10^{-2}	2.70×10^{-2}
3	6.77×10^{-5}	2.81×10^{-4}	1.11×10^{-3}	4.25×10^{-3}
4	3.84×10^{-6}	2.60×10^{-5}	1.67×10^{-4}	1.06×10^{-3}
5	3.20×10^{-7}	3.28×10^{-6}	3.38×10^{-5}	3.57×10^{-4}
l	$T^* = 0.6$	0.55	0.5	0.45
2	5.90×10^{-2}	8.72×10^{-2}	1.29×10^{-1}	1.96×10^{-1}
3	1.63×10^{-2}	3.20×10^{-2}	6.32×10^{-2}	1.33×10^{-1}
4	6.62×10^{-3}	1.68×10^{-2}	4.28×10^{-2}	1.30×10^{-1}
5	3.62×10^{-3}	1.19×10^{-2}	3.94×10^{-2}	1.56×10^{-1}
6	2.08×10^{-3}	8.76×10^{-3}	3.72×10^{-2}	2.24×10^{-1}

Table 8.2 shows the optimal fraction of bosons in the system that are in the ground momentum state as a function of temperature. The mean field theory, equation (8.41), was used together with the results in table 8.1. The results represent the stable solutions. At $T^* = 0.45$, changing l_max from 6 to 5 changes $\bar{N}_0/N$ from 0.682 to 0.684.

Table 8.2. Saturation density and fraction of ground momentum state bosons in Lennard-Jones He^4. From liquid drop Monte Carlo simulations with $N_* = N = 500$ at an overall density of $\rho\sigma^3 = 0.3$.

$k_\mathrm{B}T/\varepsilon$	$\rho\sigma^3$	$\rho\Lambda^3$	T (K)[†]	ρ_m (kg m^{-3})[†]	$\bar{N}_0/N$
1	0.62	0.76	10.220	246.76	0
0.9	0.73	1.04	9.198	290.54	0.027

$k_{\mathrm{B}}T/\varepsilon$	$\rho\sigma^3$	$\rho\Lambda^3$	T (K)†	ρ_{m} (kg m^{-3})†	$\bar{N}_0/N$
0.8	0.77	1.31	8.176	306.46	0.220
0.7	0.81	1.68	7.154	322.38	0.381
0.6	0.88	2.31	6.132	350.24	0.536
0.55	0.87	2.60	5.621	346.26	0.583
0.5	0.91	3.13	5.110	366.16	0.635
0.45	0.95	3.82	4.599	376.91	0.682
0.4	0.98	4.72	4.088	390.04	0.709

†$\varepsilon_{\mathrm{He}} = 10.22k_{\mathrm{B}}$ J and $\sigma_{\mathrm{He}} = 0.255\,6$ nm.

One can compare these results to the exact calculations for ideal bosons given in section 8.3. As reported in Attard (2021 section 5.2) for $N = 500$ about 20% of the bosons are in the ground momentum state at the λ-transition. For $N = 1000$ it is about 10%, and for $N = 5000$ it is about 5%. In the thermodynamic limit the ideal bosons apparently show a continuous condensation transition, beginning with zero ground momentum state occupation at the transition itself, followed by a continuous increase in occupancy below the transition. In contrast, the present results for the Lennard-Jones liquid, table 8.2, have a substantial fraction of bosons occupying the ground momentum state at the transition. The present prediction that the fraction of condensed interacting bosons, N_0/N, is nonzero and independent of system size contradicts the ideal boson prediction that the number of condensed (or ground momentum state) bosons is intensive (i.e. the fraction goes to zero in the thermodynamic limit), section 8.2.4. The present prediction is predicated on the binary division approximation. The result is qualitatively consistent with reported experimental results that the fraction is nonzero after the transition, although those experimental measurements give a zero fraction at the λ-transition itself (Donnelly and Barenghi 1998). This matter is further discussed in section 8.6.6. The present mean field results indicate that the ground momentum state occupancy increases continuously through the transition, which prediction will be revisited in the following results for mixed permutation loops, section 8.5.

The difference between the present results for interacting bosons and those for ideal bosons in sections 8.2 and 8.3 shows that the interaction potential has a quantitative effect on the λ-transition. One concludes that although the transition and condensation may be dominated by behavior at long-range, short-range effects are not completely negligible. This raises the possibility that the commutation function may also play a role in these phenomena, and that it would be useful to quantify this. Since the commutation function substantially raises the pressure above the classical prediction for Lennard-Jones ^{4}He, figure 7.5, the saturation liquid density would be lower in an accurate quantum calculation than the present classical value, which would reduce the contribution of the position permutation loops at a given temperature.

8.5 Mixed loops

In order to go beyond the mean field (i.e. no mixing) approach one must include mixed permutation loops that contain both ground and excited momentum state bosons. Such

mixed loops are essential because they provide the nucleating mechanism for ground momentum state condensation. Despite the strong inference based on the experimental and computational evidence that mixed loops nucleate condensation, their very existence poses a surprising conceptual and mathematical challenge.

For ideal bosons, only pure permutation loops are allowed: momentum states cannot be mixed in any one loop. The prohibition on mixing for ideal bosons is due to the orthogonality of the momentum eigenfunctions; the ideal boson configuration integral is the same as the expectation value integral. For interacting bosons this restriction does not apply directly because the interaction potential in the Maxwell–Boltzmann factor means that the configuration integral differs from the expectation value integral in which orthogonality would otherwise hold.

This observation might lead one to conclude that mixed ground and excited momentum permutation loops for interacting bosons are fully permitted. But there is a complication, namely that at large separations the interaction potential goes to zero, which means that when far enough apart the bosons behave ideally with respect to each other. From this point of view mixed ground and excited momentum state loops would be forbidden also for interacting bosons because the large separation regime dominates the configuration integral.

There is an issue to be resolved here that turns on the F London (1938) approximation that separates the discrete ground state from the excited momentum continuum, section 8.2.1, and the nature of the generalized functions in each case. In the discrete momentum picture,

$$\int_V \mathrm{d}\mathbf{q}_{12} \ \mathrm{e}^{-\mathbf{p}_{12}\cdot\mathbf{q}_{12}/\mathrm{i}\hbar} = V\delta_{\mathbf{p}_1,\mathbf{p}_2},$$

(8.4 2)

whereas in the continuum momentum picture

$$\int \mathrm{d}\mathbf{q}_{12} \ \mathrm{e}^{-\mathbf{p}_{12}\cdot\mathbf{q}_{12}/\mathrm{i}\hbar} = (2\pi\hbar)^3\delta(\mathbf{p}_1 - \mathbf{p}_2).$$

(8.4 3)

In these Kronecker and Dirac δ-functions, respectively, appear. Because of this difference, one can only apply the London approximation *after* performing the position integral for the particle density asymptote, as will be demonstrated.

8.5.1 Mixed dimer

In order to account for mixed permutation loops write the symmetrization function as

$$\eta(\mathbf{q},\mathbf{p}) = \eta_0(\mathbf{q}^{N_0},\mathbf{p}^{N_0})\eta_*(\mathbf{q}^{N_*},\mathbf{p}^{N_*})[1 + \eta_{\mathrm{mix}}(\mathbf{q},\mathbf{p})].$$

(8.4 4)

Here $\eta_0 = N_0!$ is the pure ground momentum state symmetrization function, η_* is the pure excited momentum state symmetrization function, and η_{mix} is the sum of mixed loop weights.

The ground momentum state bosons in any one mixed loop are inaccessible to η_0 and their number have to be removed from it. This is most easily done by fixing $\eta_0 = N_0!$ and multiplying each mixed symmetrization loop with n_0 ground momentum state bosons by $(N_0 - n_0)!/N_0!$. The excited loops are compact, and so we can neglect any similar dependence on n_* in the thermodynamic limit.

Here we focus on the leading mixed contribution, which is the mixed dimer,

$$\eta_{0*}^{(1;1)}(\mathbf{q}, \mathbf{p}) = \sum_j^{N_0} \sum_k^{N_*} e^{-\mathbf{p}_k \cdot \mathbf{q}_{kj}/i\hbar}.$$

This is also the dimer chain, $\tilde{\eta}^{(2)}$, and it belongs to the class of singly excited mixed loops, $\eta^{(n;1)}$, both of which are treated below. As mentioned, we can account for the fact that there are now $N_0 - 1$ ground momentum state bosons available for pure loops by dividing this by N_0, which will then leave $\eta_0 = N_0!$ unchanged.

We suppose that we can factorize the classical average of the symmetrization function

$$\left\langle \eta_0 \eta_* \eta_{0*}^{(1;1)}/N_0 \right\rangle_{z,\mathrm{cl}} = \left\langle \eta_0 \eta_* \right\rangle_{z,\mathrm{cl}} \left\langle \eta_{0*}^{(1;1)}/N_0 \right\rangle_{z,\mathrm{cl}}.$$

The first average on the right-hand side is the unmixed product of pure loops, which was dealt with in section 8.4.2.

Using for convenience a canonical rather than a grand canonical average, in the discrete momentum case one can write for the mixed dimer average

$$\left\langle \eta_{0*}^{(1;1)}/N_0 \right\rangle_{N_0, N_*, \mathrm{cl}} = \frac{\Lambda^{3N_*}}{V^{N_*}} \prod_j^{N_0} \delta_{\mathbf{p}_j, 0} \prod_k^{N_*} \sum_{\mathbf{p}_k}^{(p_k > 0)} e^{-\beta p_k^2/2\,m}$$

$$\times \left\langle \frac{1}{N_0} \sum_j^{N_0} \sum_k^{N_*} e^{-\mathbf{p}_k \cdot \mathbf{q}_{kj}/i\hbar} \right\rangle_{N,\mathrm{cl}}$$

$$= \frac{\Lambda^{3N_*}}{V^{N_*}} \frac{N_*}{N(N-1)} \prod_j^{N_0} \delta_{\mathbf{p}_j, 0} \prod_k^{N_*} \sum_{\mathbf{p}_k}^{(p_k > 0)} e^{-\beta p_k^2/2\,m}$$

$$\times \int_V d\mathbf{q}_1 \, d\mathbf{q}_2 \, \rho_N^{(2)}(\mathbf{q}_1, \mathbf{q}_2) e^{-\mathbf{p}_1 \cdot \mathbf{q}_{12}/i\hbar}.$$

The angular brackets represent the configurational average, the momentum average being explicit: the prefactor for the first equality is the normalization for the momentum states, which invokes the continuum integral for the excited states because in this case there is no issue with generalized functions. The second equality converts the configuration average to an integral over the canonical pair density, which is normalized as $\int_V d\mathbf{q}_1 \, d\mathbf{q}_2 \, \rho_N^{(2)}(\mathbf{q}_1, \mathbf{q}_2) = N(N-1)$. In this case particle 1 is in an excited momentum state and particle 2 is in the momentum ground state, although this makes no difference to the configuration integral.

The asymptotic contribution to the configurational integral is

$$\rho^2 \int_V d\mathbf{q}_1 \, d\mathbf{q}_2 \, e^{-\mathbf{p}_1 \cdot \mathbf{q}_{12}/i\hbar} = \rho^2 V \delta_{\mathbf{p}_1, 0}.$$

Since $p_1 > 0$ by design, this vanishes. Therefore, in order to transform the sum over excited states to the integral over the momentum continuum and interchange the order of integration we must first replace the pair density by its connected part, the total correlation function,

$$\rho^2 h_N^{(2)}(\mathbf{q}_1, \mathbf{q}_2) = \rho_N^{(2)}(\mathbf{q}_1, \mathbf{q}_2) - \rho^2.$$

(8.49)

By subtracting the asymptote we ensure that this goes to zero at large separations. Here we assume a homogeneous system, $\rho_N^{(1)}(\mathbf{q}_1) = \rho = N/V$, in which case the total correlation function is a function only of the particle separation, $h_N^{(2)}(\mathbf{q}_1, \mathbf{q}_2) = h_N^{(2)}(q_{12})$.

Subtracting the asymptote we are left with a short-ranged integrand and therefore no generalized function. As we have just seen, the asymptotic contribution is zero when integrated over position configurations for fixed nonzero momentum, and so we are permitted to subtract it. Having subtracted the asymptote, we may transform to the momentum continuum, interchange the order of integration, and perform the integral over the excited states. These give

$$\begin{aligned}
\left\langle \eta_{0*}^{(1;1)}/N_0 \right\rangle_{N_0, N_*, \mathrm{cl}} &= \frac{\Lambda^{3N_*}}{V^{N_*}} \frac{N_*}{N(N-1)} \prod_j^{N_0} \delta_{\mathbf{p}_j, 0} \prod_k^{N_*} \sum_{\mathbf{p}_k}^{(p_k > 0)} e^{-\beta p_k^2 / 2\, m} \\
&\quad \times \int d\mathbf{q}_1\, d\mathbf{q}_2\, \rho^2 h_N^{(2)}(\mathbf{q}_1, \mathbf{q}_2) e^{-\mathbf{p}_1 \cdot \mathbf{q}_{12}/i\hbar} \\
&= \frac{\Lambda^{3N_*}}{V^{N_*}} \frac{N_*}{N(N-1)} \Delta_p^{-3N_*} \int d\mathbf{p}^{N_*}\, e^{-\beta \mathcal{K}(\mathbf{p}^{N_*})} \\
&\quad \times \int d\mathbf{q}_1\, d\mathbf{q}_2\, \rho^2 h_N^{(2)}(\mathbf{q}_1, \mathbf{q}_2) e^{-\mathbf{p}_1 \cdot \mathbf{q}_{12}/i\hbar} \\
&= \frac{\rho^2 N_*}{N(N-1)} \int d\mathbf{q}_1\, d\mathbf{q}_2\, h_N^{(2)}(\mathbf{q}_1, \mathbf{q}_2) e^{-\pi q_{12}^2 / \Lambda^2} \\
&= \frac{N_*}{V} \int d\mathbf{q}_{12}\, h_N^{(2)}(q_{12}) e^{-\pi q_{12}^2 / \Lambda^2}.
\end{aligned}$$

(8.50)

This can also be written

$$\begin{aligned}
\left\langle \eta_{0*}^{(1;1)}/N_0 \right\rangle_{N_0, N_*, \mathrm{cl}} &= \frac{N_* V}{N^2} \int d\mathbf{q}_{12}\, \rho_N^{(2)}(q_{12}) e^{-\pi q_{12}^2 / \Lambda^2} - \frac{N_* V}{N^2} \rho^2 \Lambda^3 \\
&= \frac{N_*}{N} \left\langle \frac{2}{N} \sum_{j<k}^N e^{-\pi q_{jk}^2 / \Lambda^2} \right\rangle_{N, \mathrm{cl}} - \frac{N_* V}{N^2} \rho^2 \Lambda^3 \\
&= \frac{N_*}{N} \left\langle \frac{1}{N} \sum_{j,k}^N {}' \left[e^{-\pi q_{jk}^2 / \Lambda^2} - \frac{\rho \Lambda^3}{N} \right] \right\rangle_{N, \mathrm{cl}} \\
&\equiv \frac{N_*}{N} \left\langle \tilde{\eta}^{(2)}/N \right\rangle_{N, \mathrm{cl}}^{\mathrm{corr}}.
\end{aligned}$$

(8.51)

The final average, which is in a canonical classical system of N particles, has the asymptotic correction as defined by the preceding equality. The result is intensive.

For high temperatures such that thermal wavelength is less than about the core diameter, $\Lambda \lesssim \sigma$, the integrand of the first equality vanishes because the factors are nowhere simultaneously nonzero: the density vanishes in the core, $\rho_N^{(2)}(q_{12}) = 0$, $q_{12} \lesssim \sigma$,

and the Gaussian vanishes beyond the core, $\mathrm{e}^{-\pi q_{12}^2/\Lambda^2} \approx 0$, $q_{12} \lesssim \sigma$. Hence at high temperatures $\langle \eta_{0*}^{(1;1)}/N_0 \rangle_{N_0,N_*,\mathrm{cl}} \sim -\rho\Lambda^3$. That this is negative means that such mixed dimers are entropically unfavorable in the high temperature regime. This suppresses ground momentum state occupation, which will turn out to be significant.

8.5.2 Singularly exciting mixed loops

8.5.2.1 Analysis
Consider mixed l-loops with one excited momentum state boson, labeled 1 or j_1, and $l-1$ ground momentum state bosons, labeled $2, \ldots, l$ or $k_2, \ldots, k_l$, whose weight we shall denote as $\eta_{0*}^{(l-1;1)}$. A particular such loop with particle 1 excited has symmetrization factor $\mathrm{e}^{-\mathbf{p}_1 \cdot \mathbf{q}_{12}/\mathrm{i}\hbar}$, which is independent of the positions of all the ground momentum state bosons in the loop, $\mathbf{q}_3, \mathbf{q}_4, \ldots, \mathbf{q}_l$, except the adjacent one labeled 2. Therefore, we can arrange them in $(l-2)!$ ways without changing the value of the symmetrization factor. The total weight involving all such singly excited mixed loops is

$$
\begin{aligned}
\eta_0 \, \eta_* \eta_{0*}^{(l-1;1)} &= (N_0 - l + 1)! \, \eta_*(N_*) \sum_{k_2,\ldots,k_l}^{N_0} {}' \sum_{j_1}^{N_*} \mathrm{e}^{-\mathbf{p}_{j_1} \cdot \mathbf{q}_{j_1,k_2}/\mathrm{i}\hbar} \\
&= (N_0 - l + 1)! \, \eta_*(N_*) \frac{(N_0 - 1)!}{(N_0 - l + 1)!} \sum_{k_2}^{N_0} \sum_{j_1}^{N_*} \mathrm{e}^{-\mathbf{p}_{j_1} \cdot \mathbf{q}_{j_1,k_2}/\mathrm{i}\hbar} \\
&= \eta_0(N_0) \, \eta_*(N_*) \frac{1}{N_0} \sum_{k_2}^{N_0} \sum_{j_1}^{N_*} \mathrm{e}^{-\mathbf{p}_{j_1} \cdot \mathbf{q}_{j_1,k_2}/\mathrm{i}\hbar}.
\end{aligned}
$$

(8.52)

The mixed factor is obviously intensive (i.e. independent of N_0 and of N_*) in the thermodynamic limit, since for each ground momentum state boson there is a limited number of excited momentum state bosons in the neighborhood that will give a nonzero weight after averaging.

The singularly excited mixed loop phase function defined by the final equality here, $\eta_{0*}^{(l-1;1)}(\mathbf{q}, \mathbf{p})$, is identical to the mixed dimer phase function, equation (8.45). Hence its average is also the same. Importantly, it is independent of l, which means that their total contribution is

$$
\begin{aligned}
-\beta\Omega_{\mathrm{mix}}^{(1)} &= \sum_{l=2}^{N_0+1} \left\langle \eta_{0*}^{(l-1;1)}/N_0 \right\rangle_{N_0,N_*,\mathrm{cl}} \\
&= N_0 \left\langle \eta_{0*}^{(1;1)}/N_0 \right\rangle_{N_0,N_*,\mathrm{cl}} \\
&= \frac{N_0 N_*}{N} \left\langle \tilde{\eta}^{(2)}/N \right\rangle_{N,\mathrm{cl}}^{\mathrm{corr}},
\end{aligned}
$$

(8.53)

where the final factor on the right-hand side is given in equation (8.51). This is the total from all mixed symmetrization loops that have a single excited momentum state boson.

One can write the grand potential as $\Omega = \Omega_{\mathrm{cl}} + \Omega_* + \Omega_{\mathrm{mix}}$, with the first two terms being given by equations (8.38) and (8.39), respectively. The present $\Omega_{\mathrm{mix}}^{(1)}$ is the leading order contribution to the mixed term. Its derivative at constant N is

$$
\left(\frac{\partial\left(-\beta\Omega_{\mathrm{mix}}^{(1)} \right)}{\partial N_*} \right)_N = \frac{N - 2N_*}{N} \left\langle \tilde{\eta}^{(2)}/N \right\rangle_{N,\mathrm{cl}}^{\mathrm{corr}}.
$$

(8.54)

The average is independent of N_*. The prefactor is negative for $N_* > N/2$, which is the case at high temperatures. As mentioned above, the average of the mixed dimer must be negative for high temperatures. These two facts mean that at high temperatures this derivative is positive, which increases the occupation of the excited states compared to the classical and pure terms alone.

8.5.2.2 Numerical results

Simulation results for the mixed dimer are shown in table 8.3. The difference in the Lennard-Jones saturation density compared to that in table 8.2 is partly due to the larger system size and lower overall density here, which means that the periodic boundary conditions have less influence on the shape of the liquid phase. The primary difference is that here the averages were taken over a sphere of radius 10σ about the center of mass, whereas in table 8.2 they were taken over the whole system, of which approximately 20% of the bosons were in the vapor phase or interfacial region. For the present larger system at $T^* = 0.5$, the radius of the interface at half density is $\approx 10.6\sigma$. The central density measured within 2σ of the center of mass is $\rho\sigma^3 = 0.943(1)$, whereas the density measured within 10σ is $\rho\sigma^3 = 0.911\,6(5)$. For $T^* \lesssim 0.5$ the system was somewhat glassy, with limited or no macroscopic diffusion of the particles.

Table 8.3. Liquid density, mixed dimer weight, and optimum ground momentum state fraction for saturated Lennard-Jones He^4 at various temperatures, $T^* \equiv k_{\mathrm{B}}T/\varepsilon$. From liquid drop Monte Carlo simulations with $N = 5000$, overall density of $\rho\sigma^3 = 0.2$, using the central liquid volume of radius 10σ for the averages.

T^*	$\rho\sigma^3$	$\rho\Lambda^3$	$\dfrac{\langle\tilde{\eta}^{(2)}\rangle_{N,\mathrm{cl}}^{\mathrm{corr}}}{N}$	$\dfrac{\bar{N}_0}{N}$	$\dfrac{\bar{N}_0}{N}$
				stable	unstable
1.00	0.67	0.81	-0.60	0	–
0.90	0.73	1.05	-0.70	0	–
0.80	0.79	1.34	-0.79	0	–
0.70	0.83	1.73	-0.87	0	–
0.65	0.86	1.99	-0.90	0	–
0.60	0.88	2.29	-0.93	0.640	0.35
0.55	0.90	2.67	-0.96	0.754	0.35
0.50	0.91	3.14	-0.99	0.818	0.35

T^*	$\rho\sigma^3$	$\rho\Lambda^3$	$\dfrac{\langle\tilde{\eta}^{(2)}\rangle_{N,\mathrm{cl}}^{\mathrm{corr}}}{N}$	$\dfrac{\bar{N}_0}{N}$	$\dfrac{\bar{N}_0}{N}$
0.45	0.97	3.93	-1.04	0.872	0.45
0.40	0.99	4.74	-1.06	0.902	0.55

It can be seen in table 8.3 that the mixed dimer weight is negative at all temperatures. This means that the asymptotic correction dominates. The weight increases in magnitude with decreasing temperature, although the dependence on temperature is rather weak.

The mixed dimer weight was combined with the classical pure loop grand potential, $\Omega = \Omega_{\mathrm{cl}} + \Omega_{\mathrm{pure}} + \Omega_{\mathrm{mix}}^{(1)}$, with the first two terms being given by equations (8.38) and (8.39), and the singly excited mixed grand potential by equation (8.53). The pure loops used $l_{\max} = 4$ for $T^* \geqslant 0.7$ and $l_{\max} = 5$ for $T^* \leqslant 0.65$. The results were little different when $l_{\max}$ was increased by one. The optimum fraction of bosons in the ground momentum state, $\bar{N}_0/N$, corresponds to the vanishing of the derivative of the grand potential, equation (8.41) plus equation (8.54). It can be seen in table 8.3 that for $T^* \geqslant 0.65$ there is only one stable solution, namely all the bosons in the system are in excited states. For temperatures $T^* \leqslant 0.6$ (6.13 K), there are two zeros for the derivative, the higher fraction being the stable solution (i.e. the minimum in the grand potential). At $T^* = 0.6$ the stable fraction of ground momentum state bosons is $\bar{N}_0/N = 0.640$, which is a rather abrupt change from $\bar{N}_0/N = 0$ at $T^* = 0.65$. It would be fair to call this ground momentum state condensation. The transition more or less coincides with the passage of $\bar{N}_0/N$ from greater than one half to less than one half, as given by the pure loops only in table 8.2. At $T^* = 0.6$, the value $\bar{N}_0/N = 0.640$ obtained by including the mixed dimer is greater than the value $\bar{N}_0/N = 0.536$ in table 8.2 obtained with the pure loops only, as is expected for a negative values of $\langle\tilde{\eta}^{(2)}\rangle_{N,\mathrm{cl}}^{\mathrm{corr}}$ (cf the remarks following equation (8.54)).

In table 8.3 it can be seen that $\langle\tilde{\eta}^{(2)}\rangle_{N,\mathrm{cl}}^{\mathrm{corr}}/N > -1$ for $T^* > 0.45$. Obviously there is some uncertainty in where it first exceeds this bound as it is the difference between two positive comparable quantities, and it therefore has a larger relative error than either alone. Minus one is significant because it marks the point where $1 + (N_*/N)\langle\tilde{\eta}^{(2)}\rangle_{N,\mathrm{cl}}^{\mathrm{corr}}/N$, becomes negative if the system is fully excited, $N_* = N$. At the dimer level of approximation, beyond this point it is not possible to have a fully excited system (cf dressed bosons below in section 8.5.3).

In any case the present mixed loop calculations provide a credible mechanism for condensation. (Within the present binary division approximation, the condensation is necessarily into the ground momentum state.) The unmixed results in table 8.2 show condensation that grows continuously from zero at $\rho\Lambda^3 = 1$. In the present mixed case the condensation is rather sudden, and it occurs at $\rho\Lambda^3 = 2.29$. Since mixed loops are forbidden for ideal bosons this effect is specific to interacting bosons. It shows that the growth of excited state position permutation loops nucleate, or are responsible for, condensation (Attard 2021 section 5.6.1.7). Such coupling and nucleation means that for interacting bosons the number of condensed bosons must be extensive, just like the number of excited momentum state bosons. Again the present binary division approximation means that the condensation is necessarily into the ground momentum state. The issue of the extensivity of the ground momentum state is discussed in sections 8.2.4 and 8.6.6.

8.5.3 Dressed bosons

In the preceding subsection the non-local nature of the ground momentum state was exploited to incorporate an arbitrary number of ground momentum state bosons into the mixed dimer permutation loop to obtain the so-called singly excited grand potential as the leading order contribution from mixed ground and excited momentum state permutation loops. In this subsection this idea is generalized to form mixed permutation loops by concatenating permutation chains, which consist of a ground momentum state boson as the head and excited momentum state bosons as the tail. The series of such chains may be called a dressed ground momentum state boson, or dressed boson for short. The theory based on them is like the ideal solution theory of physical chemistry, with the dressed ground momentum state bosons being the solute that forms a dilute solution in the fluid of excited momentum state bosons.

Consider an *l*-chain, with the ground momentum state boson at the head, which we designate as position *l*, and with $l-1$ excited momentum state bosons forming the tail. A particular chain has symmetrization function

$$\tilde{\eta}^{(l)}_{j_1,\ldots,j_l} = \mathrm{e}^{-\mathbf{p}_{j_1}\cdot\mathbf{q}_{j_1,j_2}/\mathrm{i}\hbar}\mathrm{e}^{-\mathbf{p}_{j_2}\cdot\mathbf{q}_{j_2,j_3}/\mathrm{i}\hbar}\cdots\mathrm{e}^{-\mathbf{p}_{j_{l-1}}\cdot\mathbf{q}_{j_{l-1},j_l}/\mathrm{i}\hbar}.$$

(8.55)

The sum of all possible dimer chains is the same as the mixed dimer symmetrization function, equation (8.45),

$$\eta^{(1;1)}_{0*}(\mathbf{q},\mathbf{p}) = \sum_{j}^{N_0}\sum_{k}^{N_*}\mathrm{e}^{-\mathbf{p}_k\cdot\mathbf{q}_{kj}/\mathrm{i}\hbar}$$

$$= \sum_{j}^{N_0}\sum_{k}^{N_*}\tilde{\eta}^{(2)}_{k,j}.$$

(8.56)

The average of this correcting for the asymptote is given as equation (8.51).

The sum of all possible trimer chains is the same as the mixed trimer symmetrization function with one ground and two excited momentum state bosons, equation (A.1) of Attard (2022),

$$\eta^{(1;2)}_{0*}(\mathbf{q},\mathbf{p}) = \sum_{j}^{N_0}\sum_{k,l}^{N_*}{}'\,\mathrm{e}^{-\mathbf{p}_k\cdot\mathbf{q}_{kl}/\mathrm{i}\hbar}\mathrm{e}^{-\mathbf{p}_l\cdot\mathbf{q}_{lj}/\mathrm{i}\hbar}$$

$$= \sum_{j}^{N_0}\sum_{k,l}^{N_*}{}'\,\tilde{\eta}^{(3)}_{k,l,j}.$$

(8.57)

The prime on the sum indicates that no two indeces may be equal. The average of this correcting for the asymptote is given as equations (A.5) and (A.6) of Attard (2022),

$$\left\langle \eta_{0*}^{(1;2)}/N_0 \right\rangle_{N_0,N_*,\mathrm{cl}} = \frac{N_*(N_*-1)}{N(N-1)(N-2)}\left\{ \rho^3 \int \mathrm{d}\mathbf{q}_1 \, \mathrm{d}\mathbf{q}_2 \, \mathrm{d}\mathbf{q}_3 \, h_N^{(3)}(\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3)\mathrm{e}^{-\pi q_{12}^2/\Lambda^2}\mathrm{e} \right.$$

$$\left. + \, 2^{-3/2}\rho^3 V\Lambda^3 \int_V \mathrm{d}\mathbf{q}_{13} \, h_N^{(2)}(q_{13})\mathrm{e}^{-\pi q_{13}^2/2\Lambda^2} \right\}$$

$$= \frac{N_*^2}{N^2}\left\langle \frac{1}{N}{\sum_{j,k,l}}' \left[\mathrm{e}^{-\pi q_{jk}^2/\Lambda^2} - \frac{\rho\Lambda^3}{N}\right]\left[\mathrm{e}^{-\pi q_{kl}^2/\Lambda^2} - \frac{\rho\Lambda^3}{N}\right] \right\rangle_{N,\mathrm{cl}}$$

$$\equiv \frac{N_*^2}{N^2}\left\langle \tilde{\eta}^{(3)}/N \right\rangle_{N,\mathrm{cl}}^{\mathrm{corr}}.$$

(8.58)

For a ground state boson k, chains of a given length may be summed over all excited momentum state bosons,

$$\tilde{\eta}_k^{(l)}(\mathbf{q},\mathbf{p}) = \sum_{j_1,\ldots,j_{l-1}}{}' \, \tilde{\eta}_{j_1,\ldots,j_{l-1},k}^{(l)}(\mathbf{q},\mathbf{p}).$$

(8.59)

These in turn may be summed over all lengths

$$\tilde{\eta}_k(\mathbf{q},\mathbf{p}) = \sum_{l=2}^{N_*+1} \tilde{\eta}_k^{(l)}(\mathbf{q},\mathbf{p}).$$

(8.60)

The quantity $\tilde{\eta}_k(\mathbf{q},\mathbf{p})$ is the symmetrization weight of the dressed ground momentum state boson k at this point in phase space. We may average this weight, either before or after summing over chain length, and treat the dressed ground momentum state bosons as independent, which is valid if they are dilute compared to the excited momentum state bosons, $N_0(l_{\max}-1) \ll N_*$.

One can form $N_0!$ permutations of the dressed bosons. But for this to work one needs to show that the weight upon concatenation of two chains into a permutation loop is the same as the product of the weights of the two individual permutation chains. To this end, consider an l-chain $j_1, j_2, \ldots, j_l$ and an m-chain, $k_1, k_2, \ldots, k_m$. The symmetrization function for the permutation loop formed from them is

$$\tilde{\eta}_{j^l,k^m}^{(l+m)}(\mathbf{q},\mathbf{p}) = \mathrm{e}^{-\mathbf{p}_{j_1}\cdot\mathbf{q}_{j_1,j_2}/\mathrm{i}\hbar}\cdots\mathrm{e}^{-\mathbf{p}_{j_{l-1}}\cdot\mathbf{q}_{j_{l-1},j_l}/\mathrm{i}\hbar}$$

$$\times \, \mathrm{e}^{-\mathbf{p}_{k_1}\cdot\mathbf{q}_{k_1,k_2}/\mathrm{i}\hbar}\cdots\mathrm{e}^{-\mathbf{p}_{k_{m-1}}\cdot\mathbf{q}_{k_{m-1},k_m}/\mathrm{i}\hbar}$$

$$= \tilde{\eta}_{j^l}^{(l)}(\mathbf{q},\mathbf{p})\,\tilde{\eta}_{k^m}^{(m)}(\mathbf{q},\mathbf{p}).$$

(8.61)

The two transposition factors that link the chains, $\mathrm{e}^{-\mathbf{p}_{j_l}\cdot\mathbf{q}_{j_l,k_1}/\mathrm{i}\hbar}$ and $\mathrm{e}^{-\mathbf{p}_{k_m}\cdot\mathbf{q}_{k_m,j_1}/\mathrm{i}\hbar}$, are unity because the head bosons are in the ground momentum state, $\mathbf{p}_{j_l} = \mathbf{p}_{k_m} = 0$. Hence the permutation loop function formed by concatenating the two chains is just the product of the two chain functions, with the restriction that no index can be the same. Since the chains are compact, and since it is assumed that $N_0 \ll N_*$ (dilute solution), this restriction can be dropped with negligible error. From this it follows that the $N_0!$ permutations of

particular chains form legitimate permutation loops, and that all $N_0!$ permutations have the same total weight, which is equal to the product of the weights of the individual chains.

The average weight of a dressed ground momentum state boson is

$$\langle \tilde{\eta}(\mathbf{q}, \mathbf{p}) \rangle_{N_0, N_*, \mathrm{cl}} = \left\langle \frac{1}{N_0} \sum_{k}^{N_0} \tilde{\eta}_k(\mathbf{q}, \mathbf{p}) \right\rangle_{N_0, N_*, \mathrm{cl}}$$

$$= \sum_{l=2}^{\infty} \left\langle \frac{1}{N_0} \sum_{k}^{N_0} \tilde{\eta}_k^{(l)}(\mathbf{q}, \mathbf{p}) \right\rangle_{N_0, N_*, \mathrm{cl}}$$

$$= \sum_{l=2}^{\infty} \frac{N_*^{l-1}}{N^l} \left\langle \tilde{G}^{(l)} \right\rangle_{N,\mathrm{cl}}^{\mathrm{corr}}$$

$$= \sum_{l=2}^{\infty} \frac{N_*^{l-1}}{N^{l-1}} \tilde{g}^{(l)}.$$

(8.62)

The $l = 2$ term is the mixed dimer, equation (8.51), and the $l = 3$ term is the mixed trimer, equation (8.58). The chain Gaussian is

$$\tilde{G}^{(l)}(\mathbf{q}^N) = \sum_{j_1, \ldots, j_l}{}' \prod_{k=1}^{l-1} \mathrm{e}^{-\pi q_{j_k, j_{k+1}}/\Lambda^2}.$$

(8.63)

The prime indicates that no two indices may be equal.

The product of the dressed weights of all the bosons gives the mixed grand partition function,

$$\tilde{\Xi} = 1 + \eta_{\mathrm{mix}} = \left[1 + \langle \tilde{\eta}(\mathbf{q}, \mathbf{p}) \rangle_{N_0, N_*, \mathrm{cl}} \right]^{N_0},$$

(8.64)

where the bare boson (monomer chain), $\tilde{\eta}^{(1)} = 1$, appears explicitly. The logarithm gives the mixed grand potential

$$-\beta \tilde{\Omega} = N_0 \ln \left[1 + \langle \tilde{\eta}(\mathbf{q}, \mathbf{p}) \rangle_{N_0, N_*, \mathrm{cl}} \right].$$

(8.65)

The total grand potential is $\Omega = \Omega_{\mathrm{cl}} + \Omega_* + \tilde{\Omega}$. Implicitly, $N_0 = \bar{N}_0(z, V, T)$ and $N_* = \bar{N}_*(z, V, T)$, although as a variational formulation this is a second order effect. The $N_0!$ permutations of the chains, which is the pure ground momentum state symmetrization weight, cancels with the same factor in the denominator of the partition function. The total weight of the ground momentum state bosons is the product of their average dressed weight, assuming that they do not overlap.

If one linearizes the expression for the mixed grand potential one obtains

$$-\beta\tilde{\Omega} = N_0 \sum_{l=2}^{\infty} \left\langle \frac{1}{N_0} \sum_{k}^{N_0} \tilde{\eta}_k^{(l)}(\mathbf{q}, \mathbf{p}) \right\rangle_{N_0, N_*, \mathrm{cl}}$$

$$= N_0 \sum_{l=2}^{\infty} \frac{N_*^{l-1}}{N^{l-1}} \tilde{g}^{(l)}$$

$$= -\beta \sum_{l=2}^{\infty} \Omega_{\mathrm{mix}}^{(l-1)}.$$

(8.66)

The dimer term is just equation (8.53), and the trimer term is just equation (A.7) of Attard (2022). This shows how the present dressed boson theory is the non-linear generalization of the singly and doubly excited loops of section 8.5.1 and appendix A of Attard (2022).

The derivative of the non-linear mixed grand potential is

$$\left(\frac{\partial\left(\beta\tilde{\Omega}\right)}{\partial N_*} \right)_N = -\ln\left[1 + \sum_{l=2}^{\infty} \frac{N_*^{l-1}}{N^{l-1}} \tilde{g}^{(l)} \right] + \frac{\frac{N_0}{N} \sum_{l=2}^{\infty} (l-1) \frac{N_*^{l-2}}{N^{l-2}} \tilde{g}^{(l)}}{1 + \sum_{l=2}^{\infty} \frac{N_*^{l-1}}{N^{l-1}} \tilde{g}^{(l)}}.$$

(8.67)

Add this to the unmixed result, equation (8.41), to obtain the fraction of excited momentum state bosons in the system. One can also linearize this.
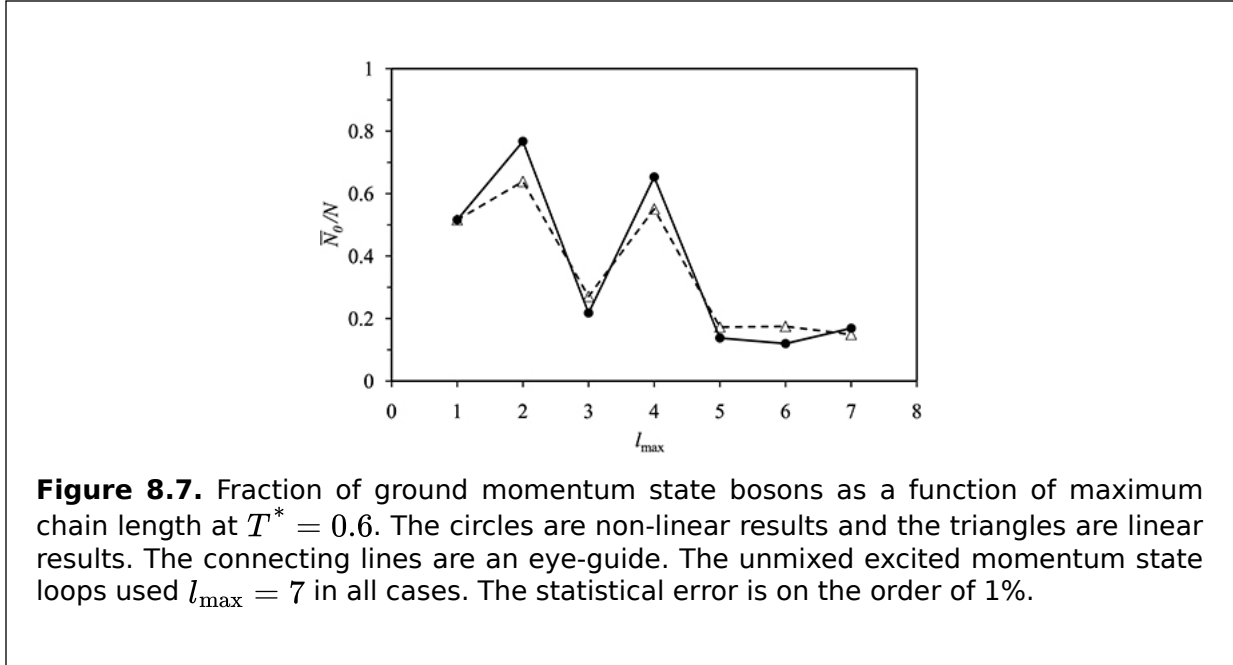
Table 8.4 shows results for chains up to $l_{\max} = 4$. That the chain weights alternate in sign poses convergence problems. For $T^* \lesssim 0.55$ it is not clear that the chain series is converging (although of course each term must be multiplied by $(N_*/N)^{l-1}$). Retaining the first four terms in the series (including the monomer term of unity) gives a result for the optimum fraction of ground momentum state bosons in broad agreement with those predicted retaining two terms only, table 8.3. The suppression of the occupation of the ground momentum state for $T^* \geqslant 0.65$ is likewise clear in the two tables. There is little difference between the linear and non-linear results, which perhaps suggests that the linear theory is exact. The location of the condensation, $T^* = 0.6$, $\rho\Lambda^3 = 2.29$, is relatively insensitive to the level of the theory.

Table 8.4. Chain weights, $\tilde{g}^{(l)} = \langle\tilde{\eta}^{(l)}\rangle_{N,\mathrm{cl}}^{\mathrm{corr}}/N$, and optimum fraction of ground momentum state bosons for saturated Lennard-Jones He^4 at various temperatures.

T^*	$\frac{\langle\tilde{\eta}^{(2)}\rangle_{N,\mathrm{cl}}^{\mathrm{corr}}}{N}$	$\frac{\langle\tilde{\eta}^{(3)}\rangle_{N,\mathrm{cl}}^{\mathrm{corr}}}{N}$	$\frac{\langle\tilde{\eta}^{(4)}\rangle_{N,\mathrm{cl}}^{\mathrm{corr}}}{N}$	$\frac{\bar{N}_0}{N}$	$\frac{\bar{N}_0}{N}$
				linear	non-linear
1.00	−0.599	0.357	−0.213	0	0
0.90	−0.700	0.486	−0.339	0	0
0.80	−0.790	0.613	−0.485	0	0
0.70	−0.866	0.728	−0.650	0	0

T^*	$\dfrac{\langle\tilde{\eta}^{(2)}\rangle_{N,\mathrm{cl}}^{\mathrm{corr}}}{N}$	$\dfrac{\langle\tilde{\eta}^{(3)}\rangle_{N,\mathrm{cl}}^{\mathrm{corr}}}{N}$	$\dfrac{\langle\tilde{\eta}^{(4)}\rangle_{N,\mathrm{cl}}^{\mathrm{corr}}}{N}$	$\dfrac{\bar{N}_0}{N}$	$\dfrac{\bar{N}_0}{N}$
0.65	−0.904	0.782	−0.754	0	0.527
0.60	−0.934	0.820	−0.875	0.560	0.658
0.55	−0.963	0.852	−1.058	0.677	0.745
0.50	−0.992	0.889	−1.398	0.771	0.812
0.45	−1.038	1.089	−2.496	0.845	0.868
0.40	−1.063	1.180	−4.069	0.887	0.900

Figure 8.7 shows the optimum fraction of ground momentum state bosons as a function of the maximum chain length at $T^* = 0.6$. There is clearly an even/odd effect for $l_{\max} \leqslant 5$, but for larger values the fraction appears to have converged to $\bar{N}_0/N \approx 0.15$. Condensation appears to have occurred by $T^* = 0.6$, although for individual choices of $l_{\max}$ it can occur by $T^* = 0.65$ (cf table 8.4 for $l_{\max} = 4$, non-linear).

Figure 8.7. Fraction of ground momentum state bosons as a function of maximum chain length at $T^* = 0.6$. The circles are non-linear results and the triangles are linear results. The connecting lines are an eye-guide. The unmixed excited momentum state loops used $l_{\max} = 7$ in all cases. The statistical error is on the order of 1%.

Similar behavior as a function of $l_{\max}$ occurred for $T^* = 0.62$, 0.60, and 0.55. Condensation had occurred for almost all $l_{\max}$, linear and non-linear; the two exceptions in the 42 cases calculated happened at low $l_{\max}$. Using $l_{\max} = 7$, the following ground momentum state bosons fractions were found: For $T^* = 0.62$, $\bar{N}_0/N = 0.148\ 0(3)$ (linear) and 0.157 7(18) (non-linear). For $T^* = 0.6$, $\bar{N}_0/N = 0.148\ 7(7)$ (linear) and 0.169 5(23) (non-linear). And for $T^* = 0.55$ $\bar{N}_0/N = 0.140\ 5(4)$ (linear) and 0.198 1(14) (non-linear). Since the change in the condensed fraction over these temperature intervals is negligible compared to the change from zero at $T^* = 0.65$, one can conclude that the

condensation transition is discontinuous. The transition occurs somewhere in the range $0.62 \leqslant T_{\mathrm{c}}^{*} < 0.65$.

Experimentally, the fraction of condensed bosons (i.e. He II) inferred from viscosity measurements at the λ-transition is zero (Donnelly and Barenghi 1998).

8.5.4 Energy and heat capacity in the condensed regime

In these results there is either no increase or else a slow increase in the fraction of ground momentum state bosons with decreasing temperature after the transition. The slowness of the increase appears to be an artefact that arises from assuming that the dressed bosons do not interfere with each other. The condition for non-interference, $N_0(l_{\mathrm{max}} - 1) \ll N_*$, holds barely in the present case, if it holds at all; dividing through by N, the left-hand side is typically $0.15 \times 5 = 0.75$, whereas the right-hand side is 0.85. If interference were taken into account the contribution from mixed loops would be reduced (because many loops currently counted would be prohibited). Since the mixed loop contribution suppresses the number of ground momentum state bosons compared to the pure loop prediction, table 8.2, interference between the mixed loops would reduce their negative contribution and hence increase the number of ground momentum state bosons above that calculated above. This means that the pure loop result becomes increasingly accurate below the condensation transition, as is its prediction that the ground momentum state is increasingly occupied as the temperature is decreased, table 8.2.

This discussion leads to the conclusion that the mixed loop theory is probably accurate for predicting the location of the condensation transition, but that it underestimates the number of ground momentum state bosons, and its rate of increase with decreasing temperature, once condensation has occurred. The pure loop theory is likely more accurate in the condensed regime. In calculating the effect of condensation on the energy and on the heat capacity, our strategy will be to use the mixed loop theory to locate the transition, and the pure loop theory to estimate the number of ground momentum state bosons in the condensed regime.

The use of different theories on either side of the transition is not entirely satisfactory. It leads directly to the energy discontinuity and latent heat that is now discussed. These ultimately are an artefact of the binary division approximation (see sections 8.6.6, 9.3, and 9.2).

The average energy is the sum of loop energies, $\bar{\mathcal{H}} = \sum_{l=1}^{\infty} \bar{\mathrm{E}}^{(l)}$. The averages are classical. One can relate each average loop energy for a system of $\bar{N}_0$ ground momentum state bosons and $\bar{N}_* = N - \bar{N}_0$ excited momentum state bosons, $\bar{\mathrm{E}}^{(l)}(\bar{N}_0, \bar{N}_*)$, to one obtained in a classical system of N bosons, $\bar{\mathrm{E}}^{(l)}(N)$, as follows. The classical or monomer contribution is the sum of the classical kinetic energy and the potential energy,

$$\bar{\mathrm{E}}^{(1)}(\bar{N}_0, \bar{N}_*) = \frac{3}{2}\bar{N}_* k_{\mathrm{B}}T + \langle U \rangle_{N,V,T}^{\mathrm{cl}}.$$
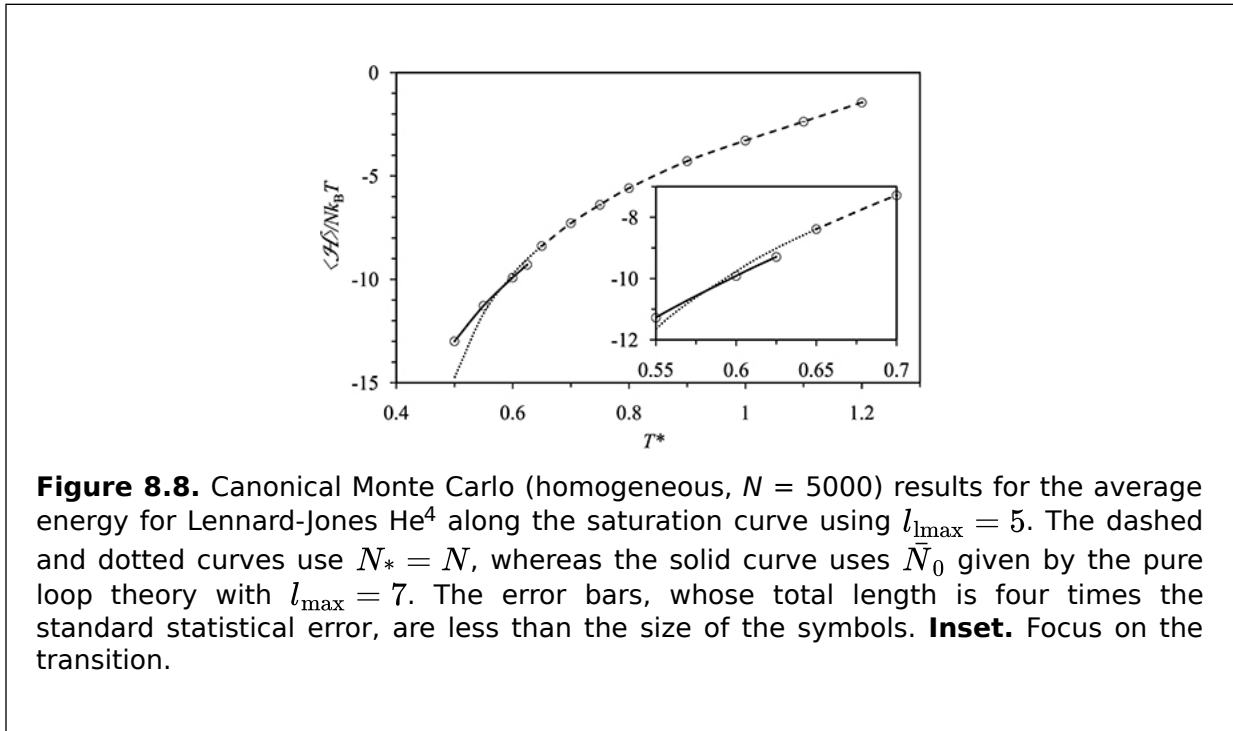
(8.68)

Obviously only excited momentum state bosons contribute to the kinetic energy. The potential energy does not distinguish ground and excited momentum state bosons. The dimer and higher loop energies are

$$\bar{\mathrm{E}}^{(l)}(\bar{N}_0, \bar{N}_*) = \left(\frac{\bar{N}_*}{N} \right)^{l} \bar{\mathrm{E}}^{(l)}(N), \quad l \geqslant 2.$$

(8.69)

Above the condensation transition there are only excited momentum state bosons, $\bar{N}_* = N$, which means that the two systems are the same.

The average energy resulting from these formulae is shown in figure 8.8. These data are from homogeneous canonical Monte Carlo simulations along the saturated liquid density curve using the densities from table 8.3. The condensation transition was set at $T_c^* = 0.625$, as determined by the mixed loop theory described in the preceding subsection. The fraction of ground momentum state bosons below the condensation transition was determined by the unmixed theory of section 8.4.2 at the heptamer level of approximation. One can see in figure 8.8 that the condensed branch energy (solid curve) does not coincide with the non-condensed branch extrapolated to lower temperatures (dotted curve). At the highest condensed temperature in the figure, $T^* = 0.625$, the value of the extrapolated energy that neglects the difference between ground and excited momentum state bosons is $\beta\langle\mathcal{H}\rangle/N = -9.017(2)$, whereas the actual condensed energy is -9.295. Hence the latent heat for the Lennard-Jones condensation transition is $E_{\text{latent}} = 0.3Nk_{\text{B}}T$, which is 3% of the total energy. This surprisingly small value results from three effects: first, the kinetic energy, which is positive, is lower in magnitude in the condensed regime. Second, the loop energies, which are negative, are also lower in magnitude in the condensed regime. And third, the classical potential energy, $\beta\bar{U}^{(1)}/N = -9.958(1)$, which is 90% of the total, is unchanged. Because at the transition the fraction of ground momentum bosons is on the order of 50% (according to the unmixed theory), each individual change in energy is large relative to its contribution to the total. But their relative contribution to the total is only about 10%, and the two changes partially cancel each other. This leaves a residual latent heat of 3% of the total energy.

Figure 8.8. Canonical Monte Carlo (homogeneous, $N = 5000$) results for the average energy for Lennard-Jones He^4 along the saturation curve using $l_{\text{lmax}} = 5$. The dashed and dotted curves use $N_* = N$, whereas the solid curve uses $\bar{N}_0$ given by the pure loop theory with $l_{\text{max}} = 7$. The error bars, whose total length is four times the standard statistical error, are less than the size of the symbols. **Inset.** Focus on the transition.

This predicted latent heat is entirely dependent on the simple binary division into ground or else the set of excited momentum states. There is reason to believe that condensation actually represents the shift to pure momentum loops and increased occupancy of low-lying momentum states rather than just the ground momentum state

(see sections 8.6.6, 9.3, and 9.2). These suggest that the present discontinuity in ground momentum state occupancy and the consequent latent heat for the λ-transition may be an artefact of the binary division approximation.
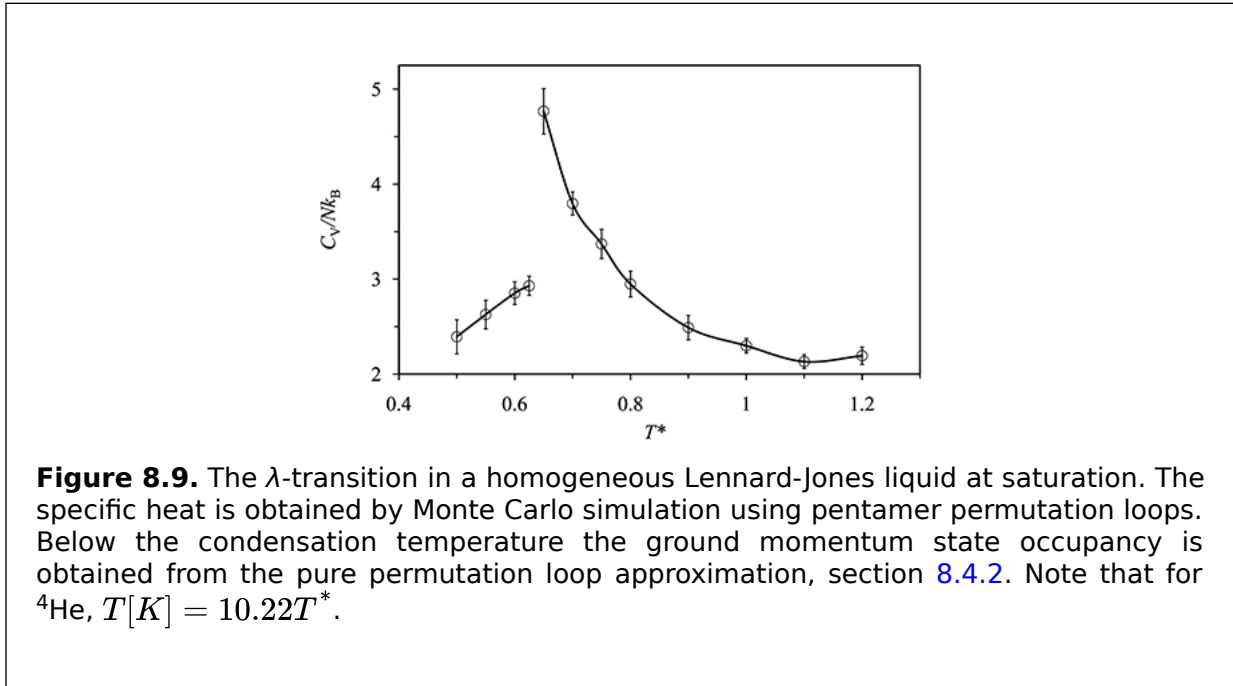
The heat capacity, which is the temperature derivative of the energy, can be similarly analyzed. The monomer term is

$$C_V^{(1)}(\bar{N}_0, \bar{N}_*) = \frac{3\bar{N}_* k_B}{2} + \frac{3k_B T}{2} \frac{\partial \bar{N}_*}{\partial T} + C_V^{cl,ex}(N).$$

(8.70)

The dimer and higher loops, $l \geqslant 2$, contribute

$$C_V^{(l)}(\bar{N}_0, \bar{N}_*) = l\left(\frac{\bar{N}_*}{N}\right)^{l-1} \frac{\partial \bar{N}_*}{N \partial T} \bar{E}^{(l)}(N) + \left(\frac{\bar{N}_*}{N}\right)^l C_V^{(l)}(N).$$

(8.71)

The results are plotted in figure 8.9. The data are from homogeneous canonical Monte Carlo simulations at the saturated density. The difference with figure 8.6 is that those results are $C_V(N)$, which assume that all N bosons are excited, whereas the results in figure 8.9 are $C_V(\bar{N}_0, \bar{N}_*)$, which take into account the number of ground momentum state bosons given by the pure loop theory below the condensation transition given by the mixed loop theory. The qualitative and quantitative resemblance to the measured λ-transition in liquid helium-4 is unmistakable.

Figure 8.9. The λ-transition in a homogeneous Lennard-Jones liquid at saturation. The specific heat is obtained by Monte Carlo simulation using pentamer permutation loops. Below the condensation temperature the ground momentum state occupancy is obtained from the pure permutation loop approximation, section 8.4.2. Note that for ^{4}He, $T[K] = 10.22T^*$.

The divergence and discontinuity in the heat capacity appear to result from using the simplest approximation, namely the binary division into either the ground momentum state or else the set of excited momentum states. There is evidence (figure 8.4, section 8.6.2, 8.6.6, 9.3, and 9.2) that condensation in fact represents the formation of pure momentum loops in the low-lying momentum states. By instead forcing all-or-nothing macroscopic occupation of the ground momentum state, the binary division creates a large loss in kinetic energy, a latent heat, and a divergence in the heat capacity at the λ-transition.

The measured heat capacity diverges at the λ-transition, going like $C_{\text{sat}} \sim A \mid 1 - T/T_\lambda \mid^{-0.013}$ (Lipa *et al* 1996). However, this is an integrable singularity, which means that the energy is continuous and there is no latent heat at the λ-transition. The experimentally measured fraction of condensed bosons (i.e. He II) at the λ-transition is reported to be zero (Donnelly and Barenghi 1998), which is difficult to reconcile with the known discontinuous transition to superfluidity.

8.5.5 Dressed bosons are ideal

The non-linear form for the weight of dressed ground momentum state bosons has the appearance of a fugacity, and so in the grand canonical system one can make the replacement $z^{N_0} \Rightarrow \left(z[1 + \langle\tilde{\eta}\rangle_{N_0, N_*, \text{cl}}]\right)^{N_0}$. One then has a form of ideal solution theory, with the dressed ground momentum state bosons acting as dilute solutes in a fluid of excited momentum state bosons. In this case the ground momentum state contribution to the grand potential, including the mixed loops with excited momentum state bosons, is that of an effective ideal gas,

$$-\beta\tilde{\Omega}(z, V, T) \approx -\ln\left\{1 - z[1 + \langle\tilde{\eta}\rangle_{\bar{N}_0, \bar{N}_*, \text{cl}}]\right\}.$$

(8.72)

Compare this with the first term in equation (8.7). This shows formally how interactions between bosons modify the ground momentum state contribution to the ideal gas grand potential. To this should be added the classical and the pure excited state contributions, $\Omega_{\text{cl}} + \Omega_*$.

If $-1 < \langle\tilde{\eta}\rangle_{\bar{N}_0, \bar{N}_*, \text{cl}} < 0$, which is likely the case on the high temperature side of the transition, then the effective fugacity is less than the actual fugacity, and the number of ground momentum state bosons would be less than predicted by neglecting mixed loops. For $\langle\tilde{\eta}\rangle_{\bar{N}_0, \bar{N}_*, \text{cl}} < -1$, the effective fugacity would be negative, which would be problematic. Each term in the series of chain weights in table 8.4 has to be multiplied by $(N_*/N)^{l-1}$, and so one cannot say *a priori* that the total is less than -1 without taking this into account. On the other hand, what one *can* see in table 8.4 is that for the fully excited system, $N_*/N = 1$, the sum $l \in [2, 4]$ is less than -1 for $T^* < 0.6$. This is thus the spinodal limit for full excitation, at least at this level of approximation. (The sum $l \in [2, 7]$ at $T^* = 0.6$ is $1.4 \pm .1$, which says that the spinodal limit has not yet been reached at this temperature at the $l_{\text{max}} = 7$ level of approximation.)

8.5.6 Dressed excited momentum state bosons

Above it has been discussed how mixed permutation loops can be resummed as dressed ground momentum state bosons. The permutations between these are non-local, and therefore they make a significant contribution to the grand potential. The non-locality was demonstrated by considering the concatenation of two chains, equation (8.61). It was shown that the exponents of the linking Fourier factors vanished, which is to say that it did not matter how far apart the two chains were.

In fact, a similar result holds for any two bosons in the same momentum state provided that they belong to a permutation loop rather than a chain. That is, for the *l*-loop, with boson in momentum state $\mathbf{p}_{j_l}$ at the head, suppose that the distance between head and tail is small, $q_{j_1, j_l} \lesssim \Lambda$. And similarly for the *m*-loop, with boson in momentum state $\mathbf{p}_{k_m} = \mathbf{p}_{j_l}$ at the head, $q_{k_1, k_m} \lesssim \Lambda$. The product of the two transposition factors that link the two loops is

$$e^{-\mathbf{p}_{j_l} \cdot \mathbf{q}_{j_l,k_1}/i\hbar} e^{-\mathbf{p}_{k_m} \cdot \mathbf{q}_{k_m,j_1}/i\hbar} = e^{-\mathbf{p}_{j_l} \cdot [\mathbf{q}_{j_l} - \mathbf{q}_{k_1} + \mathbf{q}_{k_m} - \mathbf{q}_{j_1}]/i\hbar}$$

$$= e^{-\mathbf{p}_{j_l} \cdot \mathbf{q}_{j_l,j_1}/i\hbar} e^{-\mathbf{p}_{k_m} \cdot \mathbf{q}_{k_m,k_1}/i\hbar}.$$

<div align="right">(8.73)</div>

This says that the symmetrization function for the two loops factorizes.

Hence one can regard each of the bosons in a given multiply-occupied momentum state as dressed by summing over the bosons in singly occupied momentum states that form a permutation loop with each. These dressed excited momentum state bosons may then be permuted non-locally, exactly as for ground momentum state bosons. The main difference between ground momentum state bosons and bosons in multiply-occupied momentum states is that the dress of the former consists of permutation chains, whereas that of the latter consists of permutation loops. The weight of loops is much less than that of chains; compare $g^{(l)}$ in table 8.1 with $\tilde{g}^{(l)}$ in table 8.4. There are of course many nonzero momentum states able to be multiply-occupied, whereas there is just one ground momentum state. These different distinct advantages make both possibilities relevant for the discussion of condensation in subsection 8.6.6.

8.6 Discussion

8.6.1 Ideal versus interacting bosons

For ideal bosons with the mass of ^{4}He, condensation into the ground state (energy is equivalent to momentum), which marks the λ-transition, begins at $T_\lambda = 3.13$ K, with $\rho\Lambda^3 = 2.61$, and $C_V^{\mathrm{id}}/Nk_B = 1.93$. These are close to the measured values for the λ-transition: $T_\lambda = 2.17$ K, with $\rho\Lambda^3 = 4.57$ and $C_{\mathrm{sat}}/Nk_B = 5$–12 (Donnelly and Barenghi 1998). (The values depend on how close to the transition it is measured since the heat capacity is actually divergent (Lipa *et al* 1996).) One can conclude from this agreement that the λ-transition is indeed a manifestation of Bose–Einstein condensation. One can further conclude that the phenomenon is dominated by ideal statistics, which must mean that it comes from long-range effects and that the interaction potential between the bosons is a perturbing rather than a determining factor.

The analytic results for ideal bosons based on the F London (1938) approximation, section 8.2, are in good agreement with those obtained by exact enumeration of momentum states (section 8.3) at least in the thermodynamic limit. This largely vindicates F London's (1938) binary division into ground and excited momentum states, as well the present mathematical formulation in terms of momentum loops. However, one should be aware that only a finite number of excited states are included in the exact enumeration calculations, which is significant for the discussion of the relative occupancy of the ground momentum state discussed in sections 8.2.4 and 8.6.6.

The ideal boson analysis predicts that at the λ-transition the heat capacity is continuous and finite, with a discontinuity in its temperature first derivative. This means that there is no latent heat for the transition, which is to say that the energy is continuous. Related to this is that for ideal bosons ground state occupancy is also continuous, since it begins at zero at the λ-transition, and that number is extensive below the transition (section 8.2.4). Experimental data indicate that the heat capacity is divergent at the λ-transition (Lipa *et al* 1996). The divergence is integrable, which means that the energy and the enthalpy are continuous and that there is no latent heat associated with the λ-transition. Also, the experimentally deduced fraction of He II begins at zero at the λ-transition (Donnelly and Barenghi 1998). On the other hand, measurement shows that superfluidity begins discontinuously at the λ-transition. If one accepts that superfluidity is due to condensed bosons, and that condensed bosons are ground state bosons, this would appear to indicate that the fraction of ground state bosons must be nonzero at the λ-transition, and that their number must be extensive at the λ-transition. One would expect that any such

discontinuity in ground state occupancy would give a discontinuity in the energy. As is discussed in section 8.6.6, it seems likely that these contradictions in the ideal bosons results and in the interpretation of the experimental data can be traced to the simplest binary division approximation and the assertion that condensation means occupation solely of the momentum ground state.

Including interactions between the bosons changes aspects of the ideal boson picture of Bose–Einstein condensation and the λ-transition both quantitatively and qualitatively. The computer simulations for Lennard-Jones ^{4}He use the F London (1938) binary division approximation (i.e. excited momentum state continuum plus the explicit ground momentum state). These can be discussed in the first instance in terms of pure (unmixed) loops, in which individual permutations are restricted to amongst bosons of the same type (i.e. either ground state or excited state). The simulations show a rapid growth in the loop series and in the heat capacity approaching $T^* = 0.5$ ($T = 5$ K for the Lennard-Jones parameters of ^{4}He) from higher temperatures. Terminating the series at pentamer loops and setting the number of ground momentum state bosons to zero gives a heat capacity $C_V^{\mathrm{LJ}}/Nk_{\mathrm{B}} = 17.6$ (with $\rho\Lambda^3 = 3.21$ here). The loop series appears to be divergent using $N_0 = 0$. Optimizing the number of ground momentum state bosons in the unmixed approximation gives the fraction of condensed bosons as $\bar{N}_0/N = 0.635$ at $T^* = 0.5$. If taken into account this would substantially lower the heat capacity and it would also likely make the loop series convergent. The unmixed approximation gives a continuous increase in the optimum fraction of condensed bosons with decreasing temperature from $\bar{N}_0/N = 0$ at $\rho\Lambda^3 = 1$.

Compared to the result for unmixed loops alone, including also mixed permutation loops, which consist of both ground and excited momentum state bosons, initially suppresses ground momentum state occupancy as the temperature is decreased from the point at which $\rho\Lambda^3 = 1$. Such loops comprise concatenated chains (dressed ground momentum state bosons), and they are specific to interacting bosons since they are forbidden for ideal bosons. Using up to the heptamer loops, it is found that occupancy of the ground momentum state occurs discontinuously at a point within the temperature range $0.65 > T_\lambda^* \geqslant 0.625$. At $T^* = 0.65$, $\rho\Lambda^3 = 2.02$, one has $\bar{N}_0/N = 0$ and $C_V^{\mathrm{LJ}}/Nk_{\mathrm{B}} = 4.77$. At $T^* = 0.625$, $\rho\Lambda^3 = 2.16$, the condensed fraction using the mixed approximation is $\bar{N}_0/N = 0.15$, whereas using the unmixed approximation it is $\bar{N}_0/N = 0.49$. Using the former, $C_V^{\mathrm{LJ}}/Nk_{\mathrm{B}} = 4.04$, whereas using the latter, $C_V^{\mathrm{LJ}}/Nk_{\mathrm{B}} = 2.93$. (The more accurately the λ-transition is pinned down, the higher is the one-sided heat capacity on either side.)

Of course there is a certain inconsistency in using the mixed theory to determine the location of the condensation transition and on the high temperature side of it, and then using the unmixed theory on the low temperature side of the transition to determine the fraction of condensed bosons and the consequent heat capacity. One can only argue in defence of this that the mixed theory invokes independent dressed ground momentum state bosons, and it is expected to fail when the fraction of condensed bosons becomes large. The qualitative picture of a barrier to condensation provided by the mixed theory can be expected to be robust. The quantitative values for the fraction of ground momentum state bosons following condensation ($\bar{N}_0/N = 0.49$ unmixed, versus $\bar{N}_0/N = 0.15$ mixed) are questionable, in part because of the handling of mixed loops does not forbid intersections, and in part because of limited computer resources (i.e. the termination of the loop series at the heptamer level). The latent heat at the λ-transition only occurs within the context of the binary division approximation and the mixed loop analysis. The unmixed theory, which is exact for ideal bosons, is expected to become more accurate for

interacting bosons as the temperature approaches absolute zero and the system is dominated by condensed bosons.

Approaching the λ-transition from the high temperature side, the heat capacity increases more rapidly for interacting than for ideal bosons. The increase in heat capacity is due to the increase in size and number of permutation loops. For both interacting and ideal bosons the dominant or only contribution comes from the excited state momentum continuum, which is embodied in the thermal wavelength Λ, which appears in the Gaussian that arises after momentum integration of the kinetic energy Maxwell–Boltzmann factor. The difference between ideal and interacting bosons is the increase in classical configurational (position) structure of the liquid with decreasing temperature for interacting bosons. When the thermal wavelength exceeds the location of the growing peak in the structure factor, a type of percolation transition takes place wherein individual permutation loops can become macroscopic.

The coupling of loop entropy and liquid structure for interacting bosons is consistent with the experimental result that the liquid saturation density for ^{4}He is a maximum at the λ-transition (Donnelly and Barenghi 1998, Sachdeva and Nuss 2010). Usually in a simple liquid the density increases monotonically with decreasing temperature because the thermal energy makes the nearest-neighbor separation on average greater than the minimum in the pair potential, which effect decreases with decreasing temperature. The relatively sharp increase in density in liquid helium approaching the λ-transition from the high temperature side is consistent with the increase in position permutation loops, since these create an attractive effective potential between bosons, equation (7.165). The more gradual decrease in density with decreasing temperature on the low temperature side of the transition suggests that this extra attraction is diminishing in part due to the decline in position permutation loops themselves. Also, the attractive force due to the Gaussian effective potential weakens once the thermal wavelength significantly exceeds the separation at which the pair potential is a minimum as further increase does little to alter the structure of the liquid. In any case, the measured density appears to be continuous, but its temperature derivative appears discontinuous at T_λ (Donnelly and Barenghi 1998, Sachdeva and Nuss 2010).

In the following section it is explained that permutations cause the formation of position loops on the high temperature side of the transition, and momentum loops on the low. Momentum loops depend upon the quantized momentum picture, and they are independent of the thermal wavelength and configurational structure of the liquid. There is a competition between the two types of loops that explains the maximum in the heat capacity at the λ-transition, since the growth in heat capacity is due to the growth in position loops, and this comes to an abrupt end when momentum loops take over. Within the context of the binary division approximation, the mixed loop calculations explain the coincidence of the λ-transition (the peak in the heat capacity) and the Bose–Einstein condensation transition, since ground momentum state occupancy is initially suppressed allowing position permutation loops to grow, until it occurs discontinuously, marking both transitions.

8.6.2 Position and momentum symmetrization loops

A symmetrization loop, in either the discrete quantum picture or in the continuum phase space prior to momentum integration, consists of loops of Fourier factors that can be written either as $\mathrm{e}^{-\mathbf{q}_{j_k,j_{k+1}}\cdot\mathbf{p}_{j_k}/\mathrm{i}\hbar}$, or as $\mathrm{e}^{-\mathbf{q}_{j_k}\cdot\mathbf{p}_{j_k,j_{k+1}}/\mathrm{i}\hbar}$. Although both forms yield the same value when evaluated over the loop, they serve to explain the different regions of phase space where the permutation loops contribute. The exponent must be small in magnitude in order for the loop to which it belongs to be significant upon averaging. Hence in practice there are two cases: either all $q_{j_k,j_{k+1}}$ are small (position loops) or else all $p_{j_k,j_{k+1}}$ are small

or zero (momentum loops). This division is convenient as the two types are qualitatively different.

Position loops are localized in position space. They are the ones that give rise to the loop phenomena that have been identified on the high temperature side of the λ-transition in the simulations, section 8.4. The effective pair potential derived in section 7.7.3 is a position dimer. Position loops are usually small in size l, and small in spatial extent. For high temperatures they are few in number. Approaching the transition from the high temperature side, $T \to T_\lambda^+$, their number, size, and extent increases.

Increasingly for $T < T_\lambda$, loops are localized in momentum space (momentum loops). They tend to be large in size l and large in spatial extent. (These assertions are justified below.) They are an example of the non-locality of quantum mechanics. The particles belonging to a momentum loop are highly correlated in momentum notwithstanding the fact that they are separated beyond the range of the interaction potential. With quantized momenta, all atoms in a loop can be in the same single-particle discrete momentum state, $\mathbf{p}_{j_k, j_{k+1}} = 0$. Quantum leaps in momenta around the loop limit the loop's size and spatial extent in a way that having all the bosons in the same momentum state does not. Indeed to be definite, one can define a momentum loop to be one in which all bosons are in the same momentum state, and position loops to be everything else. (By this definition the permutation loop simulations of section 8.4 invoke position loops, and the exact ideal boson analysis of section 8.3 involves only momentum loops.)

When there are many more accessible momentum states than bosons, most bosons are monomers. When the number of accessible momentum states is comparable to or less than the number of bosons, a significant number, or even the majority, of bosons share a momentum eigenvalue with others.

In the quantized picture, the number of accessible single-particle momentum states is $n_p \approx \left(\sqrt{2\pi m k_{\mathrm{B}} T}/\Delta_p\right)^3 = \Lambda^{-3} V$. This formula for n_p comes from the momentum eigenvalues and the exponent of the Maxwell–Boltzmann factor for a single boson. The factor of 2π is an order of magnitude constant that simplifies the following estimate. Since $\Delta_p = 2\pi\hbar/L$, the number of accessible single-particle momentum states is macroscopic even at low temperatures, $n_p \propto T^{3/2} V$. However, the number of bosons is also macroscopic, and the probability that none of the N bosons is in the same momentum state as any other is vanishingly small when $n_p = \mathcal{O}(N)$ (birthday paradox). This occurs at low temperatures when $\rho\Lambda^3 = \mathcal{O}(1)$. More exactly, with the above formula, $n_p = N \Leftrightarrow \rho\Lambda^3 = 1$. In the simulations of the Lennard-Jones liquid, the λ-transition occurs at about $\rho\Lambda^3 = 2$–3. The measured value for helium-4 is $\rho\Lambda^3 = 4.57$, and that for the ideal gas is $\rho\Lambda^3 = 2.61$. These suggest that the present criterion, $\rho\Lambda^3 \gtrsim 1$ (equivalently, multiple occupancy of momentum states), is necessary but not sufficient for the λ-transition.

In summary, on the high temperature side of the λ-transition, the system is dominated by position loops, whose number, size l, and spatial extent all increase with decreasing temperature. But increasing the number of excited state position loops encourages the dressing of ground momentum state bosons, and hence of occupancy of the ground momentum state. Condensation into the momentum ground state interferes with the formation of position loops because there are fewer excited momentum state bosons available, and because a ground momentum state boson blocks the formation of excited momentum state boson position loops in the vicinity. More generally, bosons belonging to pure momentum loops interfere with the formation of position loops in their neighborhood. After the λ-transition the number, size l, and spatial extent of position loops all decrease with decreasing temperature.

8.6.3 Effect of commutation function on position loops

In the analysis in this chapter the commutation function was neglected. It will be recalled from section 7.4.2 that this is a phase space function that accounts for the non-commutativity of the position and momentum operators. One can qualitatively predict the effect of the commutation function exponent by focussing upon the two-body contribution, $w^{(2)}(\mathbf{q}_1, \mathbf{q}_2; \mathbf{p}_1, \mathbf{p}_2)$.

For the present case of a spherically symmetric pair potential, this reduces to a function of the separation and the axial component of the relative momentum,

$$\mathrm{e}^{-\beta p^2/2\mu} \mathrm{e}^{-\beta u(q)} \mathrm{e}^{w^{(2)}(p_\parallel, q)} = \mathrm{e}^{p_\parallel q/i\hbar} \mathrm{e}^{-\beta[(-\hbar^2/2\mu)\nabla_q^2 + u(q)]} \mathrm{e}^{-p_\parallel q/i\hbar}.$$

(8.74)

where $q = q_{12}$, $p_\parallel = \mathbf{p} \cdot \mathbf{q}/q = \mathbf{p}_{12} \cdot \mathbf{q}_{12}/2q_{12}$, and $\mu = m/2$.

The commutation function exponent is largest at the minimum of the pair potential. About this minimum one approximation is to use the commutation function for a simple harmonic oscillator, for which an exact analytic expression exists (Attard 2021 section 8.6). Figure 8.1 of Attard (2021) shows that at the minimum of the kinetic and potential energy, which is the maximum of the Maxwell–Boltzmann classical phase space weight, the commutation function exponent is negative, and it increases in magnitude with decreasing temperature. Figure 8.2 of Attard (2021) shows that the magnitude of the commutation function exponent decreases with increasing separation and relative momentum.

From this one concludes that at the two-body level for a system with a pair potential minimum (like the present Lennard-Jones potential), the commutation function exponent acts like a repulsive pair potential that will decrease the depth and curvature of the pair potential of mean force compared to the case where the commutation function is neglected. This will lead to a broader distribution of separations and relative momenta than would be the case classically. This is consistent with one's expectations based on the Heisenberg uncertainty principle.

Since the minimum in the pair potential of mean force essentially determines the structure of the fluid, it plays a key role in position permutation loops; the growth in size and number of the latter with decreasing temperature over and above that of the ideal gas is a direct result of the increase in fluid structure. Hence including the commutation function will reduce the contribution from position permutation loops to the various thermodynamic properties. In particular, the divergence in the loop series and in the heat capacity found here for Lennard-Jones ^{4}He at $T^* = 0.65$ ($T = 6.6$ K) would be pushed to lower temperatures if the commutation function were included.

In addition to, or as a result of, the above effects, the commutation function substantially raises the pressure above the classical prediction for Lennard-Jones ^{4}He (figure 7.5, $T^* = 0.6$). This means that the saturation liquid density would be lower in an accurate quantum calculation than the classical coexistence value used in the present calculations because the saturated vapor has low density and low pressure, which are quite well accounted for classically. The decrease in density would reduce the contribution of the position permutation loops at a given temperature compared to the present calculations.

8.6.4 Molecular contributions to the entropy of loops

Position loops do not strongly restrict momenta, although they definitely favor low-lying momentum states. Momentum loops do not strongly restrict positions. Strictly, momentum loops comprise bosons with the same momenta, which can only occur in the quantized case, and they can extend spatially throughout the whole system.

These differences in the two types of loops induce a difference in the configurational or molecular contribution to their entropy. (The combinatorial entropy of loops is treated below.) Momentum loops require highly correlated momenta, and they therefore have low momentum entropy. This is quite costly at high temperatures where there are many

momentum states available. It is much less costly at low temperatures where there are few accessible momentum states.

In contrast the spatial structure of the liquid is relatively insensitive to temperature, and so the configuration cost of the short-range spatial order of position loops changes little with temperature. This is particularly true at temperatures below the λ-transition.

What favors momentum loops is that the atoms that comprise them can range throughout the volume of the system, because the differences in consecutive momenta can (or must) be precisely zero. Momentum loops are unconstrained by space. As mentioned, they are an example of quantum non-locality. In three-dimensional space, there are many more possible loops comprising far-separated consecutive atoms than close-separated ones because the number of possible partners increases quadratically with separation.

Position loops never have the same freedom in momentum space because the atoms never attain zero separation. And also, the accessible volume of momentum space decreases with decreasing temperature. Instead there is a threshold where the thermal wave-length extends beyond the nearest-neighbor separation in which the atoms in position loops can be composed from the available momentum states.

8.6.5 Combinatorial entropy of loops

The loop grand potential in the form $-\beta\Omega^{\pm,(l)}$ measures the entropy due to loops. The reader may have trouble accepting this, particularly in view of the fact that position loops require atoms localized to small spatial neighborhoods, and momentum loops require the atoms to have highly correlated momenta. To understand the loop entropy, one has to recall that the loop expansion begins with the monomer or classical term. This accounts for in its entirety the classical picture of entropy in terms of molecular configurations in phase space. The loop entropy is in addition to the classical entropy, and it is irrelevant that the loops themselves consist of ordered sets of molecules.

Loops are distinct arrangements of molecules, and such arrangements can be counted. Hence there is an entropy associated with them (chapter 1). The partition function, for example equation (7.71) or (8.1), is the sum over states and also over wave function symmetrization permutations. The logarithm of the partition function is the total entropy. In the case of bosons, all distinct permutation states add to the entropy. In the case of fermions, some permutations preclude otherwise acceptable states, which subtracts from the total entropy. The effective pair potential, equation (7.165), $v^{\pm}(r) = -k_B T \ln[1 \pm e^{-2\pi r^2/\Lambda^2}]$, is attractive for bosons and repulsive for fermions. This comes from the dimer loop, and, as signified by its temperature-dependence, is entropic in nature.

For bosons in a given momentum state, instead of subdividing the permutations into momentum loops, it is often more convenient to focus on the occupancy of the momentum state as a whole. In the example of $N_{\mathbf{a}}$ bosons occupying the single-particle momentum state $\mathbf{a}$, there are $\chi_{\mathbf{a}}^{+}(N_{\mathbf{a}}) = N_{\mathbf{a}}!$ possible arrangements in total. The logarithm of this gives the internal entropy of the momentum state with this occupancy, $S_{\mathbf{a}}^{+,\mathrm{int}} = k_B \ln N_{\mathbf{a}}!$. Note that these permutations of bosons in the same momentum state do not depend upon their positions, which is to say that they are non-local. (This occupancy picture neglects the contribution of mixed loops of interacting bosons, which are important above and in the vicinity of the λ-transition.)

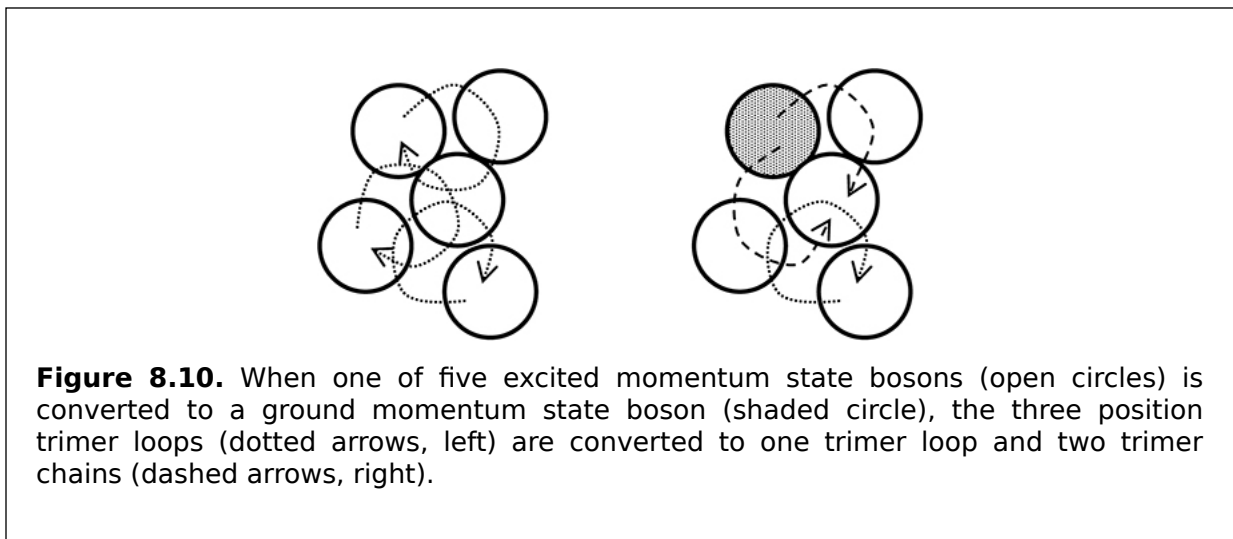## 8.6.6 Condensation is into low-lying momentum states

Permutation entropy favors bosons in the same momentum state, as these can form non-local permutation loops. From the point of view of permutation entropy, any single momentum state is as good as any other for forming non-local pure permutation loops. This raises the question: is it true that at the λ-transition bosons condense overwhelmingly

into the ground momentum state rather than into any other low-lying momentum state, or perhaps several of them?

The answer begins with the mixed loops analyzed in section 8.5. In subsection 8.5.3 it was shown that dressed bosons, which are localized permutation chains with a ground momentum state head and excited momentum state tails, could be concatenated to form non-local mixed permutation loops. This property is not unique to the ground momentum state, since section 8.5.6 shows that multiply-occupied momentum states can form dressed loops. But chains have a greater weight than loops; compare $g^{(l)}$ in table 8.1 with $\tilde{g}^{(l)}$ in table 8.4. In this regard there is a preference for the ground momentum state. Since so much entropy is available from such non-local permutations, and since there are significant numbers of such dressed ground momentum state bosons present in the vicinity of the λ-transition (because localized excited momentum state permutation loops are present in large numbers), this drives condensation into the ground momentum state at the transition.

Although condensation into the ground momentum state occurs at the λ-transition, it appears that the process does not end there. The condensation contributes to the destruction of position permutation loops at the transition because the ground momentum state bosons interfere with the ability to form loops composed only of excited momentum state bosons. Also, the open nature of dressed boson chains, and the non-local loops formed from chains, diminish the fluid structure that is formed by, and that is necessary for, position permutation loops. In short, ground momentum state bosons poison the very position permutation loops that catalyze them.

Figure 8.10 sketches this mechanism. The left-hand cluster illustrates that position trimers can only form when the excited state bosons are sufficiently close to each other. When one of these is converted to a ground momentum state boson, as in the right-hand cluster, two of the trimer loops become chains, which loosens the bond between the ground momentum state boson and the central boson because there is no longer the requirement that it be close in order for the permutation to average to a nonzero value. This means in turn that the cluster is less strongly bound together, and that it will break up more easily than if there had been no conversion to the ground momentum state.

Figure 8.10. When one of five excited momentum state bosons (open circles) is converted to a ground momentum state boson (shaded circle), the three position trimer loops (dotted arrows, left) are converted to one trimer loop and two trimer chains (dashed arrows, right).

In the context of what follows, it is important to note that this mechanism for the destabilization of position permutation loops also results from any boson that is in a multiply-occupied momentum state, which is to say that it participates in pure momentum permutation loops. Loops formed from dressed bosons in such multiply-occupied

momentum states, as in section 8.5.6, also upset position permutation loops. Non-local permutations confer a momentum stability to the dressed boson that is anathema to position permutation loops, which are necessarily composed of bosons that can be summed over multiple momentum states. Although chains have a greater weight than loops (tables 8.1 and 8.4), there are an extensive number of nonzero momentum states, each of which can head a dressed boson loop. Thus in the competition with position loops, both ground and excited state momentum loops may contribute comparably.

Since position permutation loops both require and encourage bosons to be in different momentum states (the thermal wavelength results from independent integration over each boson's momentum), the reduction in position permutation loops must be accompanied by an increase in the number of bosons in the same momentum state. So while the λ-transition signifies the growth of dressed ground momentum state bosons and the growth of the occupancy of the momentum ground state, the accompanying loss of position permutation loops also signifies the growth in pure momentum loops and condensation into the low-lying momentum states. (It can equally be argued that the growth of some pure excited state momentum loops from the very large number of possible excited momentum states cause the decline in the position permutation loops.) This makes sense, because in practical terms the energy gap between the momentum ground state and the low-lying momentum excited states is orders of magnitude smaller than the thermal energy, and it is not possible to argue that there should be any sharp distinction between the ground momentum state and the low-lying momentum states. As discussed in the first paragraph of this section, any pure momentum loop is non-local, and the permutation entropy contributed by a momentum state depends only upon its occupancy, not upon the value of the momentum.

One has to conclude from this that the λ-transition, and Bose–Einstein condensation more generally, does not represent a marked increase in the occupancy solely of the momentum ground state. Rather, the condensation represents a marked increase in the occupancies of occupied low-lying momentum states. Equivalently, it marks the shift in dominance of position to momentum permutation loops.

This idea is sketched in figure 8.11, which shows that at a given temperature the range of accessible momentum states is about the same for non-condensed bosons, otherwise known as He I, as it is for condensed bosons, otherwise known as He II. Further, the envelope of the occupancy distribution has more or less the same shape in the two cases, which explains how there can be no latent heat at the λ-transition despite macroscopic condensation occurring discontinuously. What is different is that for condensed bosons the instantaneous occupancy of occupied states is much greater than for non-condensed bosons, and the instantaneous number of empty momentum states within the range of accessible states is much greater below the condensation transition than above.
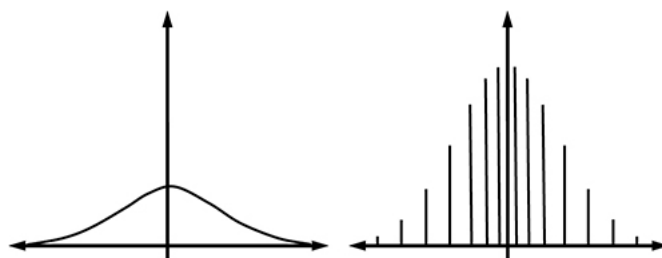

Figure 8.11. Occupancy of momentum states by non-condensed bosons (He I, left), and by condensed bosons (He II, instantaneous, right).

In the conventional binary division approximation, He I comprises the excited momentum (or energy) state bosons and He II comprises the ground momentum (or energy) state bosons. In the picture advocated here, He I comprises bosons in few-occupied momentum states, and He II comprises bosons in highly-occupied momentum states.

The 'spiked' occupancy diagram in figure 8.11 is consistent with the fluctuations in single-particle energy state occupancy for bosons, equation (7.156). (For ideal bosons, momentum states are the same as energy states, and there is reason to believe that ideal behavior dominates below the condensation transition (section 9.3).). Following equation (7.156) it was pointed out that the fluctuations in the occupancy of states with high average occupancy were as large as the occupancy itself, which means that there is a non-negligible probability of finding them empty or occupied by only a few bosons. There is also a non-negligible probability of some states having occupancy much greater than their average. In this sense the spikes represent an instantaneous snapshot of the condensed system rather than the average. The lifetime of the individual fluctuations is not specified *a priori*, but in the light of the discussion of the mechanism for superfluidity discussed in section 9.4, it is probably relatively long.

With He II being the condensed bosons in highly-occupied momentum states, and He I being non-condensed bosons in singly (or few) occupied momentum states, one can see that it is possible for the two to coexist, and that ^{4}He below the λ-transition can be considered a mixture of the two. Superfluidity is carried by the bosons in the highly-occupied momentum states, as is discussed in section 9.2.

It will be recalled from section 8.5.4 that the mixed loop computations for Lennard-Jones ^{4}He produced a latent heat for the λ-transition, which gives a divergence in the heat capacity. The divergence in the experimentally measured heat capacity is integrable (Lipa *et al* 1996), which means that the energy and the enthalpy are continuous and that there is no latent heat at the λ-transition. The latent heat computed in section 8.5.4 is dependent on using mixed and unmixed loops on either side of the transition, and on the binary division approximation. The assertion that condensation is into the ground state forces a discontinuity in the kinetic energy at the transition. The present picture of condensation into multiple highly occupied momentum states allows the occupancy to be discontinuous while the kinetic energy can remain continuous at the transition. In other words, unlike the binary division approximation, there is likely no latent heat in the present more sophisticated model of Bose–Einstein condensation.

This picture suggests that the binary division of bosons as belonging to either the ground momentum state or else the set of excited momentum states (F London 1938), which was used for the ideal and interacting boson results in this chapter, is only a first approximation. It produces reasonable results on the high-temperature side of the λ-transition, and also for the location of the transition. A more sophisticated model would account for the fact that below the λ-transition pure momentum loops form, and some or all of the low-lying momentum states are each occupied by large numbers of bosons (see sections 9.3 and 9.4).

It is possible that the molecular mechanism for the λ-transition discussed above—excited momentum state position permutation loops nucleating ground momentum state condensation, and ground momentum state bosons breaking up position permutation loops—may be exaggerated by the binary division approximation. In the more nuanced picture, multiple multiply-occupied low-lying momentum states occur at low enough temperatures, and the entropy from the non-local nature of their permutation loops may be enough to overcome the entropy of the localised position permutation loops without needing the specific intervention of ground momentum state bosons.

To the extent that the chain weight is greater than the loop weight, $\tilde{g}^{(l)} > g^{(l)}$ (tables 8.1 and 8.4), and that non-condensed bosons dress the condensed bosons below the λ-transition, the ground momentum state will enjoy greater weight than any individual

excited momentum state. This means that below the λ-transition the average occupancy of the ground momentum state will be greater than that of any individual excited momentum state, even though the total number of condensed bosons in excited momentum states exceeds those in the ground momentum state.

Following this last point, it is worth mentioning that a significant difference between Bose–Einstein condensation in liquid helium and that in Bose–Einstein condensates is the size of the system. The former has on the order of Avogadro's number of atoms, whereas the latter typically has $\mathcal{O}(10^2\text{–}10^5)$ atoms (Wikipedia 2022), which is fewer than in even one occupied low-lying momentum state of ^{4}He. Because of the size difference, the conclusions drawn here for liquid helium below the λ-transition do not automatically carry over to Bose–Einstein condensates. Nevertheless it does appear that even in such condensates the condensation is into a range of low-lying momentum states rather than solely into the ground momentum state (le Bellac *et al* 2004 figure 5.17, Wikipedia 2022 velocity-distribution data).

The common conception and experimental measurement of a specific fraction of He II below the λ-transition may be interpreted in the light of this more sophisticated model of condensation as occupancy of multiple low-lying momentum states. In this new picture He I consists of bosons instantaneously in few-occupied momentum states, and He II consists of bosons instantaneously in highly-occupied momentum states. Whether there is a sharp boundary between the two is unclear. This more nuanced model of condensation is discussed in greater detail in the context of superfluidity in chapter 9.

Summary

- Momentum states are better than energy states for representing interacting particles. In the first approximation (binary division), condensation is into the ground momentum state rather than into the ground energy state. Ground momentum state permutations are non-local.
- Ideal bosons semi-quantitatively capture the λ-transition in ^{4}He, which confirms that it is a manifestation of Bose–Einstein condensation, and that it is dominated by long-range particle statistics that are independent of the boson interactions to leading order.
- Interactions between bosons significantly contribute to the magnitude and growth rate of the heat capacity approaching the λ-transition from the high temperature side in liquid ^{4}He. They couple position permutation loops to liquid structure. They also give rise to ground and excited momentum state mixed permutation loops, which are prohibited for ideal bosons, and which initially create an entropy barrier to condensation. The change in dominance from position to momentum loops defines the λ-transition, after which the position loops begin to decline.
- Experimental measurements at the λ-transition show that the heat capacity is integrably divergent, which implies that the energy is continuous, that the density is continuous, and that the fraction of He II grows continuously from zero. Experiments also show that superfluidity is discontinuous and nonzero at the λ-transition. The present theory for interacting bosons using mixed loops and defining condensation as a binary division shows discontinuous ground momentum state occupancy at the λ-transition, which implies energy, density, and superfluid discontinuities.
- The binary division into the ground momentum state and the excited momentum state continuum is a good first approximation for locating the condensation transition, and also for describing the high temperature side of it. However, inconsistencies on the low temperature side predicted for both ideal and for interacting bosons suggest that the condensed state actually consists instantaneously of multiple highly-occupied momentum states.

References

Abramowitz M and Stegun I A 1970 *Handbook of Mathematical Functions* 9th printing edn (New York: Dover)

Attard P 2015 *Quantum Statistical Mechanics: Equilibrium and Non-Equilibrium Theory from First Principles* (Bristol: IOP Publishing)

Attard P 2018 Quantum statistical mechanics in classical phase space. expressions for the multi-particle density, the average energy, and the virial pressure arXiv:1811.00730

Attard P 2021 *Quantum Statistical Mechanics in Classical Phase Space* (Bristol: IOP Publishing)

Attard P 2022 Bose-Einstein condensation, the lambda transition, and superfluidity for interacting bosons arXiv:220107382

Balibar S 2014 Superfluidity: how quantum mechanics became visible *History of Artificial Cold, Scientific, Technological and Cultural Issues* ed Gavroglu K (ed) (Dordrecht: Springer) pp 93–117

Balibar S 2017 Laszlo Tisza and the two-fluid model of superfluidity *C. R. Phys.* **18** 586

Donnelly R J 1995 The discovery of superfluidity *Phys. Today* **48** 30

Donnelly R J and Barenghi C F 1998 The observed properties of liquid Helium at the saturated vapor pressure *J. Phys. Chem. Ref. Data* **27** 1217

Donnelly R J 2009 The two-fluid theory and second sound in liquid helium *Phys. Today* **62** 34

Feynman R P 1954 Atomic theory of the two-fluid model of liquid helium *Phys. Rev.* **94** 262

Griffin A 1999 A brief history of our understanding of BEC: from Bose to Beliaev arXiv:cond-mat/9901123v1

Kawatra M P and Pathria R K 1966 Quantized vortices in an imperfect Bose gas and the breakdown of superfluidity in liquid helium II *Phys. Rev.* **151** 132

Keesom W H and Wolfke M 1927 Two different liquid states of helium *Commun. Phys. Lab. Leiden* **22**

Keesom W H, Clusius K and Keesom A P 1932 Ueber die specifische warme des fussigen heliums *Commun. Phys. Lab. Leiden* **219e** and 221d

Landsberg P T 1954 On Bose–Einstein condensation *Proc. Camb. Phil. Soc.* **50** 65

Landau L D 1941 Theory of the superfluidity of helium II *Phys. Rev.* **60** 356

le Bellac M, Mortessagne F and Batrouni G G 2004 *Equilibrium and Non-Equilibrium Statistical Thermodynamics* (Cambridge: Cambridge University Press)

Lipa J A, Swanson D R, Nissen J A, Chui T C P and Israelsson U E 1996 *Phys. Rev. Lett.* **76** 944

London F 1938 λ-phenomenon of liquid helium and the Bose–Einstein degeneracy *Nature* **141** 643

London F 1950 *Superfluids* **vol I** (New York: Wiley)

Merzbacher E 1970 *Quantum Mechanics* 2nd edn (New York: Wiley)

Messiah A 1961 *Quantum Mechanics* **vols 1** (Amsterdam: North-Holland)

Pathria R K 1972 *Statistical Mechanics* (Oxford: Pergamon)

Sachdeva D and Nuss M 2010 Density of liquid ^{4}He as a function of temperature (University of Minnesota, Methods of Experimental Physics) http://spa-mxpweb.spa.umn.edu/s10/Projects/S10_DensityofLiquidHe/

van Sciver S W 2012 *Helium Cryogenics* 2nd edn (New York: Springer)

Tisza L 1938 Transport phenomena in helium II *Nature* **141** 913

Wikipedia 2022 Bose–Einstein_condensate https://en.wikipedia.org/wiki/Bose--Einstein_condensate (accessed 4 August 2022)

Entropy Beyond the Second Law (Second Edition)

Thermodynamics and statistical mechanics for equilibrium, non-equilibrium, classical, and quantum systems

Phil Attard

Chapter 9

Superfluidity: flow without entropy dissipation

'helium below the λ-point enters a special state which might be called superfluid'

P Kapitza (1938)

'A verification of equ. (11) should be the easiest method of proving the hypothesis that the helium atoms moving through narrow capillaries are in their lowest quantized state.'

H London (1939)

[F. London and] 'L. Tisza suggested that helium II should be considered as a degenerate ideal Bose gas.... This point of view, however, cannot be considered as satisfactory... nothing would prevent atoms in a normal state from colliding with excited atoms, i.e. when moving through the liquid they would experience a friction and there would be no superfluidity at all. In this way the explanation advanced by Tisza not only has no foundation in his suggestions but is in direct contradiction to them.'

L Landau (1941)

'The division into a normal fluid and a superfluid although yielding a simple model for understanding the final equations appears artificial from a microscopic point of view.'

R P Feynman (1954)

9.1 Introduction

Superfluid He II is commonly believed to be bosons condensed in the energy ground state. It is further asserted that this necessarily forbids momentum exchange, which explains why they flow without viscous dissipation. What happens in a collision with a non-condensed boson or a wall is not explained. In this chapter evidence is presented that following Bose–Einstein condensation, the system instantaneously consists of multiple highly occupied momentum states. Equations of motion are given that explain that He II does not dissipate momentum in collisions because of the non-local nature and the large permutation entropy of each instantaneously highly occupied momentum state.

Superfluidity, which is flow without viscosity, was discovered in helium-4 by Kapitza (1938) and by Allen and Misener (1938). The transition to superfluidity coincides with the λ-transition. Therefore it should be clear that superfluidity is connected with Bose–Einstein condensation, which is the accepted basis for the λ-transition (see chapter 8). Readable histories of superfluidity have been given by Donnelly (1995, 2009) and by Balibar (2014, 2017).

Superfluid helium was originally called He II, being the fluid whose properties become noticeable on the low temperature side of the liquid-liquid λ-transition (Keesom *et al* 1927, 1932). The ideal boson two-fluid theory of F London (1938) and Tsiza (1938) asserts that below the λ-transition the system consists of coexisting superfluid (He II) and normal (He I) fluid. The superfluid consists of bosons in the ground energy state, which are said to have zero viscosity, and the normal fluid comprises bosons in excited states having the usual viscosity (Tsiza 1938). The superfluid is said to be able to flow unimpeded through the normal fluid, although the precise mechanism for this is unstated. The equilibrium thermodynamics of the two-component mixture is given by the ideal boson approach of F London (1938), as detailed in section 8.2.

The phonon-roton theory of Landau (1941) is also a type of two-fluid scheme, with the superfluid again being bosons in the ground energy state, forming a quiescent background, and again said to have zero viscosity. But in this case the normal fluid consists of a gas of elementary energy excitations, which are related to quantized sound waves, which are termed phonons and rotons, and which are taken to be bosons. The theory is said to be applicable far below the λ-transition close to absolute zero (Pathria 1972), although of course superfluidity sets in at the λ-transition itself. The postulated rotons, whose energy spectrum Landau (1941) fitted to measured heat capacity data, have been interpreted as quantized vortices (Feynman 1954, Kawatra and Pathria 1966, Pathria 1972), although direct evidence for this is lacking. Landau (1941) never accepted Bose–Einstein condensation as playing a role in the λ-transition (Balibar 2014, 2017) (see the quote at the head of the chapter). It is difficult to imagine that his phonon-roton theory can have any relevance to superfluidity if the latter is accepted to be a consequence of Bose–Einstein condensation and the λ-transition. Bogolubov (1947) gave for superfluidity a quantum mechanical wave function theory that included weak interactions but which was explicitly based on Bose–Einstein condensation. A summary of these and other conventional theoretical treatments of Bose–Einstein condensation and of superfluidity has been given by Griffin (1999).

Measurements of superfluidity generally involve non-equilibrium phenomena, and so a complete analysis would invoke the formulations of chapters 3, 4, and 6. The analysis of superfluidity in the present chapter is incomplete. Section 9.2 makes the case for a new interpretation of the Bose–Einstein condensed state. Section 9.3 presents an analysis of interacting bosons far below λ-transition where pure momentum permutation loops dominate. Section 9.4, based on the evidence for the state of ^{4}He below the λ-transition, discusses in detail the molecular mechanism for inviscid superfluid flow, and compares the predicted and measured critical velocities. Section 9.4.4 gives superfluid equations of motion, and offers a physical explanation of how condensed bosons collide without dissipating momentum. Section 9.5 treats the fountain effect in liquid ^{4}He, which arises in two chambers held at different temperatures and connected by a thin capillary. Although this appears to be merely structural, it is in actual fact a steady-state phenomenon that reveals the principle that drives superfluid flow.

9.2 Nature of the Bose–Einstein condensed state

As mentioned above, it is universally accepted that Bose–Einstein condensation is into the energy ground state. My view however is that after condensation a

snapshot of the system would reveal multiple highly occupied low-lying momentum states (as well as a majority of states that are empty, and also states that are few-occupied, the number of which depends on the proximity to the transition). Perhaps Feynman's observation, as quoted at the head of this chapter, was remarkably prescient. Because my picture of Bose–Einstein condensation is contrary to the widespread and longstanding understanding, it is incumbent upon me to lay out my evidence in detail.

The argument was largely made in section 8.6.6, based on the computer simulations of the λ-transition in ^{4}He. These revealed the dominance of position permutation loops leading up to the transition, the nucleation of ground momentum state occupancy at the transition, and the growth of momentum permutation loops at the expense of position permutation loops after the transition. Roughly speaking, in the conventional language of the binary division approximation, position loops are associated with excited state bosons (non-condensed He I), and momentum loops are associated with condensed bosons (He II) (section 8.6.2). Here I reprise some of those arguments and present new ones for condensation into low-lying momentum states, as opposed to condensation solely into the ground momentum state (equivalently, the ground energy state).

First to extensivity, which is important because superfluid flow must scale with the size of the system. If one doubles the size of the subsystem $V = L^3$ at constant intensive variables, then one doubles the number of particles. Since the volume of each momentum state, $\Delta_p^3 = (2\pi\hbar/L)^3$, is halved, there are now twice as many accessible states, which means that the occupancy of each momentum state must remain unchanged. This includes the ground momentum state and individual excited momentum states. Occupancy of any one individual state is thus an intensive variable (cf section 8.2.4).

This argument that the occupancy of any one momentum state must be an intensive variable implies that the fractional occupancy of a state must go to zero in the thermodynamic limit. This can be seen in figure 9.1, which results from the exact enumeration over momentum states for ideal bosons (section 8.3). The figure shows that the fraction of ideal bosons in the momentum ground state decreases with increasing system size at fixed temperature and density. The same trend holds for each excited momentum state. Obviously, however, the number of momentum states within a given energy band increases with increasing system size. Whilst the number of bosons in the ground momentum state is an intensive variable, the total number of condensed bosons (i.e. those in low-lying multiply-occupied momentum states) is extensive.
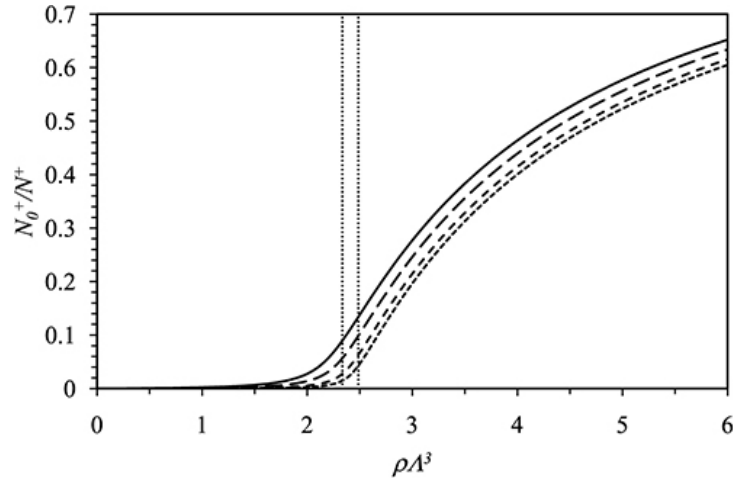
Figure 9.1. Average fraction of ideal bosons in the ground momentum state for total $N^+ = 1000$ (solid curve), $N^+ = 2000$ (long dashed curve), $N^+ = 5000$ (dashed curve), and for $N^+ = 10\,000$ (short dashed curve). The abscissa is essentially $T^{-3/2}$. The dotted lines locate the heat capacity maximum for the smallest (left) and largest (right) system. Exact enumeration (section 8.3) is used, and the density is $\rho = 0.3$.

Average occupancy is also practically continuous, because the spacing between momentum states scales inversely with subsystem edge length. There seems nothing special about the ground momentum state, at least not on the far side of the λ-transition. The pure momentum loops do not single out the ground momentum state, since the permutation entropy contributed by pure momentum loops in any momentum state is equally non-local. Condensation and superfluidity appear to be carried by the pure momentum loops of the low-lying momentum states, because the total occupancy of these is an extensive variable. Since the occupancy of the ground momentum state is intensive, it alone cannot carry superfluidity.

It was F London (1938) who originally devised the binary division into the ground and continuum excited momentum states for his ideal boson treatment. As pointed out in section 8.2.4, this predicts that the occupancy of the ground momentum state is a function of the fugacity, $\bar{N}_0^{+,\mathrm{id}} = z/[1-z]$, which is an intensive variable. It also predicts that the ground momentum state occupancy is extensive, $\bar{N}_0^{+,\mathrm{id}} \sim N[\rho\Lambda^3 - \zeta(3/2)]/\rho\Lambda^3$. This contradiction arises because the binary division approximation forces the ground momentum state to perform two roles: one is as literally the ground momentum state, in which the occupancy is intensive. And the other is as the condensed state, which in the present language is the set of low-lying, multiply-occupied momentum states, whose total occupancy is extensive. The divergence of the F London (1938) expression for the ground momentum state occupancy as $z \to 1^-$, seems to be an artefact of the continuum approximation for the excited momentum states, which severely limits the number of bosons that can

be accommodated. (The divergence would be physical if it applied to the condensed states as an extensive whole, rather than to the ground state as an intensive single state.) The exact enumeration of discrete momentum states for ideal bosons confirms that the fraction of bosons in the momentum ground state goes to zero in the thermodynamic limit, figure 9.1. If one does not make the continuum approximation, or indeed the binary division, then there would be no limit on the number of bosons that could occupy the low-lying momentum states. Consequently there would be no excess bosons that would be forced to condense into the momentum ground state.

In the present picture of condensation into the discrete low-lying momentum states, it might be argued that due to the growth of the permutation entropy with the occupancy of a state, there could be a non-linear effect in which all the bosons with low-lying momenta condense into a single state. Such a non-linear effect alone would not single out the momentum ground state, since it applies equally to any single low-lying momentum state. Since the spacing between kinetic energy states is on the order of $10^{-14}k_{\mathrm{B}}T$ for a macroscopic subsystem, it's difficult to argue that the thermal preference for the momentum ground state is large enough to drive the putative non-linear condensation into it. Even symmetry arguments, which say that the average momentum must be zero, do not persuade that there is a marked preference for condensation into the momentum ground state.

It can also be pointed out that the experimentally measured enthalpy is continuous at the λ-transition (Donnelly and Barenghi 1998), which is consistent with the fact that the divergence in the heat capacity is integrable (Lipa *et al* 1996). This suggests that there remains substantial occupation of the excited low-lying momentum states immediately after the transition (i.e. the kinetic energy is more or less unchanged by the λ-transition). Therefore the bosons in the low-lying momentum states cannot all condense into the ground momentum state (or indeed into any one momentum state) by the non-linear effect canvassed in the preceding paragraph because this would give a discontinuity in the kinetic energy and hence a latent heat for the λ-transition. In order to avoid such a latent heat the envelope of occupancy of momentum states must be more or less the same before and after the λ-transition (cf figure 8.11).

As mentioned in conjunction with figure 8.11, the relative fluctuations in highly occupied states are of order unity, equation (7.156), which implies that there is a non-negligible probability of finding some empty or occupied by only a few bosons, and others with an occupancy many times their average. These instantaneous spikes in the occupancy of low-lying momentum states, as depicted in figure 8.11, enable the present picture that condensation consists of multiple multiply-occupied momentum states, at least instantaneously. In fact the lifetime of highly occupied states could well be extended by their non-local permutation entropy (section 9.3).

The arguments against condensation being solely into the ground state also include mechanical and trajectory considerations. It is difficult to envisage forces and collisions that would conspire to cause a macroscopic number of bosons to lose their momenta and condense into the zero momentum state instantaneously at the λ-transition. If instead such condensation occurred continuously over a finite time interval, then some of the intermediate low-lying momentum states must become highly occupied, in which case it is difficult to design forces and collisions that would subsequently empty them to achieve zero momentum. As will be shown

below, it is highly unlikely for forces to occur that will empty a state one boson at a time (sections 9.4.1, 9.4.3, and 9.4.4).

A subsystem of size 1 cm³ at 1 K has $N = 2.2 \times 10^{22}$ helium atoms, and on the order of $\left(2\,mk_{\mathrm{B}}T/\Delta_p^2\right)^{3/2} = 2^{3/2}L^3/\Lambda^3 = 2.7 \times 10^{20}$ accessible momentum states. This averages out to 10^2 helium atoms in each accessible momentum state, independent of subsystem size. This is a crude estimate based on setting the Maxwell–Boltzmann kinetic energy exponent to unity.

As mentioned, fluctuations mean that at any instant many low-lying momentum states are empty or few-occupied, equation (7.156). Due to the relatively small gap between kinetic energy states, the permutation entropy might coalesce the bosons in any kinetic energy neighborhood into a single momentum state. Provided that there were many such condensed states, the overall kinetic energy distribution would remain unchanged while the majority of low-lying momentum states would be unoccupied. This is what is sketched in figure 8.11. The advantage of this picture is that it would leave the total kinetic energy unchanged, in accordance with the absence of a measured latent heat, it would enable many large pure momentum loops to form because now large numbers of bosons are in the same momentum state, albeit with very many such distinct occupied states, and it would necessarily leave many if not most low-lying momentum states empty. Evidence is presented in subsection 9.4.3 that the great majority of low-lying momentum states are instantaneously empty. These arguments suggest that the instantaneously occupied low-lying momentum states have much greater than 10^2 bosons in each.

9.3 Interactions on the far side

The treatment of the λ-transition for interacting bosons with computer simulations in chapter 8 is reliable on the high temperature side, which is dominated by the position loops that form the basis of the loop expansion algorithm. The thermal wavelength that defines those loops arises from independent integration over the excited momentum state continuum, which means that the bosons are in different momentum states. The binary division into ground or else the set of excited momentum states does no harm on the high temperature side of the transition.

Below the transition the dominance of momentum loops implies the increasing occupancy of individual low-lying momentum states, which makes the binary division problematic. It is no longer appropriate to make the continuum approximation for the excited states because then one cannot identify the occupancy of individual momentum states. Nor is it appropriate to assert that the ground momentum state is the only discrete momentum state. In order to explore the behavior of the system at very low temperatures one has to carry out the analysis for discrete momentum states.

The analytic but approximate approach pursued here is motivated on physical grounds. It partially explains mathematically the otherwise puzzling success of the ideal boson model (section 8.2 and 8.3) applied to interacting ^{4}He by demonstrating the insensitivity of the system to the interaction potential at low temperatures.

9.3.1 Factorization and the heat capacity

We proceed with discrete momentum states. The canonical equilibrium partition function for bosons, equation (7.71) or equation (8.1), neglecting the commutation

function, and with a momentum eigenvalue being $p_{j\alpha} = n_{j\alpha}\Delta_p$, $\Delta_p = 2\pi\hbar/L$, is

$$
\begin{aligned}
Z^+(N,V,T) &= \frac{1}{N!} \sum_{\mathbf{n}} \sum_{\hat{P}} \langle \mathbf{n} \mid e^{-\beta\hat{\mathcal{H}}} \mid \hat{P}\mathbf{n} \rangle \\
&\approx \frac{1}{N!} \sum_{\mathbf{n}} \sum_{\hat{P}} \frac{1}{V^N} \int d\mathbf{q}\, e^{-\beta\mathcal{H}(\mathbf{q},\mathbf{p_n})} e^{\mathbf{q}\cdot\mathbf{p_n}/i\hbar} e^{-\mathbf{q}\cdot\hat{P}\mathbf{p_n}/i\hbar} \\
&= \frac{1}{N!} \sum_{\hat{P}} \frac{1}{V^N} \int d\mathbf{q}\, e^{-\beta U(\mathbf{q})} \sum_{\mathbf{n}} e^{-\beta\Delta_p^2 n^2/2m} e^{\mathbf{q}\cdot(\mathbf{n}-\mathbf{n'})\Delta_p/i\hbar}.
\end{aligned}
$$

(9.1)

Here $j = 1, 2, \ldots, N$, $\alpha = x, y, z$, $\sum_{\mathbf{n}} = \prod_{j,\alpha} \sum_{n_{j\alpha}=-\infty}^{\infty}$, and $\mathbf{n'} \equiv \hat{P}\mathbf{n}$, where $\hat{P}$ is the permutation operator. The second equality sets the commutation function to unity, as argued in the introduction to section 8.4. The symmetrization function used here may be the complex conjugate of that used in the formulation of quantum statistical mechanics in chapter 7, $\eta_p = \eta_q^*$, which is not important because the commutation function here is real (it is unity), and since the partition function is real, the imaginary parts must sum to zero.

This section focusses on low temperatures, $\beta \to \infty$. In this case there is a limited number of accessible momentum states available, say $\mid n_{j\alpha} \mid \lesssim L/\Lambda$, where L is the cubic subsystem edge length and $\Lambda = \sqrt{2\pi\hbar^2\beta/m}$ is the thermal wavelength. The number of accessible momentum states per particle at the λ-transition can be estimated as $n_p = \Lambda^{-3}V \lesssim N$. This gives up to $\mathcal{O}(10^2)$ ^{4}He atoms on average in each accessible momentum state. In subsection 9.4.3 it will be estimated that below the λ-transition each occupied state is on average occupied by $\mathcal{O}(10^6)$ ^{4}He atoms. Under these circumstances the permutation loops are predominantly momentum loops in which all particles in a loop are in the same momentum state. In this case the momentum sum and the position integral factorize (because if $n_{j\alpha} - n_{j\alpha}' = 0$, then the Fourier factor is unity for all values of $\mathbf{q}$). The momentum factor that arises is

$$
Z^{+,\mathrm{id}}(N,V,T) = \frac{1}{N!} \sum_{\hat{P}} \prod_{j,\alpha} \sum_{n_{j\alpha}=-\infty}^{\infty} e^{-\beta\Delta_p^2 n_{j\alpha}^2/2m} \delta(n_{j\alpha} - n_{j\alpha}').
$$

(9.2)

This neglects mixed momentum permutation loops, which do not necessarily contribute zero upon integration because of the interaction potential in the integrand. However, when pure permutation loops give the dominant contribution, as is assumed here, then all the wave function symmetrization effects are contained within this momentum factor. Because of the factorization, this is purely an ideal gas result, and it is directly connected with the grand canonical ideal bosons results of section 8.2 and 8.3. What is different to that analysis is that here in the low temperature regime we avoid the binary division continuum approximation, and instead focus on the occupancy of the discrete momentum states.

The remaining factor contains the classical configuration integral, $Q^{\mathrm{cl}}(N,V,T) = \int \mathrm{d}\mathbf{q}\ \mathrm{e}^{-\beta U(\mathbf{q})}$, and so one has

$$Z^+(N,V,T) = Z^{+,\mathrm{id}}(N,V,T)\frac{Q^{\mathrm{cl}}(N,V,T)}{V^N}, \quad \beta \to \infty.$$

(9.3)

It must be emphasized that this factorization is predicated upon the neglect of the commutation function, and upon the explicit restriction to pure momentum permutation loops.

The average energy is just the logarithmic temperature derivative of this,

$$\begin{aligned}
\langle E \rangle^+_{N,V,T} &= \frac{-V^{-N}}{Z^+(N,V,T)}\left\{ Q^{\mathrm{cl}}(N,V,T)\frac{\partial Z^{+,\mathrm{id}}(N,V,T)}{\partial \beta} \right.\\
&\quad \left. + Z^{+,\mathrm{id}}(N,V,T)\frac{\partial Q^{\mathrm{cl}}(N,V,T)}{\partial \beta} \right\}\\
&= \langle E \rangle^{+,\mathrm{id}}_{N,V,T} + \langle U \rangle^{\mathrm{cl}}_{N,V,T}.
\end{aligned}$$

(9.4)

This is the sum of the quantum average energy for ideal bosons plus the classical average potential energy of their interactions. Hence the heat capacity at constant density is similarly the sum of the quantum ideal part and the classical interaction part. The former in binary division approximation (ground state plus continuum excited states) is derived in subsection 8.2.3.

Approaching absolute zero, the potential energy is practically constant as the structure of the fluid is insensitive to temperature, $\langle U \rangle^{\mathrm{cl}}_{N,V,T} \to$ const. Hence its contribution to the heat capacity must be negligible, and to a good approximation the dominant energy contribution to the heat capacity at constant density is that from ideal bosons. In other words, at low temperatures approaching absolute zero, the present analysis of interacting bosons shows that the classical excess contributes negligibly to the heat capacity. The ideal boson expression, equation (8.20), dominates even for interacting bosons. One concludes that interactions make little difference to the heat capacity deep below the transition temperature.

It should be emphasized that the main approximation in the above analysis is to restrict the permutation loops to those consisting of bosons in the same momentum state (pure momentum loops). This neglects permutation loops composed from bosons in different momentum states, which are the ones that give rise to the thermal wavelength in the momentum continuum. These give the dominant contribution to the heat capacity in the vicinity of the λ-transition, and so the present analysis cannot be expected to be accurate in that regime.

9.3.2 Permutation entropy for condensation

There is an alternative and arguably better way to view the preceding factorization that brings out its explicit relationship to entropy and that sheds light on the nature of the system in the superfluid regime.

The microstates of the system are $\mathbf{n} = \{\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_N\}$. The elements here are single-particle momentum states, with $n_{j\alpha} = 0, \pm 1, \pm 2, \ldots$, with $j = 1, 2, \ldots, N$ and $\alpha = x, y, z$, and the corresponding the phase space momentum is $\mathbf{p} = \Delta_p \mathbf{n}$, $\Delta_p = 2\pi\hbar/L$. A microstate has internal subsystem entropy due to permutations,

$$S_{\mathbf{n}}^{\mathrm{s},+} = k_{\mathrm{B}} \ln \chi_{\mathbf{n}}^+.$$
(9.5)

Here $\chi_{\mathbf{n}}^+$ is the symmetrization factor for bosons, which appears in the expression for the partition function, equation (7.71) or equation (8.1). It is the number of allowed permutations. Since the momentum state is the product of single-particle states this is (Attard 2021 equation (2.20))

$$\chi_{\mathbf{n}}^+ = \sum_{\hat{\mathrm{P}}} \langle \hat{\mathrm{P}} \mathbf{n} \mid \mathbf{n} \rangle = \prod_{\mathbf{a}} N_{\mathbf{a}}(\mathbf{n})!.$$
(9.6)

The only nonzero permutations are between bosons in the same momentum state, the number of which is equal to the occupancy of the state,

$$N_{\mathbf{a}}(\mathbf{n}) = \sum_{j=1}^{N} \delta_{\mathbf{n}_j, \mathbf{a}}.$$
(9.7)

Here $\mathbf{a} = \{a_x, a_y, a_z\}$ and $a_\alpha = 0, \pm 1, \pm 2, \ldots$. Since the logarithm of the partition function is the total entropy, and since the Maxwell–Boltzmann factor is the reservoir entropy, one can identify the symmetrization factor with the internal or subsystem entropy for the momentum states due to permutations.

Obviously we add this internal entropy to the canonical equilibrium reservoir entropy to get the total entropy of the subsystem microstate,

$$S_{\mathbf{n}}^+ = S_{\mathbf{n}}^{\mathrm{s},+} + S_{\mathbf{n}}^{\mathrm{r}} = k_{\mathrm{B}} \ln \chi_{\mathbf{n}}^+ - \frac{\mathcal{H}(\mathbf{q}, \mathbf{p})}{T}.$$
(9.8)

Here $\mathcal{H}(\mathbf{q}, \mathbf{p}) = \mathcal{K}(\mathbf{p}) + U(\mathbf{q})$ is the classical Hamiltonian phase space function.

The specific contribution from the permutation entropy can be elucidated in terms of a trial momentum move for particle j, $\mathbf{p}_j \to \mathbf{p}_j\prime$. If $\mathbf{p}_j = \mathbf{a}\Delta_p$ and $\mathbf{p}_j\prime = \mathbf{b}\Delta_p$, then this may be written

$$\mathbf{a} \to \mathbf{b}, \quad \text{or} \quad \{N_{\mathbf{a}}, N_{\mathbf{b}}\} \to \{N_{\mathbf{a}} - 1, N_{\mathbf{b}} + 1\}.$$
(9.9)

Obviously $N_{\mathbf{a}} \geqslant 1$. The change in the internal subsystem entropy for this is

$$\Delta S_{\mathbf{n}}^{\mathrm{s},+}(\mathbf{a} \to \mathbf{b}) = k_{\mathrm{B}} \ln \frac{N_{\mathbf{b}} + 1}{N_{\mathbf{a}}}.$$
(9.10)

This is positive if the number of bosons in the destination state ends up being greater than the number in the initial state (the rich get richer). This what drives the bosons' momenta to become highly correlated at low temperatures, the formation of large momentum loops, and increased momentum state occupancy.

The change in reservoir entropy, $\Delta S_{\mathbf{n}}^{\mathrm{r}}(\mathbf{a} \to \mathbf{b})/k_{\mathrm{B}} = -\beta(b^2 - a^2)\Delta_p^2/2m$, should be added to this. This tends to drive the particles into the momentum ground state. Assuming that structure, not dynamics, are of interest, the particles positions can be independently changed, $\Delta S^{\mathrm{r}}(\mathbf{q} \to \mathbf{q}')/k_{\mathrm{B}} = -\beta(U(\mathbf{q}') - U(\mathbf{q}))$. The changes in momentum are conveniently done one particle at a time, and similarly for position. The changes in entropy can be used in conjunction with, for example, the Metropolis algorithm, to generate a trajectory through (semi-discrete) phase space (which includes the ground momentum state) that ensures that the trajectory average equals the equilibrium average. They can also be used to formulate stochastic dissipative equations of motion (see sections 3.2.4, 4.1, and 5.4). More sophisticated equations of motion for superfluid flow are given in section 9.4.4.

The canonical equilibrium partition function for bosons, equation (9.1), is

$$
\begin{aligned}
Z^+(N, V, T) &= \frac{1}{N!} \sum_{\mathbf{n}} \sum_{\hat{P}} \langle \phi_{\mathbf{n}}(\mathbf{q}) \mid e^{-\beta\hat{\mathcal{H}}} \mid \phi_{\mathbf{n}}(\hat{P}\mathbf{q}) \rangle \\
&= \frac{1}{N!V^N} \sum_{\mathbf{n}} \sum_{\hat{P}} \int d\mathbf{q} \; e^{\mathbf{q}\cdot\mathbf{p_n}/i\hbar} e^{-\beta\hat{\mathcal{H}}} e^{\mathbf{q}\cdot\hat{P}\mathbf{p_n}/i\hbar} \\
&\approx \frac{1}{N!V^N} \sum_{\mathbf{n}} \sum_{\hat{P}} \int d\mathbf{q} \; e^{\mathbf{q}\cdot\mathbf{p_n}/i\hbar} e^{-\beta\mathcal{H}(\mathbf{q},\mathbf{p_n})} e^{\mathbf{q}\cdot\hat{P}\mathbf{p_n}/i\hbar} \\
&\approx \frac{1}{N!V^N} \sum_{\mathbf{n}} \int d\mathbf{q} \; e^{-\beta\mathcal{H}(\mathbf{q},\mathbf{p_n})} \chi_{\mathbf{n}}^+ \\
&= \frac{1}{N!V^N} \sum_{\mathbf{n}} e^{-\beta\mathcal{K}_{\mathbf{n}}} e^{S_{\mathbf{n}}^{\mathrm{s},+}/k_{\mathrm{B}}} \int d\mathbf{q} \; e^{-\beta U(\mathbf{q})}.
\end{aligned}
$$

(9.1
1)

The third equality neglects the commutation function. The fourth equality makes the same approximation as in the preceding subsection, namely it neglects permutations between bosons in different momentum states on the grounds that these give highly oscillatory Fourier factors. This means that we only need retain permutations amongst bosons in the same momentum state, the number of which is given by $\chi_{\mathbf{n}}^+$, and the logarithm of which number gives the internal subsystem entropy of the microstate. The final factor is just the configurational integral, $Q^{\mathrm{cl}}(N, V, T)$, and so this expression is the same as that given in the preceding subsection, equation (9.3), except that here the sum over permutations has been performed explicitly to give the subsystem internal microstate entropy.

The symmetrization factor differs from unity only at low temperatures when there are a restricted number of momentum states accessible so that there is high probability of some being occupied by multiple bosons. It should be noted that the

permutation entropy of a momentum state depends only upon its occupancy, not upon the value of its momentum. In this sense there is no qualitative difference between the ground momentum state and low-lying excited momentum states (section 8.6.6 and 9.2). It is the entropy embodied in the symmetrization factor, and its increase with increasing momentum correlation, that is ultimately responsible for superfluidity below the λ-transition.

9.4 Entropy of superfluidity

The puzzle with superfluidity is the physical mechanism by which bosons flow without viscosity. One hint to the answer is the fact that the onset of superfluidity coincides with the λ-transition. As shown in chapter 8, in the first approximation this signifies Bose–Einstein condensation into the zero momentum state. More detailed arguments in sections 8.6.6, 9.2, and 9.3.2 extend the simple binary picture to say that condensation represents the occupancy of multiple low-lying momentum states, not just the ground momentum state.

The coexistence of He II and He I below the λ-transition is consistent with the condensed bosons being the ones in highly occupied momentum states, and the non-condensed bosons being the ones in singly (or few) occupied momentum states. This binary division based upon occupancy is likely a better approximation than the binary division developed by F London (1938) based upon ground and excited states. One can see that below the λ-transition ^{4}He is a mixture of bosons in highly occupied (He II) and in few-occupied (He I) momentum states, with the proportion of He II increasing with decreasing temperature. Superfluidity is carried by the bosons in the highly occupied momentum states, as is now discussed.

9.4.1 Momentum loop entropy in collisions

In section 9.3.2 the permutation entropy for condensation into pure momentum permutation loops was analyzed. The change in the total entropy for a change in the momentum state of a single boson was given as (cf equations (9.8) and (9.10))

$$k_{\mathrm{B}}^{-1}\Delta S_{\mathbf{n}}^{+}(\mathbf{a}\rightarrow\mathbf{b})=\ln\frac{N_{\mathbf{b}}+1}{N_{\mathbf{a}}}-\frac{\beta\Delta_p^2}{2m}(b^2-a^2).$$

(9.12)

The internal permutation entropy, the first term on the right-hand side, will increase if the destination state ends up with greater numbers than were in the original state. The reservoir entropy, the second term on the right-hand side, will increase if the destination state has lower kinetic energy than the original state. These two terms are the deterministic or dissipative contribution to the evolution and maintenance of the equilibrium distribution of momentum states. One can also include the mechanical forces if one were to implement these to generate a trajectory over time. This is a bowdlerized version of the stochastic dissipative equations of motion (3.86), (4.21), and (5.50). A more complete description of the equations of motion for superfluid flow is given in section 9.4.4.

One can draw two important conclusions from this result. First, if the two states have the same momentum magnitude, then this tends to drive the boson into the most highly occupied state. This supports the suggestion above that it is possible that only one or a few momentum states in each kinetic energy neighborhood are

occupied. Second, the reservoir term tends to drive bosons into the momentum ground state. However, the reservoir term is extremely weak, being $\mathcal{O}(10^{-14})$.

On both these points one has to be aware that there has to be added a stochastic contribution to the equations of motion (see sections 3.2.4, 4.1, and 5.4). The randomizing effect of this tends to drive bosons out of the ground momentum state, and out of a single momentum state in any given kinetic energy neighborhood. So despite the fact that the dissipative terms given explicitly above favor ground momentum state occupation, the reality is that the picture enunciated above in section 9.2, and earlier in subsection 8.6.6, is most likely correct: condensation occurs into the low-lying momentum states, rather than solely into the ground momentum state. The superfluid equations of motion conform to this picture (section 9.4.4).

In view of these arguments and the very small spacing between momentum states estimated for ^{4}He below the λ-transition, $\beta\Delta_p^2/2\,m = 2 \times 10^{-14}$ for $T = 1$ K and $L = 1$ cm, the equilibrium probability distribution for momentum states likely takes the form of a Gaussian peaked at the ground momentum state, $\bar{n} = \{0, 0, 0\}$, with a rather broad width that encompasses on the order of 10^{14} low-lying momentum states in each direction. It is probably the case that at any one instant the majority of the individual momentum states within this Gaussian envelope are unoccupied.

For the case of steady flow down a narrow channel or capillary driven by an external force or pressure gradient, the most likely momentum state is of the form $\bar{n} = \{\bar{n}_x, 0, 0\}$, for some $\bar{n}_x > 0$. Again there would be a broad, essentially one-dimensional, Gaussian envelope of momentum states about this, most of which are unoccupied at any one time. (The transverse width of the Gaussian would be $\mathcal{O}(\bar{n}_x L_y^2/L_x^2)$.) Since the occupied states have a high number of bosons in them, their permutation entropy is high. Unless the collision force is large, or the momentum state is high, any attempt to shift an individual boson out of an occupied state and into an unoccupied state would most likely not succeed because of the loss of part of the permutation entropy of the occupied state. The probability is that such a force would not occur in the first place. If the forces were such as to attempt to transfer a boson from one occupied state to another occupied state, then there may be a change of momentum state, depending upon the relative occupancies. Of course the only destination states where this could happen are those likely occupied in accordance with the probability distribution centered on $\bar{n}$. A more sophisticated presentation of the equations of motion for superfluid flow is given in section 9.4.4.

9.4.2 Non-local momentum correlations and plug flow

The non-local nature of the momentum correlations induced by the pure momentum permutation loops is the fundamental cause of superfluidity. Since the λ-transition marks the change in dominance from position to momentum permutation loops (section 8.6.2), and since the former are localized in space and the latter aren't, it is clear that non-localization must be essential for superfluidity.

Recall the molecular origin of shear viscosity in a classical fluid. In shear flow, the inhomogeneous momentum flux is dissipated by molecular collisions. In effect faster moving fluid layers are slowed and adjacent slower layers are accelerated as

momentum is transferred between them. This momentum dissipation has the effect of increasing the subsystem entropy, since the order represented by spatial variations in momentum flux is a state of low configurational entropy (cf section 3.4 for heat flow, and Attard (2012 section 9.5) for shear flow).

In general for a classical fluid, or for a quantum fluid above the λ-transition where position permutation loops dominate, the momentum correlations must be spatially localized, which gives rise to shear flow. For example, for Poiseuille flow, which is laminar flow in a pipe or channel under a pressure gradient, the flow is zero at the walls and increases continuously toward the center.

In the case of bosons below the λ-transition, those bosons in the same highly occupied momentum state are correlated without regard to separation. Such non-local momentum correlations prohibit the Poiseuille parabolic profile, since the high momentum state in the center of the channel would induce the same state in the bosons near the walls. Non-local momentum correlations imply momentum spatial homogeneity: the only mathematical solution for nonzero flow is plug flow in which the most likely momentum state of the bosons is uniform across the channel. This of course is also the classical solution for inviscid hydrodynamic flow.
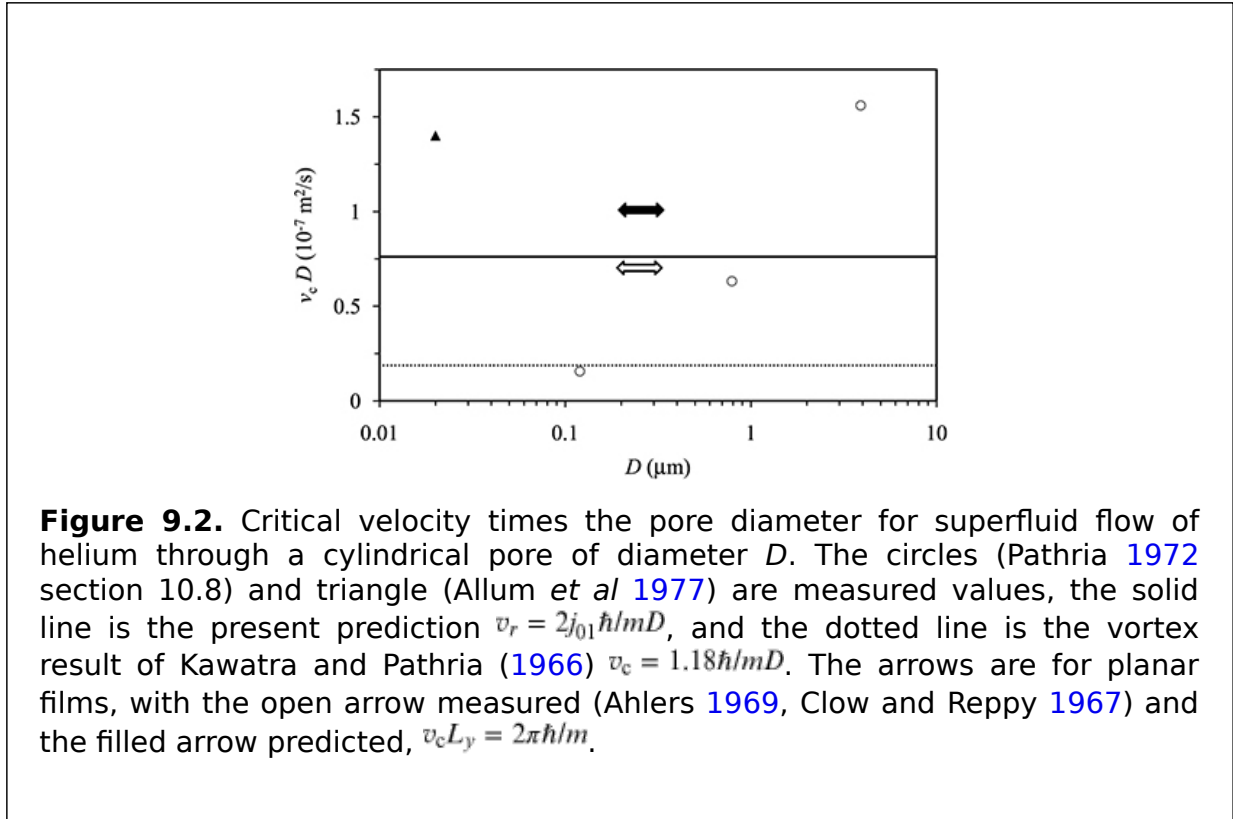
9.4.3 Critical velocity

Laboratory measurements show the existence of a critical velocity above which superfluid flow through pores is destroyed. Pathria (1972 section 10.8) gives measured experimental values for the critical velocity of $v_c=$ 0.13, 0.08, and 0.04 m for capillary diameter $D = 0.12, 0.79$, and 3.9 μm, respectively. The common explanation is that above a critical velocity the production and growth of excitations destroy the superfluid. The type of excitations generally invoked are associated with the rotons of Landau's (1941) theory, which are commonly interpreted as vortex rings.

In the context of the present approach to the λ-transition based upon momentum state occupancy, the fact that the critical velocity increases with decreasing pore diameter suggests instead that we should compare it to the velocity of the first excited transverse momentum state. (The theoretical basis for this prediction is derived below following the comparison with measured data.) For the rectangular geometry that we have focussed on thus far this is $v_y = \Delta_{py}/m = 2\pi\hbar/mL_y$. (This would also apply to a liquid film, where the vanishing of the wave function at the solid surface and at the liquid–vapor interface creates just such a transverse momentum gap.) For a cylinder of diameter D, the transverse momentum gap is $\Delta_{pr} = 2j_{01}\hbar/D$, where $j_{01} = 2.4$ is the first zero of the zeroth order Bessel function (Blinder 2011). Hence according to the present theory the critical velocity for a cylindrical pore is $v_r = 2j_{01}\hbar/mD$.

This predicted critical velocity is tested in figure 9.2. Over some three orders of magnitude in pore diameter it remains within a factor of three of the measured data. The Landau stability criterion for superfluid flow gives a critical velocity of about 60 m s^{-1}, which is several orders of magnitude larger than the measured values (Batrouni *et al* 2004, Balibar 2017). The vortex prediction of Kawatra and Pathria (1966), which implements Feynman's (1954) suggestion that the rotons invented by Landau (1941) in his theory of superfluidity were in fact quantized vortices, is also compared in the figure. Although limited, the present theory lies closer to the experimental data in figure 9.2 than does the vortex/roton theory.

For the case of a planar (actually annular) film, Ahlers (1969) measured the critical velocity times the thickness, which was estimated as 0.25 μm. Ahlers (1969) found that sufficiently below the λ-transition this approaches a constant value of $v_c L_y = 7.1 \times 10^{-8}$ m² s⁻¹. The present prediction, $v_c L_y = 2\pi\hbar/m = 9.98 \times 10^{-8}$ m² s⁻¹, compares agreeably with the measured value (figure 9.2). The results of Ahlers (1969) agree with those obtained by Clow and Reppy (1967) for thickness 0.2 μm using a different method.

Figure 9.2. Critical velocity times the pore diameter for superfluid flow of helium through a cylindrical pore of diameter D. The circles (Pathria 1972 section 10.8) and triangle (Allum *et al* 1977) are measured values, the solid line is the present prediction $v_r = 2j_{01}\hbar/mD$, and the dotted line is the vortex result of Kawatra and Pathria (1966) $v_c = 1.18\hbar/mD$. The arrows are for planar films, with the open arrow measured (Ahlers 1969, Clow and Reppy 1967) and the filled arrow predicted, $v_c L_y = 2\pi\hbar/m$.

It should be mentioned that for superfluid flow in a film, collisions that change momentum parallel to the surface do not dissipate momentum or create shear flow. It is only collisions that change momentum normal to the surface that must be suppressed. Hence it is the momentum gap normal to the surface that determines the critical velocity.

As discussed in subsection 9.4.1, superfluid flow down a channel in the *z*-direction most likely occurs in a ground transverse momentum state, $\bar{\mathbf{n}} = \{0, 0, \bar{n}_z\}$. The probability distribution of momentum states is peaked about this, with a relatively broad width within which many momentum state are probably empty. The occupied states in the vicinity of $\bar{\mathbf{n}}$ likely have large numbers of bosons in each. Since the channel is usually much longer than it is wide, the spacing of momentum states in the flow direction is much less than those in the transverse direction. Again, as discussed in subsection 9.4.1, such flow with bosons in the transverse momentum ground state is arguably dissipationless, in order that the flow be inviscid. It will be shown shortly that when the flow velocity exceeds the velocity of the first excited transverse momentum state, interactions and collisions with the

wall are able to excite the bosons out of the state $\bar{\mathbf{n}}$, thereby destroying the superfluid flow.

The measured critical velocities are several orders of magnitude smaller than a component of the thermal velocity of an individual boson, which at $T = 1$ K is $v_{\text{th}} \equiv \sqrt{k_{\text{B}}T/m} = 46$ m s^{-1}. But this measure is not entirely appropriate because the longitudinal momentum states with smaller quantum numbers have greater occupancy and permutation entropy; one cannot really consider them individually. The external force or pressure gradient that drives the superfluid flow in the pore gives rise to the non-equilibrium probability distribution that is peaked at $\bar{n}$.

The stochastic dissipative equations of motion provide a quantitative estimate of the effect of a collision. Let $N_{\bar{\mathbf{n}}}$ be the occupancy of the most likely state, and suppose that this is a relatively large number. Let $\bar{v}_z = \bar{n}_z \Delta_{pz}/m$ be the flow velocity for this initial state. Consider the final state to be the first excited transverse momentum state and longitudinal ground state, $\mathbf{n}' = \{1, 0, 0\}$, with velocity $v_\perp = \Delta_{p\perp}/m$. Consider two possible collisions with the wall related to this final state. In the first case one particular boson transitions to the final state. In this case the changes in internal and reservoir (wall) entropy from the collision are (cf equation (9.12))

$$\Delta S^{\text{int}} = k_{\text{B}} \ln\left[\frac{(N_{\bar{\mathbf{n}}} - 1)!}{N_{\bar{\mathbf{n}}}!} \frac{1!}{0!}\right] = -k_{\text{B}} \ln N_{\bar{\mathbf{n}}}, \quad \text{and} \quad \Delta S^{\text{r}} = \frac{-m(v_\perp^2 - \bar{v}_z^2)}{2T},$$

(9.13)

respectively. The first term is negative and large in magnitude because $N_{\bar{\mathbf{n}}}$ is large. The second term is small because the velocities $v_\perp$ and $\bar{v}_z$ are much less than the thermal velocity. Hence the total change in entropy is negative, which means that the dissipative part of the stochastic equations of motion suppress such individual boson collisions. This is a dynamic example of the tendency of bosons to occupy the same state.

The second type of collision to be considered is when all of the bosons in the pore are excited into the state $\mathbf{n}'$. This might be called an occupied state collision. In this case the respective changes in entropy are

$$\Delta S^{\text{int}} = k_{\text{B}} \ln\left[\frac{0!}{N_{\bar{\mathbf{n}}}!} \frac{N_{\bar{\mathbf{n}}}!}{0!}\right] = 0, \quad \text{and} \quad \Delta S^{\text{r}} = \frac{-mN_{\bar{\mathbf{n}}}(v_\perp^2 - \bar{v}_z^2)}{2T}.$$

(9.14)

For this occupied state collision the decrease in the internal entropy of the initial state cancels with the increase for the final state. Hence the nett change in permutation entropy is zero. In contrast, the change in reservoir entropy is now large, because it scales with the occupancy of the most likely state. It is positive if and only if the flow velocity $\bar{v}_z$ exceeds the velocity of the first excited transverse momentum state $v_\perp$. Because it is macroscopic, its exponential is effectively a Heaviside step function so that the dissipative part of the stochastic equations of motion either forbids or else mandates such an occupied state collision and transition, depending on which of the two speeds is largest.

This argument only works when the occupancy is sufficiently large, which fact can be used to estimate a lower bound. Taking the difference in the kinetic energies

to be about equal to the kinetic energy at the critical flow velocity, the change in reservoir entropy can only be much greater than unity if

$$N_{\bar{\mathbf{n}}} \gg \frac{2k_\mathrm{B}T}{mv_\mathrm{c}^2} = 2\frac{v_\mathrm{th}^2}{v_\mathrm{c}^2}.$$

(9.1 5)

With $v_\mathrm{th} = 50$ m s^{-1} and $v_\mathrm{c} = 0.1$ m s^{-1}, this gives $N_{\bar{\mathbf{n}}} \gtrsim 10^6$. This is substantially larger than the average occupancy estimated for the accessible low-lying momentum states from the thermal energy, $N_\mathbf{n} = \mathcal{O}(10^2)$. This suggests that most of the low-lying momentum states are unoccupied. Since this lower bound is very much less than the total number of bosons in the channel, one can also see that it is not necessary for only one momentum state to be occupied for superfluidity to occur. That is, the existence of a critical velocity for superfluid flow suggests that condensation need not be solely into the momentum ground state.

Although the above collision calculation has been performed for the transition between the two specific states $\bar{\mathbf{n}}$ and $\mathbf{n}\prime$, it should be clear that similar conclusions will hold for any initial occupied state in the vicinity of $\bar{\mathbf{n}}$, and any final state out of that neighborhood. Also, because forces are local but occupancy is non-local, it may be that not all of the bosons in the most likely momentum state instantaneously change state. This leads to the idea of partitions of the occupied states, as is discussed in the following presentation of the superfluid equations of motion.

The first calculation shows that when the most likely state is occupied by macroscopic numbers of bosons, then excitation or transition by individual bosons is effectively forbidden. The second calculation shows that the measured critical velocity is the velocity of the transverse momentum spacing, $v_\mathrm{c} \equiv \Delta_{p,\perp}/m$, and that the necessary and sufficient condition for stable superfluid flow is that the flow velocity must be less than this critical velocity.

9.4.4 Superfluid equations of motion

The key remaining challenge is to explain mechanistically how superfluid flow actually occurs. Whilst the preceding subsections point out that the loss of permutation entropy for highly occupied momentum states makes momentum-changing collisions improbable, they remain vague on how or whether such collisions are actually avoided. Recall Landau's (1941) criticism: 'nothing would prevent atoms in a normal state from colliding with excited atoms, ie. when moving through the liquid they would experience a friction and there would be no superfluidity at all'. To answer this a more detailed picture of the trajectory with specific equations of motion is discussed.

The general form of the stochastic dissipative equations of motion is based on the second entropy of non-equilibrium thermodynamics (Attard 2012 sections 3.6 and 7.4.5). Similar to equation (5.50), the equations of motion are

$$\boldsymbol{\Gamma}(t+\tau) = \boldsymbol{\Gamma}^0(t+\tau \mid \boldsymbol{\Gamma}(t), t) + \mathbf{R}_p(t).$$

(9.1 6)

A point in classical phase space is $\boldsymbol{\Gamma} \equiv \{\mathbf{q}, \mathbf{p}\}$ and the gradient operator is $\nabla \equiv \{\nabla_q, \nabla_p\}$. The thermostat contribution has only momentum components,

$\mathbf{R}_p\left(t\right) = \overline{\mathbf{R}}_p\left(t\right) + \widetilde{\mathbf{R}}_p\left(t\right)$, with the dissipative and stochastic forces satisfying

$$\overline{\mathbf{R}}_p(t) = \frac{\sigma^2}{2k_{\mathrm{B}}}\nabla_p S(\mathbf{\Gamma}(t)), \text{ and } \langle\widetilde{\mathbf{R}}_p(t)\widetilde{\mathbf{R}}_p(t')\rangle = \sigma^2\delta_{t,t'}\mathrm{I}_{pp},$$

(9.17)

with $\langle\widetilde{\mathbf{R}}_p\rangle = 0$. The thermostat should represent a small perturbation on the adiabatic trajectory.

In the present quantum case, $S(\mathbf{\Gamma}) = S^{\mathrm{r}}(\mathbf{\Gamma}) + S^{\mathrm{int}}(\mathbf{p})$, where the reservoir entropy is $S^{\mathrm{r}}(\mathbf{\Gamma}) = -\mathcal{H}(\mathbf{\Gamma})/T$, and the permutation entropy depends upon the occupancy of the momentum states, $S^{\mathrm{int}}(\mathbf{p}) = k_{\mathrm{B}}\sum_{\mathbf{a}}\ln N_{\mathbf{a}}!$. This expression for the total entropy is the same as in the treatment on the far-side of the λ-transition in section 9.3. It includes only permutations of bosons in the same momentum state (pure momentum loops). It precludes mixed loops, which, as discussed in section 8.5, are important on the high-temperature side of the λ-transition, where they create a barrier to condensation. The present treatment in principle allows dynamic properties to be obtained, which was not possible with the analytic factorization given in section 9.3 for the static case.

Take the quantum adiabatic velocity to be

$$\dot{\mathbf{\Gamma}}^0 = -T\nabla^{\dagger}S(\mathbf{\Gamma}), \quad \nabla^{\dagger} \equiv \{\nabla_p, -\nabla_q\}.$$

(9.18)

This formulation ensures that time averages equal phase space averages (section 5.2.3). Hamilton's equations of motion, which are the classical adiabatic equations of motion, is this with $S \Rightarrow S^{\mathrm{r}} \equiv -\mathcal{H}/T$. As in the classical case, these quantum adiabatic equations of motion are time reversible (cf section 5.4.2). As in the classical case, these equations are incompressible, $\nabla \cdot \dot{\mathbf{\Gamma}}^0 = 0$, so that volume elements are conserved on a quantum adiabatic trajectory. And also as in the classical case, the adiabatic rate of change of entropy vanishes, $\dot{\mathbf{\Gamma}}^0 \cdot \nabla S = 0$, and so the entropy is also conserved on a quantum adiabatic trajectory. Hence the quantum probability density in phase space, $\wp(\mathbf{\Gamma}) = Z^{-1}e^{S(\mathbf{\Gamma})/k_{\mathrm{B}}}$, is a constant of the quantum adiabatic motion. Of course, the need to conserve the equilibrium probability is a statistical requirement.

The momentum $\mathbf{p} = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N\}$ is a continuous variable that is not quantized. It is non-trivial to combine this with the occupation of discrete states and the permutation entropy. The gradients of the total entropy are required for the adiabatic and dissipative forces. This is straightforward for the reservoir entropy since these are just the gradients of the Hamiltonian phase space function. However the permutation entropy requires the occupancies of discrete momentum states, which are necessarily discontinuous. Whilst it is easy to obtain the number of bosons in each momentum state given the spacing Δ_p, to define the momentum gradient of the occupancy, which is necessary for the $\nabla_p S^{\mathrm{int}}(\mathbf{p})$ that appears in the above equations of motion, in practice a continuous definition of occupancy can be used. Computer simulations of molecular interactions and collisions in superfluid

helium-4 using the present equations of motion have been successfully performed (Attard 2023, to be published).

The fundamental conceptual conclusion that can be drawn from these equations of motion is that the rate of change of the permutation entropy must vanish on the quantum adiabatic trajectory,

$$\dot{S}^{\text{int},0}(\mathbf{p}) = 0.$$

(Actually it is the total entropy that must be constant. But in so far as in any significant collision this is dominated by the permutation entropy, for the present discussion one may focus on the low probability of large changes in the latter.) This means that in an adiabatic collision all bosons in an occupied state either remain in the same state, or else all change momentum to a single new state (ie. the occupied state as a whole changes). For a singly occupied state this is possible, but it is difficult to imagine for a highly occupied one. One could conceive of a singleton boson colliding with a boson in a highly occupied state, and the two swapping positions (also known as a transposition permutation). This would leave the permutation entropy unchanged. For a singleton boson that has a glancing collision with a boson in a highly occupied state, the bosons in the highly occupied state must continue on unchanged in order to preserve the permutation entropy, which, from momentum conservation, means that the singleton boson is also undeflected. If the bosons in a highly occupied state were all to identically change their state in response to, say, a head-on collision between one of them and a singleton boson or wall particle, or because the momentum gap was sufficiently small, then this would imply that the total force on the state is divided equally amongst them in a non-local fashion. In this case the bosons in a highly occupied state would behave like a perfectly rigid heavy object with mass proportional to the occupancy number.

These equations show that it is unlikely for condensed bosons (ie. those in a highly occupied momentum state) to exchange momentum with uncondensed bosons (ie. those in a singly or few-occupied momentum state). Obviously there is a probabilistic element to this. For driven superfluid flow in a channel or film, the relatively large transverse momentum gap, $\Delta_{p,\perp}$, means that in a collision with a boson in a highly occupied state, the change in transverse momentum for a condensed boson is less than $\Delta_{p,\perp}$. Thus the change in the transverse momentum state is forbidden. Since the transverse momentum of the condensed bosons cannot change, neither can that of the non-condensed boson involved in the collision, since momentum must be conserved. In these circumstances where the transverse momentum gap suppresses the lateral transfer of longitudinal momentum, one can conclude that there can be no shear force on highly occupied states. In the second entropy formulation of variational hydrodynamics (section 3.5), the most likely viscous pressure tensor (ie. the traceless momentum flux tensor) is proportional to the symmetric traceless velocity gradient tensor (Attard 2012 equation (5.61)),

$$\underline{\underline{\Pi}}^{*}\left(\mathbf{r}\right) = -2\eta \, \text{sign}\left(\tau\right) \left[\nabla \mathbf{v}(\mathbf{r})\right]^{*,\text{sym}},$$

where η is the shear viscosity. For condensed bosons we have just shown that the off-diagonal elements of the momentum flux tensor are zero, whereas the velocity gradient tensor is non-zero for uncondensed bosons in shear flow (cf section 9.5.7). This means that there is no coupling between the momentum of condensed bosons and that of the non-condensed bosons in shear flow. Since η is the proportionality constant between the two, this implies that the shear viscosity for condensed bosons vanishes.

In conclusion, these equations of motion explain dissipationless superfluid flow. They answer Landau's (1941) objection that Bose–Einstein condensation is incompatible with superfluidity because collisions between condensed and uncondensed bosons would cause friction, or so he believed. What these equations show is that the entropy is conserved on an adiabatic trajectory, which means that bosons in highly occupied momentum states are unlikely to individually change their momentum in collisions. Nor can they all change to a single new momentum state, not unless the total force on the state acts on all of them in equal fraction in a non-local fashion.

9.5 Fountain pressure

9.5.1 Introduction

One of the earliest and perhaps still the most spectacular manifestation of superfluidity is the fountain effect. In this helium vigorously spurts from the open end of a heated tube that is connected by a capillary or microporous frit to a chamber of liquid helium maintained below the condensation temperature (Allen and Misener 1938, Balibar 2017). Measuring the fountain pressure with the heated chamber closed (apart from the connecting capillary) is a common experimental technique for obtaining the entropy (Donnelly and Barenghi 1998).

There is one common qualitative explanation, and one common quantitative treatment of the fountain effect. First the explanation, which is that the physical origin of the fountain effect is an osmotic pressure (Tisza 1938, Balibar 2017). The notion is based on Tisza's two-fluid model, which in turn is based on F London's (1938) proposal that the λ-transition in liquid helium is due to Bose–Einstein condensation. The two-fluid model says that helium consists of a mixture of He I and He II, which are excited state and ground state bosons, respectively. The latter comprise the superfluid and their fraction increases with decreasing temperature below the condensation temperature. This gives rise to the notion that osmosis drives He II selectively through the capillary from the low temperature, high concentration chamber to the high temperature, low concentration one. In general in a binary solution the mixing entropy favors concentration equality, and this is what is believed to create the osmotic pressure. A common conclusion drawn from the osmotic pressure explanation is that condensed bosons are attracted to heat.

In section 9.5.9 statistical mechanical analysis and calculations are performed of the osmotic prediction for the fountain pressure using both calculated (section 8.4) and measured (Donnelly and Barenghi 1998) data for the condensed boson fraction. It is shown that the predicted fountain pressures do not agree with measured values (Hammel and Keller 1961). Although these calculations are made at the level of ideal solution theory, the quantitative error is so large that the osmotic explanation for the fountain effect appears nonviable.

Actually, the osmotic pressure explanation could have been ruled out from almost the beginning since both He I and He II pass through the capillary. It is known experimentally from heat flow measurements that there is viscous flow of He I from the high temperature, high pressure chamber to the low temperature, low pressure one (F London and Zilsel 1948, Keller and Hammel 1960). This is quite contrary to the usual osmotic pressure, which only arises because a semi-permeable membrane excludes at least one of the components of the mixture from the flow.

Now to the accepted quantitative expression for the fountain pressure. H London (1939) carried out a thermodynamic analysis of the fountain effect that made a quantitative prediction for the pressure difference for a given temperature difference. His result is historically important as the quantitative experimental verification of this formula was taken to be evidence for the picture of superfluid helium being in a state of zero entropy. To the present day his expression remains significant as fountain pressure measurements are used to establish benchmark results against which calorimetric methods for the entropy of helium may be tested (Donnelly and Barenghi 1998). The results obtained from his expression are also used as calibration standards for instruments (Hammel and Keller 1961).

Elsewhere (Attard 2022b appendix A) I have presented detailed physical and mathematical arguments that the derivation of the H London (1939) expression is flawed. Further evidence is presented here of an internal contradiction in that derivation (section 9.5.4). These criticisms raise the question of whether the H London (1939) expression for the fountain pressure is a formally exact thermodynamic result, or whether it is merely a good approximation that fits the measured data. The distinction between the two can be appreciated from the role the expression has played in establishing benchmarks for calorimetry measurements and its use in instrument calibration. Also, the expression and its derivation have provided insight into the nature of superfluid helium, and the picture that emerges ought to be either rigorously confirmed or else refuted.

In section 9.5.2 four results are established: first, that the entropy is maximized when the chemical potential divided by the temperature of the two chambers is equal (equivalently, the fugacity is equal). Second, that the energy is minimized at constant entropy when the chemical potential of the two chambers is equal. Third, that equality of chemical potential is thermodynamically equivalent to the H London (1939) expression for the fountain pressure. And fourth, that these two expressions (equality of fugacity and equality of chemical potential) can be expected to give the same numerical result for the fountain pressure to within between about one part in ten and one part in one thousand, depending upon the temperatures.

In section 9.5.6 the three expressions are tested against measured fountain pressures. These use measured thermodynamic data, which tests are not straightforward because of errors that are identified in the fitted values of the enthalpy and the entropy derived by Donnelly and Barenghi (1998) from the measured heat capacity data. When these errors are corrected and the enthalpy and entropy are recalculated, the two thermodynamically equivalent forms of the H London (1939) expression are practically indistinguishable from each other and from the measured fountain pressures. The fountain pressures resulting from fugacity equality are 3%–5% below the measured values. It is tentatively concluded that the measured data support H London (1939) as being an exact expression. An

interpretation and rationalization of energy minimization at constant entropy as the principle that underlies superfluid flow is given in section 9.5.8.

9.5.2 Thermodynamic analysis

As mentioned, H London (1939) gave an expression for the temperature derivative of the fountain pressure that has been confirmed with high accuracy by experimental measurement. An important question is whether the derivation of H London (1939) is rigorous or whether it has been fudged. One must decide whether his expression is formally exact, or whether it is merely an accurate approximation. A critical analysis of H London's (1939) derivation of this expression is given in appendix A of Attard (2022b). An additional argument is presented in section 9.5.4 that the derivation is internally inconsistent. The jury's judgement is the Scottish one: his expression is not proven.

Consider two closed chambers of helium, A and B, each in contact with its own thermal reservoir of temperature T_A and T_B, and having pressure p_A and p_B. The chambers are connected by a capillary through which fluid can flow. Chamber A in practice is at the lower temperature, and consists of saturated liquid and vapor, but these points are presently unimportant. As H London (1939) points out, in the optimum steady state the pressure of the second chamber is a function of its temperature and the pressure and temperature of the first chamber, $p_B = p(T_B; p_A, T_A)$.

The result given by H London (1939) says that the derivative of the pressure of the second chamber with respect to its temperature for fixed first chamber equals the entropy density,

$$\frac{\mathrm{d}p_B}{\mathrm{d}T_B} = \rho_B s_B.$$

(9.2 1)

Here ρ is the number density and s is the entropy per particle. (I use lower case letters to denote quantities per particle; H London (1939) uses them to denote quantities per unit mass.) The derivative of the pressure and like quantities with respect to temperature for fixed first chamber parameters are total derivatives, which means that their integral along the fountain path can be evaluated analytically.

H London (1939) purported to derive this result using a work-heat flow cycle, which derivation I have reason to doubt (Attard 2022b appendix A). (See also section 9.5.4.) I note that it is not very hard to guess this result as the left-hand side has units of Boltzmann's constant per unit volume, and the only 'pure' thermodynamic quantity with those units is the entropy density. It is possible that H London (1939) first guessed the result, and retrospectively cooked up a derivation. In any case here I show that this result is equivalent to chemical potential equality of the two chambers, which I seek to locate within the broader principles of thermodynamics.

First I clarify a subtle point regarding extensivity. In equilibrium thermodynamics it is generally recognized that quantities such as energy, entropy, and free energy are extensive, which means that the total value is the sum of the individual values for independent or quasi-independent subsystems. For the present case of two subsystems of different temperatures this general rule does not hold for the free

energy. This is easily seen by considering the two to be quasi-independent such that the probability of a joint state is the product of the individual probabilities (see chapter 2, and also Attard (2002))

$$\begin{aligned}\wp(X_A, X_B) &= \wp_A(X_A)\wp_B(X_B) \\ &\propto e^{S_A(X_A)/k_B}e^{S_B(X_B)/k_B} \\ &= e^{-F_A(X_A)/k_BT_A}e^{-F_B(X_B)/k_BT_B}.\end{aligned}$$

(9.22)

Here k_B is Boltzmann's constant, and S and F are the appropriate total entropy and free energy. From this it is clear that the entropy is simply additive, $S_{tot}(X_A, X_B) = S_A(X_A) + S_B(X_B)$, but it is the free energy divided by temperature, rather than the free energy itself, that is additive, $F_{tot}(X_A, X_B)/T_{eff} = F_A(X_A)/T_A + F_B(X_B)/T_B$. It is only in the case of subsystems all with the same temperature that the free energy is extensive (i.e. simply additive).

In contrast to the free energy, by the usual principles of mechanics, the energy is simply additive $E_{tot}(X_A, X_B) = E_A(X_A) + E_B(X_B)$. This simple additivity is independent of the temperatures of the two chambers.

In searching for the thermodynamic principle that gives the H London (1939) expression (9.21), there are two possible axioms to be considered. The first is that the total entropy of the total system is a maximum, which is of course just the Second Law of Thermodynamics, albeit applied to a non-equilibrium steady-state system (see chapter 3, and also Attard (2012)). The second is that the total energy of the subsystems is a minimum, which has the appearance of a mechanical law and which is generally without thermodynamic relevance.

A third possible axiom, that the total free energy of the total system is a minimum is not viable for two related reasons. First, as just explained, it is not free energy but free energy divided by temperature that is the relevant thermodynamic potential, so it should be the sum of the free energy divided by temperature that is a minimum. And second, the free energy divided by temperature is derived directly from the total entropy, and so the principle of free energy divided by temperature minimization is not separate to, and does not yield any new information beyond, total entropy maximization. This can be seen by the trivial relationship between the second and third equalities in the above equation. To be clear and unambiguous on this point: for the case of systems with different temperatures, there is no principle of free energy minimization. There is a principle of free energy divided by temperature minimization, but this is no different to the principle of entropy maximization.

Consider now the first possible axiom, that the total entropy of the system is a maximum. Since the systems are closed, the total entropy is (section 2.3.2)

$$S_{tot} = \frac{-F(N_A, V_A, T_A)}{T_A} - \frac{F(N_B, V_B, T_B)}{T_B},$$

(9.23)

where N is the number, V is the volume, and F is the Helmholtz free energy. With the total number of helium atoms fixed, $N = N_A + N_B$, its derivative is (see equation (2.33))

$$\frac{\partial S_{\text{tot}}}{\partial N_A} = \frac{-\mu_A}{T_A} + \frac{\mu_B}{T_B},$$

(9.24)

where μ is the chemical potential. The maximum total entropy occurs when this is zero, which gives the condition for the optimum steady state that follows from the first possible axiom as

$$\frac{\mu_A}{T_A} = \frac{\mu_B}{T_B}.$$

(9.25)

Since the fugacity is $z = e^{\beta\mu}$, this condition is equivalent to

$$z_A = z_B.$$

(9.26)

To save words, I shall refer to this result as equal fugacity rather than equal chemical potential divided by temperature.

It should be emphasized that measurements of the fountain pressure involve two closed chambers and a heater. Hence it is a non-equilibrium steady-state system. Usually maximization of the first entropy plays no direct role in determining the optimum state of such systems (see chapter 3, and also Attard (2012)). Also, there is nothing specifically superfluid about this result, which may or may not be important.

Consider now the second possible axiom, that the total energy is a minimum. The energy of each chamber is a function of its entropy, volume, and number, $E_{\text{tot}} = E(S_A, V_A, N_A) + E(S_B, V_B, N_B)$, (see sections 2.3.4 and 2.4.4). In this case the derivative at fixed total N is (table 2.2 line 1)

$$\frac{\partial E_{\text{tot}}}{\partial N_A} = \mu_A - \mu_B.$$

(9.27)

Obviously the minimum energy state corresponds to

$$\mu_A = \mu_B.$$

(9.28)

For the physical interpretation of this result, section 9.5.8, it is important to note that the chamber entropy is unchanged. There is currently no principle of energy minimization in thermodynamics or statistical mechanics. In mechanics, the force points toward the potential energy minimum (Newton 1687), but even in this case the total energy is constant on a trajectory. Further, mechanical laws have no direct isolated application to thermodynamic or statistical systems. Although one could obtain equality of chemical potential by minimizing the simple sum of the free energies of the two chambers, as discussed above such procedure is not correct as it is actually the free energy divided by temperature that is additive.

The chemical potential is the Gibbs free energy per particle, $\mu = G(N, p, T)/N$, equation (2.53). The derivative of equation (9.28) with respect to T_B at constant pressure and temperature of the first chamber, and number of the second, is (section 2.4.1)

$$0 = \frac{\mathrm{d}(G_B/N_B)}{\mathrm{d}T_B}$$

$$= \frac{\partial g_B}{\partial T_B} + \frac{\partial g_B}{\partial p_B}\frac{\mathrm{d}p_B}{\mathrm{d}T_B}$$

$$= -s_B + v_B\frac{\mathrm{d}p_B}{\mathrm{d}T_B},$$

(9.29)

where g, s, and $v = \rho^{-1}$ are the Gibbs free energy, entropy, and volume per particle, respectively. This is the same as H London's expression, equation (9.21). (Since $T_B = T_A$ implies $\mu_B = \mu_A$, the constant of integration must be zero.) Therefore, equality for chemical potential, equation (9.28) is thermodynamically equivalent to the H London's expression for the fountain pressure, equation (9.21).

9.5.3 Relationship between the two principles

One might imagine it a simple matter to distinguish the two principles invoked above by appealing to experimental measurement. Unfortunately, it is not so easy.

Differentiating the condition of constant fugacity, equation (9.26), with respect to temperature on the fountain path gives

$$0 = z_B\beta_B\frac{\mathrm{d}\mu_B}{\mathrm{d}T_B} - z_B\beta_B\mu_B\frac{1}{T_B}.$$

(9.30)

On the saturation curve the chemical potential divided by temperature must be small and negative because the liquid is in equilibrium with a gas. In fact, using measured values for the enthalpy and the entropy for ^{4}He on the saturation curve (Donnelly and Barenghi 1998), the value at $T = 1$ K is $\beta\mu^{\mathrm{sat}} = -1.26 \times 10^{-3}$, and at $T = 2.15$ K it is $\beta\mu^{\mathrm{sat}} = -1.05 \times 10^{-1}$. The fugacity for bosons must be bounded above by unity, $z < 1$, otherwise the denominator of the momentum state distribution would pass through zero. (For the case of ideal bosons, F London (1938) showed that $z^{\mathrm{id}} \to 1^-$ below the λ-transition (see section 8.2).) Since the compressibility is positive, and since the fountain pressure is greater than the saturation pressure at the same temperature, on a fountain path one must have $\mu_B^{\mathrm{sat}} \leqslant \mu_B < 0$. Hence on a fountain path

$$-1 \ll \beta_B\mu_B < 0.$$

(9.31)

This result may be confirmed using measured fountain pressures (Hammel and Keller 1961) and saturation data (Donnelly and Barenghi 1998).

This means that one can neglect the term proportional to $\beta\mu$ in the above derivative, which then gives

$$\frac{\mathrm{d}\mu_B}{\mathrm{d}T_B} = \mathcal{O}(10^{-3})k_{\mathrm{B}}.$$

This says that for practical purposes the condition of constant fugacity is equivalent (9.3 to the condition of constant chemical potential to within about one part in one 2) thousand. Closer to the λ-transition the difference is about one part in ten. One can see from this that there is some question whether current measurements have the accuracy to distinguish the two principles.

Is there a theoretical reason to prefer one of the principles over the other?

In the present case it may be said that the condition of chemical potential equality is a rigorous mathematical consequence of energy minimization at constant entropy. But not all mathematical results have physical relevance. Given the fact that energy minimization plays no role in the usual thermodynamic systems, if it is indeed the underlying principle for superfluid flow then a physical explanation is called for (section 9.5.8).

Equally, it must be conceded that there is no reason to believe that maximizing the entropy is a principle relevant to the present non-equilibrium steady-state system (chapter 3).

The tests against measured data (section 9.5.6) and the arguments that can be made (section 9.5.8) favor the principle of energy minimization. Perhaps in time it will become axiomatic that it is the principle that drives superfluid flow.

9.5.4 Inconsistency in the derivation of the H London expression

An essential step in H London's (1939) derivation of his expression for the fountain pressure is where he invokes the Nernst heat theorem to conclude that the enthalpy divided by temperature and the entropy both go to zero at absolute zero. These mean that the chemical potential must go to zero at absolute zero. H London (1939) also assumes that there exists a continuous fountain path to absolute zero from an arbitrary thermodynamic point.

As shown above, the H London (1939) expression is thermodynamically equivalent to the chemical potential being constant on the fountain path, equation (9.28). Hence the assumptions made by H London (1939) in his derivation imply that the chemical potential is zero at an arbitrary thermodynamic point. This is in general nonsense, and it is in particular wrong for saturated ^{4}He. It follows, therefore, that if the H London (1939) expression is exact, then either the Nernst heat theorem is invalid, or else that there does not exist a fountain path to absolute zero, or both. In any case, the failure of either of these assumptions renders the derivation of H London (1939) invalid.

It should be mentioned that the result of entropy maximization, namely the constancy of fugacity, equation (9.25), when combined with these two assumptions, would also yield nonsense, namely that the fugacity is always unity. However, in this case, the failure of either or both of these assumptions has no effect on the derivation of the result by entropy maximization.

It is not my intention to assess the Nernst heat theorem in detail. But I cannot help but observe that if H London (1939) is correct in that it implies that the chemical potential divided by temperature vanishes at absolute zero, then it would imply that the fugacity equalled unity at absolute zero. But the fugacity is bound to be strictly less than unity, at least for a system with single-particle energy states (section 8.2). Contrariwise, the number of bosons would be infinite. In so far as ^{4}He appears to be dominated by ideal statistics deep below the λ-transition (section 9.3), a finite-sized system must arguably violate the Nernst heat theorem.

Similarly, I don't wish to traverse step-by-step a fountain path to absolute zero, but I do note that situating the high temperature chamber on the saturation curve would require a negative pressure in the other chamber at absolute zero. (The saturated vapor pressure is relatively low, but as the pressure of the high temperature chamber, it must be much greater than the pressure in the chamber at absolute zero.) Conversely, if one insists upon a stable thermodynamic state at absolute zero, then this places a lower bound on the pressure in the high temperature chamber that would exceed the saturation pressure. In other words, not all thermodynamic state points lie on a fountain path to a stable point at absolute zero.

9.5.5 Proof that the energy is a minimum

Using analogous analysis to that of section 2.4.4, define the number-constrained total entropy for a fixed subsystem entropy system

$$S_{\text{tot}}(N \mid \mu, S, V; T) = S - \frac{1}{T} E(S, V, N) + \frac{\mu}{T} N.$$

(9.33)

Here T is a temperature parameter. The associated constrained free energy is

$$A^*(N \mid \mu, S, V; T) = E(S, V, N) - \mu N - TS.$$

(9.34)

Since we are mainly interested in the variation with number we can discard the final constant and write

$$A(N \mid \mu, S, V) = E(S, V, N) - \mu N.$$

(9.35)

This describes number fluctuations in an isentropic subsystem. It has number derivative

$$\frac{\partial A(N \mid \mu, S, V)}{\partial N} = \mu(S, V, N) - \mu,$$

(9.36)

which shows that the constrained free energy is extremised at the equilibrium number $\bar{N}(S, V, \mu)$ such that the chemical potentials are equal, $\mu(S, V, \bar{N}) = \mu$.

As usual, with $\bar{A}(\mu, S, V) = A(\bar{N} \mid \mu, S, V)$, one has

$$\mathrm{d}\bar{A} = \bar{T} \, \mathrm{d}S - \bar{p} \, \mathrm{d}V - \bar{N} \, \mathrm{d}\mu.$$

(9.37)

Derived from the total entropy, the free energy must be a minimum,

$$\frac{\partial^2 A(N \mid \mu, S, V)}{\partial N^2} = \frac{\partial^2 E(S, V, N)}{\partial N^2} = \frac{\partial \mu(S, V, N)}{\partial N} > 0.$$

(9.38)

We conclude from this that if the H London (1939) expression is exact then the fountain pressure is determined by the minimum in the subsystem energy with respect to number at constant subsystem entropy.

9.5.6 Comparison with experiment

9.5.6.1 Expressions for the fountain pressure

In practice in experimental application (Hammel and Keller 1961) the H London (1939) expression for the derivative of the fountain pressure, equation (9.21), is integrated along the saturation curve,

$$p_B - p_A = \int_{T_A}^{T_B} dT\prime\, \rho^{\mathrm{sat}}(T\prime) s^{\mathrm{sat}}(T\prime).$$

(9.39)

Since the integral for the fountain pressure should be evaluated on the fountain path rather than the saturation path, one can correct this as follows. One has, with pressure and temperature as the independent variables,

$$\begin{aligned}
\rho s &= \rho^{\mathrm{sat}} s^{\mathrm{sat}} + (p - p^{\mathrm{sat}})\left(\frac{\partial(\rho s)}{\partial p}\right)_T^{\mathrm{sat}} \\
&\approx \rho^{\mathrm{sat}} s^{\mathrm{sat}} + (p - p^{\mathrm{sat}})\rho^{\mathrm{sat}}\left(\frac{1}{T}\frac{\partial h}{\partial p} - \frac{v}{T}\right)_T^{\mathrm{sat}} \\
&= \rho^{\mathrm{sat}} s^{\mathrm{sat}} - \alpha^{\mathrm{sat}}(p - p^{\mathrm{sat}}).
\end{aligned}$$

(9.40)

In the second equality the liquid has been taken to be incompressible. The thermal expansivity is given by equation (2.118), $\alpha = -\rho^{-1}\partial\rho(p, T)/\partial T$. The final equality follows from the cross-derivative of the Gibbs free energy (cf section 2.4.5)

$$\begin{aligned}
\frac{\partial h}{\partial p} &= \frac{\partial^2(\beta G(N, p, T)/N)}{\partial p \partial \beta} \\
&= \frac{\partial(\beta v)}{\partial \beta} \\
&= v - \alpha v T.
\end{aligned}$$

(9.41)

Here and throughout $\beta = 1/k_{\mathrm{B}}T$, with k_{B} being Boltzmann's constant. With this the fountain pressure integral can be translated from the fountain path to the saturation path,

$$p_B - p_A = \oint_{T_A}^{T_B} \mathrm{d}T\prime\, \rho(T\prime)s(T\prime)$$

$$\approx \int_{T_A}^{T_B} \mathrm{d}T\prime\, \{\rho^{\mathrm{sat}}(T\prime)s^{\mathrm{sat}}(T\prime) - \alpha^{\mathrm{sat}}(T\prime)[p(T\prime; p_A, T_A) - p^{\mathrm{sat}}(T\prime)]\}.$$

(9.42)

The fountain pressure in the second term in the integrand may be approximated by that given by the uncorrected expression, equation (9.39), rather than iterated to self-consistency. Numerically using measured data this form is indistinguishable from the raw H London (1939) form on the saturation path, equation (9.39).

A second equation for the fountain pressure comes from the equality of chemical potential, equation (9.28), which is thermodynamically equivalent to the H London (1939) expression, equation (9.21). This can be used to obtain the fountain pressure by writing $\mu_B = \mu_B^{\mathrm{sat}} + (p_B - p_B^{\mathrm{sat}})v_B^{\mathrm{sat}}$, which holds for an incompressible liquid. Rearranging gives

$$p_B - p_A \approx p_B^{\mathrm{sat}} - p_A + \rho_B^{\mathrm{sat}}(\mu_B - \mu_B^{\mathrm{sat}})$$
$$= p_B^{\mathrm{sat}} - p_A + \rho_B^{\mathrm{sat}}(\mu_A - \mu_B^{\mathrm{sat}}).$$

(9.43)

Invariably the experimental measurements are performed at saturation of chamber A, and so all quantities on the right-hand side, including $\mu^{\mathrm{sat}} = h^{\mathrm{sat}} - Ts^{\mathrm{sat}}$, can be obtained from standard tables such as those given by Donnelly and Barenghi (1998). It is emphasized that any difference between the fountain pressure given by equation (9.39) and that given by equation (9.43) must be due to experimental error, and the difference between them gives a guide to the quantitative reliability of the measurements.

A third equation for the fountain pressure comes from the equality of chemical potential divided by temperature, equation (9.25). Again using the incompressible liquid expression for the departure of the chemical potential from its saturation value, this gives for the fountain pressure

$$p_B - p_A = p_B^{\mathrm{sat}} - p_A + \rho_B^{\mathrm{sat}}\left(T_B\mu_A/T_A - \mu_B^{\mathrm{sat}}\right).$$

(9.44)

9.5.6.2 An unfortunate error in the reported entropy and enthalpy

In order to test these expressions against the measured fountain pressure it is necessary to obtain measured values for the enthalpy per particle $h = H/N$, the entropy per particle $s = S/N$, and the chemical potential $\mu = h - Ts$. Unfortunately, the fitted values given by Donnelly and Barenghi (1998) are in error. This can be seen in figure 9.3, which compares the three expressions for the fountain pressure with the measured values of Hammel and Keller (1961). This figure is essentially the same as that given in Attard (2022b). Only the integral form of the H London (1939) expression, equation (9.39), agrees with the measured values. What is somewhat concerning is that the results for constant chemical potential, equation (9.43), do not agree with the original H London (1939) expression, equation (9.39), even though these are thermodynamically equivalent.

If the discrepancy were due to random experimental error in either the thermodynamic data or the fountain pressure data, then there should be a similar disagreement between the symbols and the short dashed curve in the figure. The figure confirms that there is an error in the entropy and enthalpy data given by Donnelly and Barenghi (1998) that was used in the expressions. Correct expressions for these are now derived, and the recalculated data is used below.
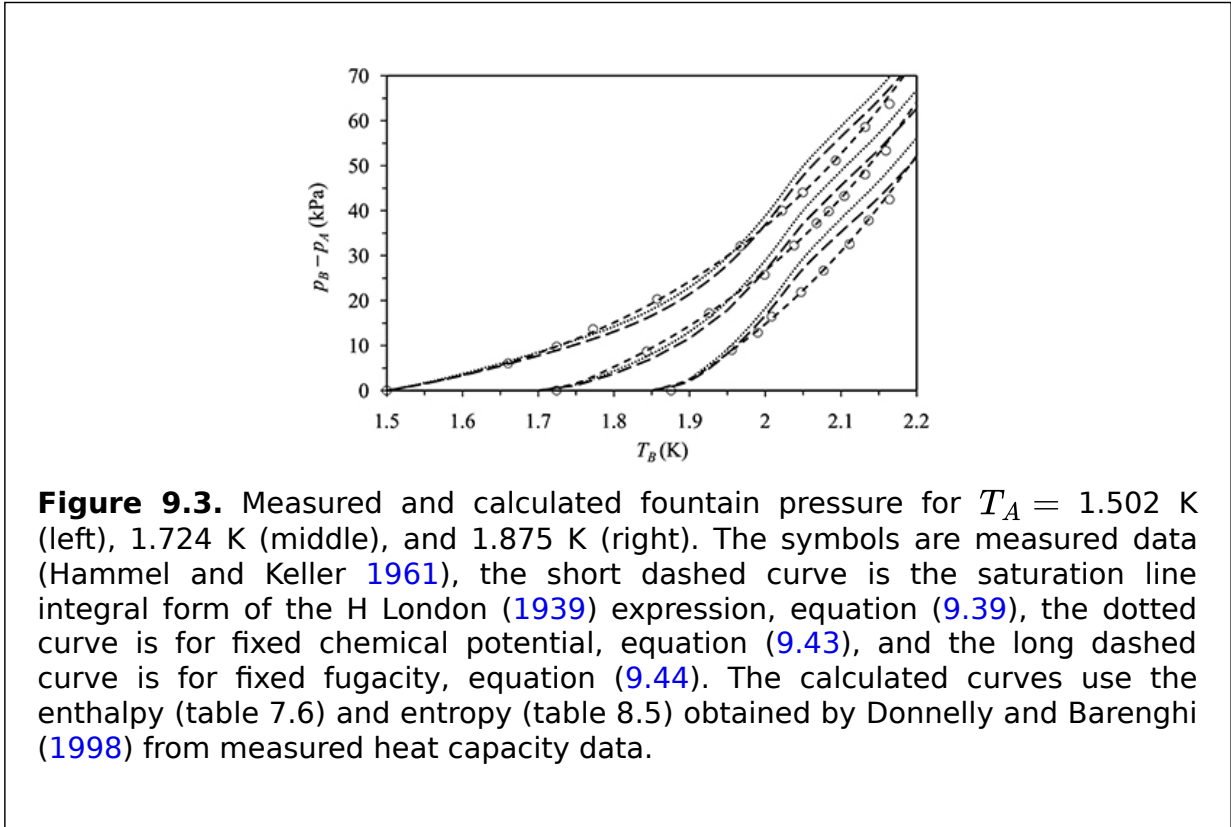

Figure 9.3. Measured and calculated fountain pressure for $T_A = 1.502$ K (left), 1.724 K (middle), and 1.875 K (right). The symbols are measured data (Hammel and Keller 1961), the short dashed curve is the saturation line integral form of the H London (1939) expression, equation (9.39), the dotted curve is for fixed chemical potential, equation (9.43), and the long dashed curve is for fixed fugacity, equation (9.44). The calculated curves use the enthalpy (table 7.6) and entropy (table 8.5) obtained by Donnelly and Barenghi (1998) from measured heat capacity data.

The heat capacity at constant pressure, equation (2.111), is

$$C_{\mathrm{p}} = \frac{-1}{T^2} \frac{\partial^2 (\bar{G}(N,p,T)/T)}{\partial (1/T)^2}$$

$$= \frac{\partial \bar{H}(N,p,T)}{\partial T}$$

$$= T \frac{\partial \bar{S}(N,p,T)}{\partial T}.$$

(9.45)

The second equality can be seen by dividing both sides of the definition of the unconstrained Gibbs free energy, $G(E,V \mid N,p,T) = E + pV - TS(E,V,N)$, by T and differentiating with respect to $1/T$, holding as usual $\bar{E}$ and $\bar{V}$ fixed. The third equality follows by expressing the first equality as a derivative with respect to T, equation (2.112).

In practice measurements are made along the saturation curve, $p^{\text{sat}}(T)$. The change in enthalpy is measured, $\Delta H = H(N, p^{\text{sat}}(T_2), T_2) - H(N, p^{\text{sat}}(T_1), T_1)$. At constant number,

$$
\begin{aligned}
C_{\text{sat}} &\equiv \left(\frac{\mathrm{d}\bar{H}(N, p, T)}{\mathrm{d}T} \right)_N \\
&= \frac{\partial \bar{H}(N, p, T)}{\partial T} + \frac{\partial \bar{H}(N, p, T)}{\partial p} \frac{\mathrm{d}p^{\text{sat}}(T)}{\mathrm{d}T} \\
&= C_{\text{p}} + \bar{V} \frac{\mathrm{d}p^{\text{sat}}(T)}{\mathrm{d}T}.
\end{aligned}
$$

(9.46)

This heat capacity at constant saturation is larger than that at constant pressure. I believe this to be the quantity reported in table 7.4 of Donnelly and Barenghi (1998) and denoted C_{s} by them. This assertion is justified by the internal consistency of the numerical results presented below.

From this one sees that the difference in enthalpy on the saturation curve is

$$
H(N, p^{\text{sat}}, T) - H(N, p_0^{\text{sat}}, T_0) = \int_{T_0}^{T} \mathrm{d}T\prime \, C_{\text{sat}}(T\prime).
$$

(9.47)

This contradicts the expression given in note 11 to section 7 of Donnelly and Barenghi (1998), who appear to have inadvertently mixed up C_{p} and C_{sat}. The results they present for the enthalpy in their table 7.6 have a relative systematic error of $\mathcal{O}(10^{-2})$. Although this is comparable to the random measurement and fitting error, because it is a systematic error, and because the chemical potential is the difference between two comparable quantities, it leads to errors on the order of 5% in the fountain pressure (figure 9.3).

Now the temperature derivative of the entropy along the saturation curve at constant number is

$$
\begin{aligned}
\left(\frac{\mathrm{d}\bar{S}(N, p, T)}{\mathrm{d}T} \right)_N &= \frac{\partial \bar{S}(N, p, T)}{\partial T} + \frac{\partial \bar{S}(\bar{E}, \bar{V}, N)}{\partial p} \frac{\mathrm{d}p^{\text{sat}}(T)}{\mathrm{d}T} \\
&= \frac{1}{T} C_{\text{p}} - \frac{\partial \bar{V}(N, p, T)}{\partial T} \frac{\mathrm{d}p^{\text{sat}}(T)}{\mathrm{d}T} \\
&= \frac{1}{T} C_{\text{p}} - \alpha \bar{V}(N, p, T) \frac{\mathrm{d}p^{\text{sat}}(T)}{\mathrm{d}T} \\
&= \frac{1}{T} C_{\text{sat}} - N\rho^{-1} \left[\frac{1}{T} + \alpha \right] \frac{\mathrm{d}p^{\text{sat}}(T)}{\mathrm{d}T}.
\end{aligned}
$$

(9.48)

Accordingly the difference in entropy on the saturation curve is

$$S(N, p^{\text{sat}}(T), T) - S(N, p^{\text{sat}}(T_0), T_0)$$
$$= \int_{T_0}^{T} dT\prime \, \frac{1}{T\prime} \left\{ C_{\text{sat}}(T\prime) - \frac{N}{\rho\prime}[1 + \alpha\prime T\prime] \frac{dp^{\text{sat}}(T\prime)}{dT\prime} \right\}.$$

(9.49)

This contradicts the expression given in note 8 to section 11 of Donnelly and Barenghi (1998), which neglects the second term in the braces. Compared to the present expression, the results for the calorimetric entropy in their table 8.5 have a relative systematic error of $\mathcal{O}(10^{-2})$. It turns out that those erroneous values of the entropy give good results for the fountain pressure when used in the integral form of the H London (1939) expression, equation (9.39), but not when used in the chemical potential equality form (figure 9.3). Neglecting α, and using either the form with the correct $C_{\text{sat}}(T\prime)$ or else the one with correct $C_{\text{p}}(T\prime)$, gives a relative systematic error of $\mathcal{O}(10^{-4})$, which is reasonable since the thermal expansivity is $\mathcal{O}(10^{-3})$.

9.5.6.3 Measured and calculated fountain pressure

Figure 9.4 tests the various equations for the fountain pressure against the measured values (Hammel and Keller 1961). The calculations use the measured data for the heat capacity at constant saturation, C_{sat} (Donnelly and Barenghi 1998 table 7.4), from which the enthalpy, equation (9.47), and the entropy, equation (9.49), and the chemical potential, $\mu = h - Ts$, are obtained. This entropy was also used for the integrated H London (1939) expression, equation (9.39).
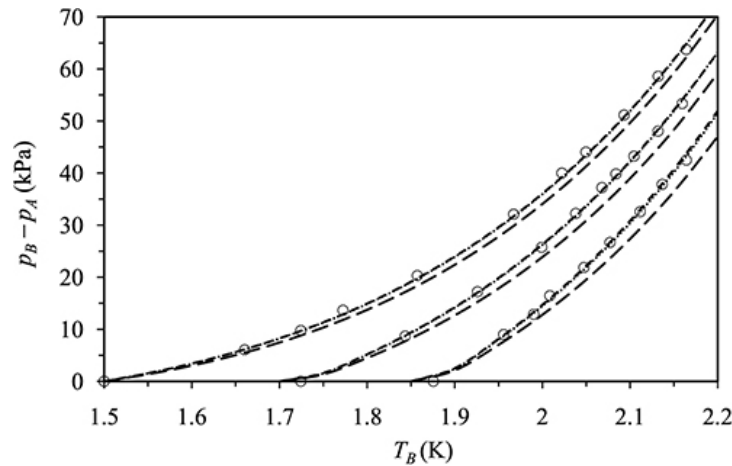

Figure 9.4. Measured and calculated fountain pressure for $T_A = 1.502$ K (left), 1.724 K (middle), and 1.875 K (right). The symbols are measured data (Hammel and Keller 1961), the short dashed curve is the saturation line integral form of the H London (1939) expression, equation (9.39), the coincident dotted curve is for fixed chemical potential, equation (9.43), and the long dashed curve is for fixed fugacity, equation (9.44), both using the incompressible fluid estimate for the departure from the saturation value. The

calculated curves use the measured saturation heat capacity, C_{sat} (Donnelly and Barenghi 1998 table 7.4), to obtain the enthalpy, equation (9.47), and the entropy, equation (9.49).

It can be seen the fountain pressure predicted by the H London expression evaluated as an integral along the fountain curve, Equation (9.39), and that evaluated by equal chemical potential, equation (9.43), are virtually indistinguishable. This confirms the thermodynamic equivalence of the two, the validity of the thermodynamic analysis that corrects the results of Donnelly and Barenghi (1998), and the reliability of the experimental data when analyzed correctly.

It can be seen that the two forms of the H London expression are in quite good agreement with the measured values of the fountain pressure (Hammel and Keller 1961). Adding the correction for the translation from the fountain path to the saturation path, equation (9.42), makes a difference of about −0.5% at the highest fountain pressure shown, which would be indistinguishable from the uncorrected result, equation (9.39), on the scale of the figure.

Hammel and Keller (1961) estimated the error in their fountain pressure measurements as on the order of 2%. Hammel and Keller (1961) compared their measurements of the fountain pressure with that predicted using the integrated form of the H London (1939) equation (9.39) together with values of the entropy measured calorimetrically (Kramers *et al* 1951, Hill and Lounasmaa 1957) and found agreement within the estimated measurement error.

Donnelly and Barenghi (1998 table 7.4), for the heat capacity at what I believe is constant saturation, give the measurement error as 1%–3% and the fitting error as 1%–2%.

The values of the fountain pressure predicted by equality of fugacity, equation (9.44), lie systematically below the measured values. The difference is on the order of 3%–5%, which appears significant compared to the various measurement errors. On the basis of the results in figure 9.4, one can tentatively conclude that the measured data favor the principle of energy minimization at constant entropy for superfluid flow, and that they likely rule out the principle of entropy maximization.

Of course experimental data is no substitute for mathematical derivation for proving a result is exact. But experimental data has historically been used to formulate general scientific principles that can then be used axiomatically to derive exact and approximate expressions to describe that and other data. It is usually the case that such principles gain acceptance over time when no contradictory evidence emerges, when they explain a range of physical phenomena, and when scientists become familiar with them. A rationale for the principle of energy minimization at constant entropy for superfluid flow is offered in section 9.5.8.

9.5.7 Convective flow

In the fountain effect with closed chambers a non-equilibrium steady state exists with viscous flow of He I from the high to the low temperature chamber and superfluid flow of He II in the other direction to maintain mass balance. These flows occur simultaneously in the same capillary. There is also net energy flow from the high temperature chamber to the low. The flow of He I is viscous Poiseuille flow

driven by the pressure difference (but see next) and it carries the energy convectively (F London and Zilsel 1948, Keller and Hammel 1960). The superfluid flow of He II arriving in the high temperature chamber is in total equal and opposite to the viscous flow of He I leaving it. A similar balance but in the opposite direction occurs for the low temperature chamber. In the steady state the total number of ^{4}He atoms in each of the two closed chambers is conserved. There is of course a net energy flux between the two chambers, with energy supplied by a heater in one and removed by a refrigerator in the other.

In normal convective flow the two species are hot and cold particles and their spatially distinct flows are driven by respective entropy gradients. One would expect similar gradients in the present fountain system. The evidence is that $\mu_A = \mu_B < 0$. Since $T_B > T_A$, this means that $(-\mu_A/T_A) > (-\mu_B/T_B)$, which means that there is an entropy gradient that drives number from B to A. (Recall $\partial S(E, V, N)/\partial N = -\mu/T$.) This is what really drives the viscous Poiseuille flow of He I.

But what drives the steady flow of condensed ^{4}He from A to B? Recall that when $\mu_A = \mu_B$ the energy is minimized and there is no driving force. One concludes that there must strictly be a gradient with $\mu_B < \mu_A$, and, to linear order, $J_0 = c_2(\mu_A - \mu_B)$. (It is possible but unusual that the right-hand side could instead be non-analytic in the difference.) In this case, then strictly speaking the measured fountain pressure should lie between that predicted by constant chemical potential and that predicted by constant fugacity. The fact that the measured fountain pressure lies so close to the prediction from $\mu_A = \mu_B$ indicates that superfluid flow is extremely efficient at eliminating gradients, $c_2 \gg \beta_A J_*$. It seems likely that the thinner the capillary, and the lower the temperature difference, the closer to equality would be the chemical potentials (because the Poiseuille flow is reduced, and a smaller balancing superfluid flow requires a smaller energy gradient). It would appear that one needs a wide slit and better than the 2% precision of current measurements to confirm or refute the hypothesis $\mu_B < \mu_A$ (equivalently, $p_B^{\text{meas}} < p_B^{\text{HLondon}}$).

In the experiments of Keller and Hammel (1960), the mean velocity of the viscous flow of He I in the case of the greatest fountain pressure is on the order of 60 times the critical velocity for superfluid flow predicted by the momentum gap for the slit (section 9.4.3). Assuming a comparable speed for the condensed bosons, this suggests that collisions are strong enough to convert a proportion of the back flow of superfluid He II to viscous He I. But in the case of the fountain effect, this does not block the capillary because the fountain pressure is so large that substantial Poiseuille flow continues.

9.5.8 Rationale for the principle of energy minimization

The H London (1939) expression for the fountain pressure agrees with the measured values within experimental error (figure 9.4). H London's (1939) expression corresponds to chemical potential equality of the two chambers (section 9.5.2). Equal chemical potential minimizes the subsystem energy at constant subsystem entropy. (In the pseudo-equilibrium thermodynamic treatment of the fountain effect, each chamber is a subsystem in contact with its own thermal

reservoir.) Two questions are addressed here: how does the fountain pressure arise, and what is the justification for energy minimization?

Equality of chemical potential offers a thermodynamic mechanism for the fountain pressure difference. Imagine that initially the high temperature chamber is at saturation, $\mu_B^{\text{ini}} = \mu^{\text{sat}}(T_B)$. The low temperature chamber is always at saturation, $\mu_A = \mu^{\text{sat}}(T_A)$. For ^{4}He, on the saturation curve the chemical potential from the measured enthalpy and entropy (Donnelly and Barenghi 1998) decreases with increasing temperature, $\mathrm{d}\mu^{\text{sat}}/\mathrm{d}T < 0$. Hence $\mu_B^{\text{sat}} < \mu_A^{\text{sat}}$. Since $\partial\mu/\partial p = v > 0$, the only way that the chemical potential of the high temperature chamber can be increased to achieve equality is by increasing the pressure beyond its saturation value. This occurs as more ^{4}He arrives in the second chamber because each atom occupies a certain impenetrable volume (i.e. the compressibility is positive). Given the goal of equalizing the chemical potentials, these are the reasons why condensed ^{4}He initially flows down the chemical potential gradient from the low temperature chamber to the high temperature chamber, and why the high temperature chamber subsequently settles at a higher pressure.

The physical basis for the principle of energy minimization at constant entropy is as follows. Superfluid flow is dissipationless flow that preserves the momentum state of each condensed boson involved and hence the permutation entropy of that state (section 9.4). To the extent that permutation entropy dominates, we may say that subsystem entropy is preserved in the superfluid transfer of a condensed boson from $A \Rightarrow B$. (The more highly occupied is a momentum state, the greater is its permutation entropy, and the more likely it contributes to superfluid flow.) Recall that each chamber is regarded as a subsystem in contact with its own thermal reservoir, and that the total energy of the two subsystems and the two reservoirs is conserved. Hence minimization of the total subsystems' energy, $E(S_A, V_A, N_A) + E(S_B, V_B, N_B)$, by number exchange at constant subsystem entropy, increases the total reservoirs' energy and hence the reservoirs' entropy. Since the subsystems' entropy is constant, this increases the entropy of the Universe. Hence the principle of subsystem energy minimization at constant subsystem entropy is just the Second Law of Thermodynamics realized by superfluid flow.

9.5.9 Osmotic pressure

A statistical mechanical analysis of the fountain pressure is now performed with a view to confirming or refuting the osmotic pressure explanation of Tisza (1938). The binary division approximation is used. I begin with the incompressible classical ideal solution in order to identify exactly what is meant by the osmotic pressure explanation.

9.5.9.1 Incompressible classical ideal solution

Consider two chambers, *A* and *B*, connected by a semi-permeable membrane transparent to the solvent 1. Assume that the solvent and solute are incompressible, so that the volume of the first chamber is $V_A = N_{A1}v_1 + N_{A2}v_2$, and similarly for the second chamber. The total number is fixed, $N_n = N_{An} + N_{Bn}$, $n = 1, 2$. Assume that the two chambers are at the same temperature. For the classical ideal gas, the constrained Gibbs free energy is, section 2.4.1

$$\beta G(N_{A1} \mid \underline{N}, p_A, p_B, T) = \sum_{n=1}^{2} \left\{ N_{An} \ln \frac{N_{An}\Lambda^3}{N_{A1}v_1 + N_{A2}v_2} - N_{An} \right.$$

$$\left. + N_{Bn} \ln \frac{N_{Bn}\Lambda^3}{N_{B1}v_1 + N_{B2}v_2} - N_{Bn} \right\}$$

$$+ \beta p_A[N_{A1}v_1 + N_{A2}v_2] + \beta p_B[N_{B1}v_1 + N_{B2}v_2].$$

(9.50)

Here $\beta = 1/k_{\mathrm{B}}T$ is the reciprocal temperature, $\Lambda = \sqrt{2\pi\hbar^2\beta/m}$ is the thermal wavelength, m is the mass, and μ is the chemical potential. Since $G = \mu N$, this shows that for an incompressible fluid, $\mu^{\mathrm{ex}} = p^{\mathrm{ex}}v$. The derivative is

$$\left(\frac{\partial \beta G}{\partial N_{A1}} \right)_{N_1} = \ln \frac{N_{A1}\Lambda^3}{N_{A1}v_1 + N_{A2}v_2} - \ln \frac{N_{B1}\Lambda^3}{N_{B1}v_1 + N_{B2}v_2}$$

$$- \frac{N_{A1}v_1}{N_{A1}v_1 + N_{A2}v_2} + \frac{N_{B1}v_1}{N_{B1}v_1 + N_{B2}v_2}$$

$$- \frac{N_{A2}v_1}{N_{A1}v_1 + N_{A2}v_2} + \frac{N_{B2}v_1}{N_{B1}v_1 + N_{B2}v_2}$$

$$+ \beta p_A v_1 - p_B \beta v_1$$

$$= \ln \frac{\rho_{A1}}{\rho_{B1}} - [\rho_{A1} + \rho_{A2} - \rho_{B1} - \rho_{B2}]v_1 + \beta[p_A - p_B]v_1.$$

(9.51)

This vanishes at equilibrium, which is therefore given by

$$\ln \frac{\bar{\rho}_{A1}}{\bar{\rho}_{B1}} - [\bar{\rho}_{A1} + \bar{\rho}_{A2} - \bar{\rho}_{B1} - \bar{\rho}_{B2}]v_1 v_1 = -\beta[p_A - p_B]v_1.$$

(9.52)

For simplicity choose $v_1 = v_2 = v$, so that $\rho = \rho_1 + \rho_2 = v^{-1}$ and $\rho_A = \rho_B$. In this case the equilibrium condition becomes

$$\ln \frac{\bar{\rho}_{A1}}{\bar{\rho}_{B1}} = \frac{-\beta}{\rho}[p_A - p_B].$$

(9.53)

Hence $p_A \geqslant p_B$ implies that $\bar{\rho}_{A1} \leqslant \bar{\rho}_{B1}$. The chamber with the greater pressure is the one with the lower solvent fraction (and hence higher solute fraction). This is the classical osmotic pressure.

9.5.9.2 Incompressible ideal bosons

Consider a system consisting of two chambers at different temperatures below the condensation temperature and containing ^{4}He. We give the free energy for a single chamber, and differentiate it to obtain an expression for the chemical potential. This allows us to equate chemical potentials, as found in section 9.5.2, and hence to obtain the fountain pressure.

The constrained Helmholtz free energy for a single system with N_0 ground momentum state bosons and N_* excited momentum state bosons is given by (chapter 8)

$$-\beta F(N_0 \mid N, V, T) = \ln \frac{\Lambda^{-3N_*} Q}{N_*! V^{N_0}} + \sum_{l=2}^{\infty} \frac{N_*^l}{N^{l-1}} g^{(l)}.$$

(9.54)

Here $N = N_0 + N_*$, $\beta \equiv 1/k_B T$, and $\Lambda \equiv \sqrt{2\pi\beta\hbar^2/m}$ is the thermal wavelength. The series contains the loop grand potentials, which is a quantum effect that arises from the symmetrization of the wave function. The logarithmic term is the classical or monomer term, and it contains the configuration integral for interacting bosons, $Q(N, V, T)$. Compared to the classical statistical mechanics of a binary mixture (see section 9.5.9.1), for the condensed boson case there is no $N_0!$ in the denominator, and instead of Λ^{3N_0} there appears V^{N_0}. These differences are a consequence of the non-locality of the permutations of ground momentum state bosons (cf equation (8.33)).

We make the incompressible liquid approximation, $V = Nv$. The constrained Gibbs free energy, $G = F + pV$, for chamber A is then given by

$$G(N_{0A} \mid N_A, p_A, T_A) = \beta_A^{-1} \left[N_{*A} \ln \frac{N_{*A} \Lambda_A^3}{N_A v_A} - N_{*A} - \ln \frac{Q_A}{V^{N_A}} \right]$$
$$- \sum_{l=2}^{\infty} \frac{\beta_A^{-1} N_{*A}^l}{N_A^{l-1}} g_A^{(l)} + p_A N_A v_A.$$

(9.55)

Recall that $N_A = N_{*A} + N_{0A}$.

Now for an incompressible fluid, since $\mu^{\text{ex}} = p^{\text{ex}} v$ (see section 9.5.9.1), the derivative of the configuration integral term vanishes,

$$\frac{\partial}{\partial N} \ln \frac{Q(N, V, T)}{V^N} = \frac{\partial(-\beta F^{\text{ex}}(N, V, T))}{\partial N} + v \frac{\partial(-\beta F^{\text{ex}}(N, V, T))}{\partial V}$$
$$= -\beta\mu^{\text{ex}} + \beta p^{\text{ex}} v = 0.$$

(9.56)

This result effectively removes the interactions between the ^{4}He atoms, and makes the analysis more or less the same as that for an ideal solution, apart from the way in which the fraction of ground and excited momentum state bosons is calculated or measured. Consequently, inserting the incompressible fluid assumption at this stage of the analysis is a rather serious approximation. It would be better to calculate the chemical potential for interacting atoms, which is quite feasible.

The derivative of the constrained Gibbs free energy gives the chemical potential, which with the vanishing of the derivative of the configurational integral, is

$$\mu_{0A} \equiv \left(\frac{\partial G_A}{\partial N_{0A}} \right)_{N_{*A}, p_A, T_A}$$

$$= \frac{-\beta_A^{-1} N_{A*} v}{V_A} + p_A v_A + \beta_A^{-1} \sum_{l=2}^{\infty} (l-1) \frac{N_{*A}^l g_A^{(l)}}{N_A^l}$$

$$= \frac{-f_{*A}}{\beta_A} + p_A \rho_A^{-1} + \beta_A^{-1} \sum_{l=2}^{\infty} (l-1)(f_{*A})^l g_A^{(l)}.$$

Here $f_{*A} = N_{*A}/N_A$ is the fraction of bosons in excited states in chamber A. One can similarly obtain an expression for the excited state chemical potential,

$$\mu_{*A} = \frac{\partial G_A}{\partial N_{*A}}$$

$$= \beta_A^{-1} \ln \frac{N_{*A} \Lambda_A^3}{V_A} - \beta_A^{-1} \frac{N_{*A} v}{V_A} + p_A v_A$$

$$+ \beta_A^{-1} \sum_{l=2}^{\infty} g_A^{(l)} \left\{ (l-1) \frac{N_{*A}^l}{N_A^l} - l \frac{N_{*A}^{l-1}}{N_A^{l-1}} \right\}.$$

In the equilibrium state, $\mu_{0A} = \mu_{*A}$, in which case one must have

$$\ln[f_{*A} \rho_A \Lambda_A^3] = \sum_{l=2}^{\infty} l g_A^{(l)} f_{*A}^{l-1}.$$

Since the right-hand side is a series of positive terms, one must have that $f_{*A} > 1/\rho_A \Lambda_A^3$, which places a lower bound on the fraction of excited state bosons (He I) as $T \to 0$. This result is predicated on the incompressible liquid, ideal solution approximation, the no mixing approximation, and on the binary division approximation.

Similar expressions hold for chamber B. Hence equating the chemical potential of the two chambers, $\mu_{0A} = \mu_{0B}$, yields the fountain pressure as

$$p_A - p_B = \rho_A k_B T_A f_{*A} - \rho_B k_B T_B f_{*B}$$

$$- k_B \sum_{l=2}^{\infty} (l-1) \left\{ \rho_A T_A g_A^{(l)} f_{*A}^l - \rho_B T_B g_B^{(l)} f_{*B}^l \right\}.$$
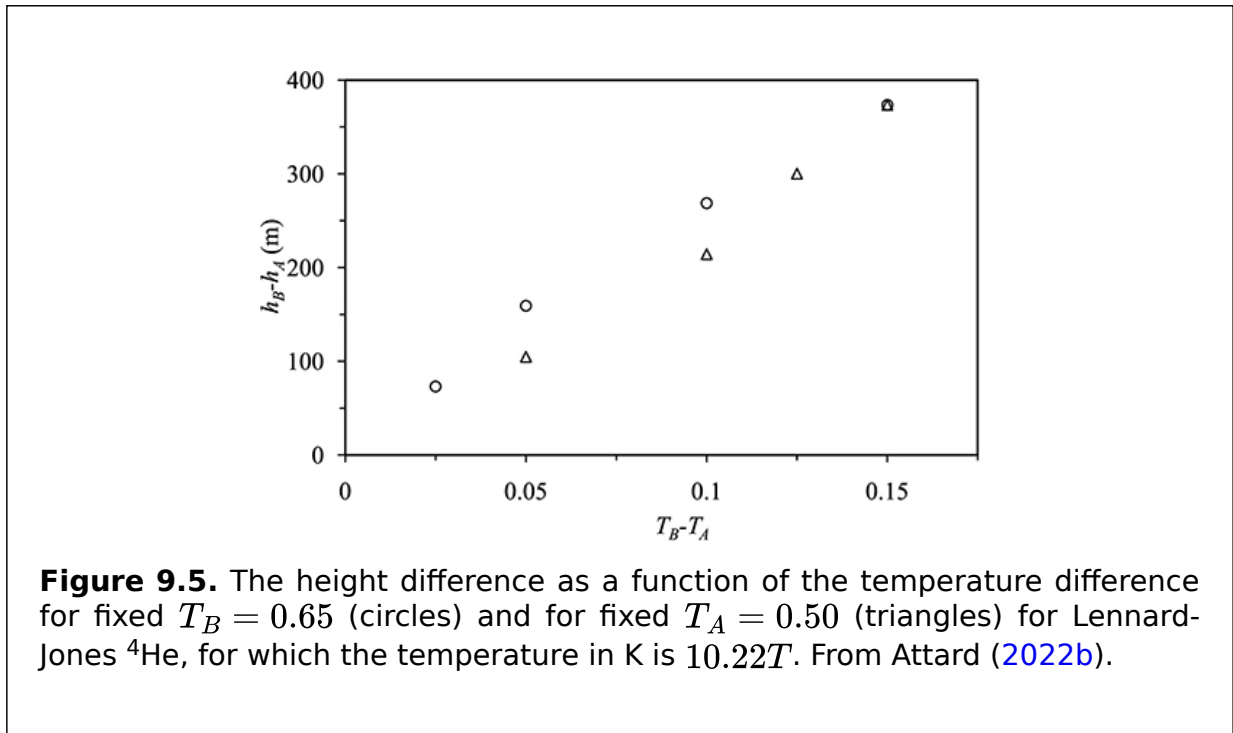
Ignoring the loop terms, and taking into account the fact that the excited momentum state fraction increases with increasing temperature, one sees that the higher temperature chamber has the higher pressure. Again neglecting the loop terms, notice the difference between this and the usual osmotic pressure equation (9.53). It is only in the linear regime that they are the same.

The left-hand side can be written as a height difference, $[p_A - p_B]\rho^{-1} \approx mg[h_A - h_B]$, where m is the molecular mass. (Note that here g is the acceleration due to gravity, not the specific Gibbs free energy, and h is the height, not the specific enthalpy.)

9.5.9.3 Numerical results

Extensive calculation of the intensive loop Gaussians $g^{(l)}$ up to $l = 7$ have been performed for a Lennard-Jones model of ^{4}He (see chapter 8, and also Attard (2021, 2022a)). The fraction of excited state bosons following condensation has been calculated in the so-called unmixed approximation (section 8.4.2). These have been used in the present expression for the pressure difference. The λ-transition in Lennard-Jones helium-4 occurs at $T = 0.65$ (section 8.5).

Figure 9.5 shows the calculated height difference as a function of the low temperature T_A for fixed high temperature $T_B = 0.65$, which is the λ- temperature in Lennard-Jones ^{4}He. It also shows the height difference for fixed low temperature $T_A = 0.50$, It can be seen that the height difference increases with increasing temperature difference. The magnitude of the fountain effect is surprisingly large considering that the largest temperature difference shown is about 1.5 K. The loop series contribution to the total height difference is on the order of 1%–5%. This means that the major part of the height differences in figure 9.5 come from the first line of the right-hand side of equation (9.60), which are essentially ideal solution terms.

Figure 9.5. The height difference as a function of the temperature difference for fixed $T_B = 0.65$ (circles) and for fixed $T_A = 0.50$ (triangles) for Lennard-Jones ^{4}He, for which the temperature in K is $10.22T$. From Attard (2022b).

It is noticeable in the figure that for the same temperature difference, $T_B - T_A = 0.05$, the higher temperature $T_B = 0.65$ has a height difference $h_B - h_A = 159.2$ m, which is about 50% larger than that at the lower temperature

$T_A = 0.50$, which has a height difference $h_B - h_A = 104.7$ m. This shows that there is a significant decrease in the ratio of height difference to temperature difference with decreasing temperature. This is because the fraction of excited state bosons decreases with decreasing temperature, and to leading order the height difference is proportional to the difference in the excited fraction.

The experimentally measured value at $T_A = 1.143$ K is 46.8 m for $T_B - T_A = 0.6$ K (Hammel and Keller 1961). The calculated height difference in figure 9.5 is 104.7 m for $T_B - T_A = 0.05 \approx 0.51$ K, which is about a factor of 2 too large. The main reason for the disagreement between the calculated and the measured height difference appears to be the use of the incompressible liquid, ideal solution approximation for the calculations.

Figure 9.6 compares the incompressible fluid osmotic pressure result, the first line of equation (9.60), with the measured fountain pressure. This uses the measured (Donnelly and Barenghi 1998) rather than calculated excited state fraction. Equating the measured fraction of He I to the excited momentum state fraction of the present analysis makes sense within the context of the binary division approximation. More generally it is not so clear what this measured quantity actually means. The calculations used for the data in figure 9.6 neglect the loop contributions, which as discussed for the previous figure are expected to be small. From the fact that the measured curves are nearly parallel, one can conclude that the fountain pressure is approximately a function of the temperature difference, $T_B - T_A$, rather than of T_B and T_A individually. In any case it can be seen that the incompressible, ideal solution result for the osmotic pressure performs quite badly.
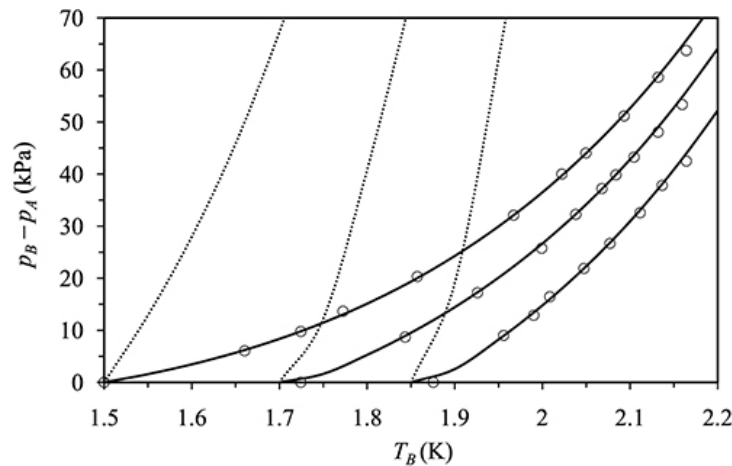

Figure 9.6. Measured and calculated fountain pressure for $T_A = 1.502$ K (left), 1.724 K (middle), and 1.875 K (right). The symbols are measured data (Hammel and Keller 1961), the full curve is the saturation line integral form of the H London (1939) expression, equation (9.39), and the dotted curve uses the incompressible fluid osmotic pressure result, the first line of equation (9.60), using the measured fraction of He I (Donnelly and Barenghi 1998). The calculated curves use measured saturation data (Donnelly and Barenghi 1998). From Attard (2022b).

Figure 9.7 shows the dimensionless parameter $f_* \rho \Lambda^3$ using measured data for the fraction of He I (Donnelly and Barenghi 1998). Again this only makes sense within the binary division approximation. As discussed following equation (9.59), equilibrium between ground and excited state bosons implies that this must be greater than unity, which the figure shows is violated for $T \leqslant 1.5$ K. The tests of the osmotic pressure prediction for the fountain pressure in figure 9.6 are for $T \geqslant 1.5$ K. Nevertheless that $f_* \rho \Lambda^3 < 1$ for $T \leqslant 1.5$ K is presumably an indication of the failure of the incompressible liquid, ideal solution, and no mixing approximations in the low temperature regime.
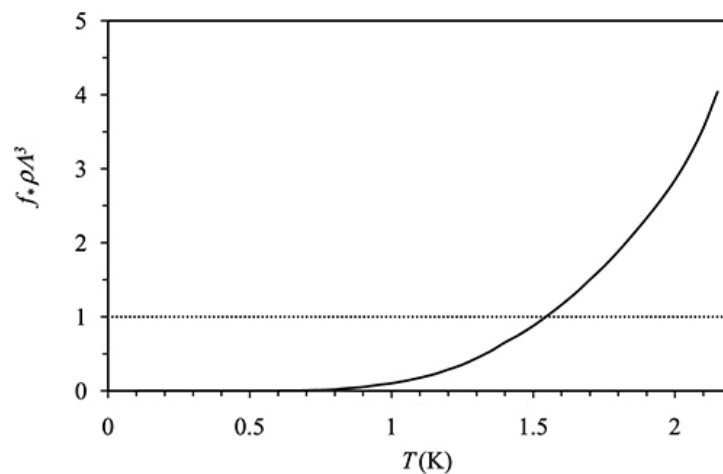

Figure 9.7. Measured (Donnelly and Barenghi 1998) fraction of He I times the measured (Donnelly and Barenghi 1998) number density and thermal wavelength cubed (solid curve). The dotted line is a guide to the eye. From Attard (2022b).

9.5.9.4 Discussion of the osmotic pressure mechanism

The osmotic pressure mechanism offered by Tisza (1938) as an explanation for the fountain pressure is problematic on physical grounds. In the usual realization of osmotic pressure for two chambers separated by a semi-permeable membrane, the solute and the solvent have distinct identities. Hence the only way to equalize the solvent chemical potential is to increase the pressure of the high concentration chamber. However, in the case of ^{4}He, the ground state bosons can become excited state bosons, and vice versa. Hence if the different fractions in each chamber somehow corresponded to a chemical potential difference due to mixing entropy, they could equalize chemical potential simply by changing their state without changing the pressure.

Also, and probably more damming, is the known experimental fact that He I passes through the capillary in the form of Poiseuille flow driven by the pressure

gradient from the high temperature chamber (F London and Zilsel 1948, Keller and Hammel 1960). In order to conserve number in this steady-state closed system, an equal and opposite superfluid flow of He II must occur *in the same capillary*. This is quite contrary to the usual osmotic pressure, which only arises because a semi-permeable membrane excludes at least one of the components of the mixture. In this non-equilibrium system, the occupied momentum states of the condensed bosons must be distributed about a peak in such a way that their flux exactly cancels the Poiseuille flow.

The present calculations show that the osmotic pressure mechanism can be an order of magnitude or more too large for the fountain pressure. Admittedly this is using the incompressible fluid, ideal solution approximation. But since the common understanding of osmotic pressure is that it is due to mixing entropy, and since the latter can be calculated exactly by ideal combinatorics, the failure of the present ideal solution calculations for the fountain pressure argue against osmotic pressure as the physical basis of the fountain effect.

Summary

- Condensation at the λ-transition signifies multiple, highly occupied, low-lying momentum states. The majority of low-lying momentum states are instantaneously empty or few-occupied. The permutation entropy makes it unlikely for collisions to occur that would change a boson's momentum from a highly occupied state into an unoccupied state. Superfluid flow is flow without momentum-changing collisions.
- The permutation entropy gives rise to a critical velocity equal to the spacing between transverse momentum states divided by the boson mass. The non-locality of momentum loop permutation entropy is essential for inviscid superfluid flow, as can be seen from its role in forming plug flow down a channel.
- Superfluid flow can be described by quantum adiabatic equations of motion in which the entropy is a constant of the motion. Collisions with uncondensed bosons do not change the momentum state of the condensed bosons, and hence the latter have zero shear viscosity.
- The measured fountain pressure due to a temperature difference is consistent with the H London (1939) expression, which is thermodynamically equivalent to equality of chemical potential, which minimizes the energy at constant entropy. The experimental data appears inconsistent with the principle of entropy maximization.
- The principle of energy minimization at constant subsystem entropy is a direct consequence of dissipationless superfluid flow: the momentum state of each condensed boson is preserved, and hence so is the permutation entropy of the state. Reducing the subsystems' energy increases the thermal reservoirs' energy and hence their entropy.

References

Ahlers G 1969 Critical heat flow in a thick He II film near the superfluid transition *J. Low Temp. Phys.* **1** 159

Allen J F and Misener A D 1938 Flow of liquid helium II *Nature* **141** 75

Allum D R, McClintock P V E and Phillips A 1977 The breakdown of superfluidity in liquid He: an experimental test of Landau's theory *Phil. Trans. R. Soc.* A **284** 179

Attard P 2002 *Thermodynamics and Statistical Mechanics: Equilibrium by Entropy Maximisation* (London: Academic)

Attard P 2012 *Non-Equilibrium Thermodynamics and Statistical Mechanics: Foundations and Applications* (Oxford: Oxford University Press)

Attard P 2015 *Quantum Statistical Mechanics: Equilibrium and Non-Equilibrium Theory from First Principles* (Bristol: IOP Publishing)

Attard P 2021 *Quantum Statistical Mechanics in Classical Phase Space* (Bristol: IOP Publishing)

Attard P 2022a Bose–Einstein condensation, the lambda transition, and superfluidity for interacting bosons arXiv:2201.07382

Attard P 2022b On the fountain effect in superfluid helium arXiv:2206.07914

Balibar S 2014 Superfluidity: how quantum mechanics became visible *History of Artificial Cold, Scientific, Technological and Cultural Issues* ed Gavroglu K (ed) (Dordrecht: Springer) pp 93–117

Balibar S 2017 Laszlo Tisza and the two-fluid model of superfluidity *C. R. Phys.* **18** 586

Batrouni G G, Ramstad T and Hansen A 2004 Free-energy landscape and the critical velocity of superfluid films *Phil. Trans. R. Soc. Lond.* A **362** 1595

Blinder S M 2011 Quantum-mechanical particle in a cylinder, Wolfram Demonstrations Project http:demonstrations.wolfram.com/QuantumMechanicalParticleInACylinder/ Accessed 23 Nov. 2021

Bogolubov N 1947 On the theory of superfluidity *J. Phys.* **11** 23

Clow J R and Reppy J D 1967 *Phys. Rev. Lett.* **19** 291

Donnelly R J 1995 The discovery of superfluidity *Phys. Today* **48** 30

Donnelly R J and Barenghi C F 1998 The observed properties of liquid Helium at the saturated vapor pressure *J. Phys. Chem. Ref. Data* **27** 1217

Donnelly R J 2009 The two-fluid theory and second sound in liquid helium *Phys. Today* **62** 34

Feynman R P 1954 Atomic theory of the two-fluid model of liquid helium *Phys. Rev.* **94** 262

Griffin A 1999 A brief history of our understanding of BEC: from Bose to Beliaev arXiv:cond-mat/9901123v1

Hammel (Jr) E F and Keller W E 1961 Fountain pressure measurements in liquid He II *Phys. Rev.* **124** 1641

Hill R W and Lounasmaa O V 1957 *Phil. Mag.* **2** 1943

Kapitza P 1938 Viscosity of liquid helium below the lambda point *Nature* **141** 74

Kawatra M P and Pathria R K 1966 *Phys. Rev.* **151** 132

Keesom W H and Wolfke M 1927 Two different liquid states of helium *Commun. Phys. Lab. Leiden* **190b** 22

Keesom W H, Clusius K and Keesom A P 1932 Ueber die specifische warme des fussigen heliums *Commun. Phys. Lab. Leiden* **219e and 221d**

Keller W E and Hammel (Jr) E F 1960 Heat conduction and fountain pressure in liquid He II *Ann. Phys.* **10** 202

Kramers H C, Wasscher J D and Gorter C J 1951 The specific heat of liouid helium between 0.25 and 1.9°K *Physica* **18** 329

Landau L D 1941 Theory of the superfluidity of helium II *Phys. Rev.* **60** 356

Lipa J A, Swanson D R, Nissen J A, Chui T C P and Israelsson U E 1996 Heat capacity and thermal relaxation of bulk helium very near the lambda point *Phys. Rev. Lett.* **76** 944

London F 1938 The λ-phenomenon of liquid helium and the Bose–Einstein degeneracy *Nature* **141** 643

London H 1939 Thermodynamics of the thermomechanical effect of liquid He II *Proc. Roy. Soc.* **A171** 484

London F and Zilsel P R 1948 Heat transfer in liquid helium II by internal convection *Phys. Rev.* **74** 1148

Newton I 1687 *Philosophiæ Naturalis Principia Mathematica* (London: S. Pepys)

Pathria R K 1972 *Statistical Mechanics* (Oxford: Pergamon)

Tisza L 1938 Transport phenomena in helium II *Nature* **141** 913

Entropy Beyond the Second Law (Second Edition)

Thermodynamics and statistical mechanics for equilibrium, non-equilibrium, classical, and quantum systems

Phil Attard

Chapter 10

High temperature superconductivity: electrons paired by entropy

'He wrapped himself in quotations as a beggar would enfold himself in the purple of emperors'

R Kipling (1918)

10.1 Introduction

Superconductivity has long been described by the Bardeen–Cooper–Schrieffer (BCS) theory (Bardeen *et al* 1957). This invokes the general notion of Cooper pairs: electrons with equal and opposite momentum, and equal and opposite spin, bound by an attractive potential (Cooper 1956). This is combined with a specific proposal for the binding potential: the attraction is due to the dynamic interaction of the electron pair with the vibrations of the solid lattice (Bardeen *et al* 1957). This explains the dependence of the transition temperature on the isotopic masses of the solid (Maxwell 1950, Reynolds *et al* 1950). The discovery of high temperature superconductors (Bednorz and Möller 1986, Wu *et al* 1987), which show no dependence on the isotopic mass, rules out the specific phonon exchange mechanism in these cases, although it leaves open the possibility for Cooper pairing if another mechanism for a binding potential could be found. Despite many proposals (Anderson 1987, Bickers *et al* 1987, Inui *et al* 1988, Gros *et al* 1988, Kotliar and Liu 1988, Mann 2011, Monthoux *et al* 1991), no consensus for such a potential has emerged, and so to date high temperature superconductivity remains unexplained.

Perhaps the clue is in the name. In reviewing the field one cannot help but be struck by the fact that existing treatments of superconductivity are exclusively quantum mechanical. And yet the name 'high temperature' surely implies that entropy is essential to the phenomenon. At low temperatures mechanical treatments suffice, and the low mass of electrons means that below 30 K or so the BCS quantum mechanical theory is applicable. But quantum mechanics is not quantum statistical mechanics, and the significantly higher temperatures that occur for the transition to superconductivity in the more recently discovered materials surely demands a qualitatively different, statistical approach.

In section 10.2 is developed the quantum statistical approach to superconductivity. Or, more accurately, it gives a theory for the condensation transition for fermions. This is entirely analogous to the approach to the λ-transition for interacting bosons given in chapter 8, including that it invokes the binary division approximation, and the no mixing approximation. Where it is specific to superconductivity is in the invocation of fermion pairs, section 10.2.2, which are more or less the same as Cooper pairs (Cooper 1956), to create effective bosons that undergo Bose–Einstein condensation. The significant difference with BCS theory is that the present theory is statistical mechanical, and here the fermion pairs are bound by an appropriate potential of mean force, which is temperature dependent and thus entropic in nature. This is what makes the present analysis applicable to high temperature superconductors.

In section 10.5 an explicit and rather attractive mechanism for the electron–electron pair potential of mean force is offered, which appears to be of the correct magnitude to account for the measured data for high temperature superconductivity. The bound fermion pairs invoked here appear qualitatively different to the Cooper pairs in BCS theory, at least

for their size and binding mechanism, which suggests that the present statistical thermodynamic theory may be the one applicable to high temperature superconductors.

10.2 General formulation and analysis

10.2.1 Symmetrization function for particles with spin

10.2.1.1 Symmetrized wave function

Consider a system of *N* particles. The set of commuting dynamical variables for one particle *j* may be taken to be $\mathbf{x}_j = \{\mathbf{q}_j, \sigma_j\}$, where $\mathbf{q}_j = \{q_{jx}, q_{jy}, q_{jz}\}$ is the position of particle *j*, and $\sigma_j \in \{-S, -S+1, \ldots, S\}$ is the *z*-component of its spin (see Messiah (1961 section 14.1), or Merzbacher (1970 section 20.5)). For electrons, $S = 1/2$ and $\sigma_j = \pm 1/2$; for fermions more generally *S* is half an odd integer. Note that here *σ* is *not* a spin operator or a Pauli spin matrix. Label the $2S + 1$ spin eigenstates of particle *j* by $s_j \in \{-S, -S+1, \ldots, S\}$, and the spin basis function by $\alpha_{s_j}(\sigma_j) = \delta_{s_j, \sigma_j}$. Note that this is *not* a spinor. For *N* particles, $\boldsymbol{\sigma} \equiv \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$, and similarly for **s** and **q**, and the basis functions for spin space are $\alpha_{\mathbf{s}}(\boldsymbol{\sigma}) = \delta_{\mathbf{s}, \boldsymbol{\sigma}} = \prod_{j=1}^N \delta_{s_j, \sigma_j}$.

Because the spin basis functions are Kronecker deltas, it is easy to show that when symmetrizing the wave functions, only permutations amongst particles with the same spin give a nonzero result. In other words, spin is one characteristic that identifies identical particles. This is the reason why two electrons with different spin can occupy the same single-particle state. That only permutations amongst same-spin particles are required is now demonstrated explicitly.

Let the number of particles with spin *s* be N_s, and $N = \sum_s N_s$. For reasons that will become clear shortly two cases will be dealt with simultaneously. The most general case allows permutations $\hat{\mathrm{P}}$ amongst all *N* particles irrespective of spin. There are $M = N!$ such permutations. The more specialized case only allows permutations amongst particles with the same spin $\hat{\mathrm{P}} = \prod_s \hat{\mathrm{P}}_s$, the factors of which commute. There are $M = \prod_s N_s!$ such permutations.

An unsymmetrized wave function $\psi(\mathbf{x})$ in general has symmetrized, normalized form (cf equation (7.65), or Attard (2021 section 6.4.1)),

$$\psi^{\pm}(\mathbf{x}) \equiv \frac{1}{\sqrt{\chi^{\pm} M}} \sum_{\hat{\mathrm{P}}} (\pm 1)^p \psi(\hat{\mathrm{P}}\mathbf{x}).$$

(10.1)

Normalization gives the symmetrization or overlap factor as

$$\chi^{\pm} \equiv \sum_{\hat{\mathrm{P}}} (\pm 1)^p \langle \psi(\hat{\mathrm{P}}\mathbf{x}) \mid \psi(\mathbf{x}) \rangle.$$

(10.2)

The upper sign is for bosons and the lower sign is for fermions. Henceforth in this chapter only the latter is considered.

Using momentum eigenfunctions with discrete momenta, the single-particle wave function for fermion *j* is

$$\Phi_{\mathbf{p}_j, s_j}(\mathbf{q}_j, \sigma_j) = \frac{1}{V^{1/2}} e^{-\mathbf{p}_j \cdot \mathbf{q}_j / i\hbar} \delta_{s_j, \sigma_j},$$

with $\mathbf{p}_j = \mathbf{n}_j \Delta_p$, $\mathbf{n}_j$ being a three-dimensional integer, and $\Delta_p = 2\pi\hbar/L$ being the spacing between momentum states, with $V = L^3$ being the volume of the cube to which the particles are confined (Messiah 1961, Merzbacher 1970).

The symmetrized, normalized, full system wave function is

$$\Phi_{\mathbf{p},\mathbf{s}}^{-}(\mathbf{q},\boldsymbol{\sigma}) = \frac{V^{-N/2}}{\sqrt{\chi_{\mathbf{p},\mathbf{s}}^{-}M}} \sum_{\hat{\mathrm{P}}} (-1)^p \prod_{j=1}^{N} \mathrm{e}^{-\mathbf{p}_j' \cdot \mathbf{q}_j/\mathrm{i}\hbar} \delta_{s_j',\sigma_j}$$

$$= \frac{V^{-N/2}}{\sqrt{\chi_{\mathbf{p},\mathbf{s}}^{-}M}} \sum_{\hat{\mathrm{P}}} (-1)^p \mathrm{e}^{-\mathbf{p}' \cdot \mathbf{q}/\mathrm{i}\hbar} \delta_{\mathbf{s}',\boldsymbol{\sigma}}.$$

The prime signifies the permuted list, $\mathbf{p}_j' = \{\hat{\mathrm{P}}\mathbf{p}\}_j$, and identically for the spin.

The Kronecker-delta $\delta_{\mathbf{s}',\boldsymbol{\sigma}}$ that appears here indicates that only permutations amongst particles with the same spin need be considered. Henceforth (until the introduction of pairs) take $\hat{\mathrm{P}} = \prod_s \hat{\mathrm{P}}_s$ and $M = \prod_s N_s!$.

10.2.1.2 Grand partition function

The grand partition function for fermions is given by equation (7.71),

$$\Xi^{-}(z,V,T) = \mathrm{TR}\prime\, \mathrm{e}^{-\beta\hat{\mathcal{H}}}$$

$$= \sum_N \frac{z^N}{M} \sum_{\mathbf{p}} \sum_{\mathbf{s}} \chi_{\mathbf{p},\mathbf{s}}^{-} \left\langle \Phi_{\mathbf{p},\mathbf{s}}^{-} | \mathrm{e}^{-\beta\hat{\mathcal{H}}} | \Phi_{\mathbf{p},\mathbf{s}}^{-} \right\rangle.$$

The symmetrization factor counts each state with the correct weight, which is equivalent to ensuring that each unique allowed state is counted once only with unit weight (see section 7.4, or Attard (2018, 2021)). Inserting the above definitions gives

$$\Xi^{-}(z,V,T) = \sum_N \frac{z^N}{\prod_s N_s!} \sum_{\hat{\mathrm{P}}} (-1)^p \sum_{\mathbf{p}} \sum_{\mathbf{s}} \left\langle \Phi_{\hat{\mathrm{P}}\mathbf{p},\hat{\mathrm{P}}\mathbf{s}} | \mathrm{e}^{-\beta\hat{\mathcal{H}}} | \Phi_{\mathbf{p},\mathbf{s}} \right\rangle$$

$$\approx \sum_N \frac{z^N}{\prod_s N_s! V^N} \sum_{\hat{\mathrm{P}}} (-1)^p \sum_{\mathbf{p},\mathbf{s}} \sum_{\boldsymbol{\sigma}} \int \mathrm{d}\mathbf{q}\, \mathrm{e}^{-\beta\mathcal{H}(\mathbf{q},\mathbf{p})} \mathrm{e}^{\mathbf{p}'\cdot\mathbf{q}/\mathrm{i}\hbar} \mathrm{e}^{-\mathbf{p}\cdot\mathbf{q}/\mathrm{i}\hbar} \delta_{\mathbf{s},\boldsymbol{\sigma}}$$

$$= \sum_N \frac{z^N}{\prod_s N_s! V^N} \sum_{\hat{\mathrm{P}}} (-1)^p \sum_{\mathbf{p},\mathbf{s}} \int \mathrm{d}\mathbf{q}\, \mathrm{e}^{-\beta\mathcal{H}(\mathbf{q},\mathbf{p})} \mathrm{e}^{-[\mathbf{p}-\mathbf{p}']\cdot\mathbf{q}/\mathrm{i}\hbar}$$

$$= \sum_N \frac{z^N}{\prod_s N_s! V^N} \sum_{\mathbf{p},\mathbf{s}} \int \mathrm{d}\mathbf{q}\, \mathrm{e}^{-\beta\mathcal{H}(\mathbf{q},\mathbf{p})} \eta^{-}(\mathbf{q},\mathbf{p},\mathbf{s}).$$

Here $\mathcal{H}(\mathbf{q},\mathbf{p}) = \mathcal{K}(\mathbf{p}) + U(\mathbf{q})$ is the Hamiltonian function of classical phase space, which at present is taken to be independent of spin.

The second equality neglects the commutation function (section 7.4.2). This is a short-ranged function and the approximation is valid when the system is dominated by long-ranged effects, which appears to be the case for Bose–Einstein condensation. The utility of

this approximation has been demonstrated for the λ-transition and for superfluidity (chapters 8 and 9). As will be discussed in section 10.4, the fermion pairs in the present theory require a short-ranged binding potential of mean force. The neglect of the commutation function may therefore have more serious consequences for the condensation of fermions than for bosons.

The symmetrization function is

$$\eta^-(\mathbf{p}, \mathbf{q}, \mathbf{s}) \equiv \prod_s \eta_s^-(\mathbf{p}^{N_s}, \mathbf{q}^{N_s})$$
$$= \prod_s \sum_{\hat{\mathrm{P}}_s} (-1)^{p_s} \mathrm{e}^{-[\mathbf{p}^{N_s} - \mathbf{p}'^{N_s}] \cdot \mathbf{q}^{N_s}/\mathrm{i}\hbar}.$$

(10.7)

Because only permutations between same spin fermions contribute, one can factorize the permutation operator, and hence also the symmetrization function.

10.2.2 Fermion pairs

10.2.2.1 Effective bosons

The usual definition of a Cooper pair is that the momenta must be equal and opposite, $\mathbf{p}_1 = -\mathbf{p}_2$, and also the spins, $s_1 = -s_2$ (Cooper 1956). This is widely accepted to act as a zero momentum, zero spin effective boson. The more general definition that will be used here likewise insists that the two fermions have equal and opposite momenta, $\mathbf{p}_1 = -\mathbf{p}_2$. But for the spins it is only necessary that $s_1 \neq s_2$ in order to prevent internal permutations within the Cooper pair, which would cancel the bosonic ones. For the case of electrons, which are spin-half fermions, the condition $s_1 \neq s_2$ is equivalent to $s_1 = -s_2$. Electrons of course are the main focus in superconductivity, in which case there is no difference in the two formulations of Cooper pairs. The more general pair may be called an effective spin-$s_1 s_2$ boson, with the convention $s_1 < s_2$. The total spin $s_1 + s_2$ is not sufficient to label the pair. There are $(2S + 1)S$ distinct species of pairs.

Consider four fermions in two pairs: $\{1, 2\}$ and $\{3, 4\}$ (figure 10.1). That is, $\mathbf{p}_1 = -\mathbf{p}_2$ and $\mathbf{p}_3 = -\mathbf{p}_4$. Suppose that $s_1 = s_3$ and $s_2 = s_4$. Then the only permutations permitted are between 1 and 3 and between 2 and 4. Hence the symmetrization function (see section 7.4, or Attard (2018, 2021)) for these four fermions is

$$\sum_{\hat{\mathrm{P}}} (-1)^p \mathrm{e}^{-\mathbf{q} \cdot [\mathbf{p} - \mathbf{p}']/\mathrm{i}\hbar} \delta_{\mathbf{s}', \mathbf{s}} = 1 - \mathrm{e}^{-\mathbf{q}_{13} \cdot \mathbf{p}_{13}/\mathrm{i}\hbar} - \mathrm{e}^{-\mathbf{q}_{24} \cdot \mathbf{p}_{24}/\mathrm{i}\hbar} + \mathrm{e}^{-\mathbf{q}_{13} \cdot \mathbf{p}_{13}/\mathrm{i}\hbar} \mathrm{e}^{-\mathbf{q}_{24} \cdot \mathbf{p}_{24}/\mathrm{i}\hbar}$$

$$\approx 1 + \mathrm{e}^{-\mathbf{q}_{12} \cdot \mathbf{p}_{13}/\mathrm{i}\hbar} \mathrm{e}^{-\mathbf{q}_{34} \cdot \mathbf{p}_{31}/\mathrm{i}\hbar}.$$

(10.8)

The prime indicates the permuted eigenvalues. The two single transposition fermionic terms have been neglected in the final equality because of their rapid fluctuation compared to the two retained bosonic terms. This is because the fermionic exponents depend upon the separation between the pair, which for most pairs is macroscopic, whereas the bosonic exponents depend upon the internal size of each pair, which is molecular. The latter assertion follows by analyzing the exponent for the double transposition, which, suppressing the factor of $-\mathrm{i}\hbar$, is

$$\mathbf{q}_{13} \cdot \mathbf{p}_{13} + \mathbf{q}_{24} \cdot \mathbf{p}_{24} = \frac{1}{2}\left[(\mathbf{q}_1 + \mathbf{q}_2) - (\mathbf{q}_3 + \mathbf{q}_4)\right] \cdot \mathbf{p}_{13} + \frac{1}{2}\left[(\mathbf{q}_1 - \mathbf{q}_2) - (\mathbf{q}_3 - \mathbf{q}_4)\right] \cdot \mathbf{p}$$

$$+ \frac{1}{2}\left[(\mathbf{q}_1 + \mathbf{q}_2) - (\mathbf{q}_3 + \mathbf{q}_4)\right] \cdot \mathbf{p}_{24} - \frac{1}{2}\left[(\mathbf{q}_1 - \mathbf{q}_2) - (\mathbf{q}_3 - \mathbf{q}_4)\right]$$

$$= \mathbf{Q}_{13} \cdot \mathbf{P}_{13} + \frac{1}{2}(\mathbf{q}_{12} - \mathbf{q}_{34}) \cdot (\mathbf{p}_{13} - \mathbf{p}_{24})$$

$$= \mathbf{q}_{12} \cdot \mathbf{p}_{13} + \mathbf{q}_{34} \cdot \mathbf{p}_{31}.$$

(10.9)

The first two equalities hold in general; the final equality holds for the Cooper pairs. The center of mass separation is $\mathbf{Q}_{13} = \mathbf{Q}_1 - \mathbf{Q}_3 = (\mathbf{q}_1 + \mathbf{q}_2)/2 - (\mathbf{q}_3 + \mathbf{q}_4)/2$, the total momentum difference is $\mathbf{P}_{13} = (\mathbf{p}_1 + \mathbf{p}_2) - (\mathbf{p}_3 + \mathbf{p}_4)$, and the 'locations' of the effective bosons are $\mathbf{q}_{12} = \mathbf{q}_1 - \mathbf{q}_2$ and $\mathbf{q}_{34} = \mathbf{q}_3 - \mathbf{q}_4$. For a Cooper pair the total momentum of each pair is identically zero, as is the difference, which gives the final equality.

Figure 10.1. Paired fermions 1 and 2 (left), and 3 and 4 (right).
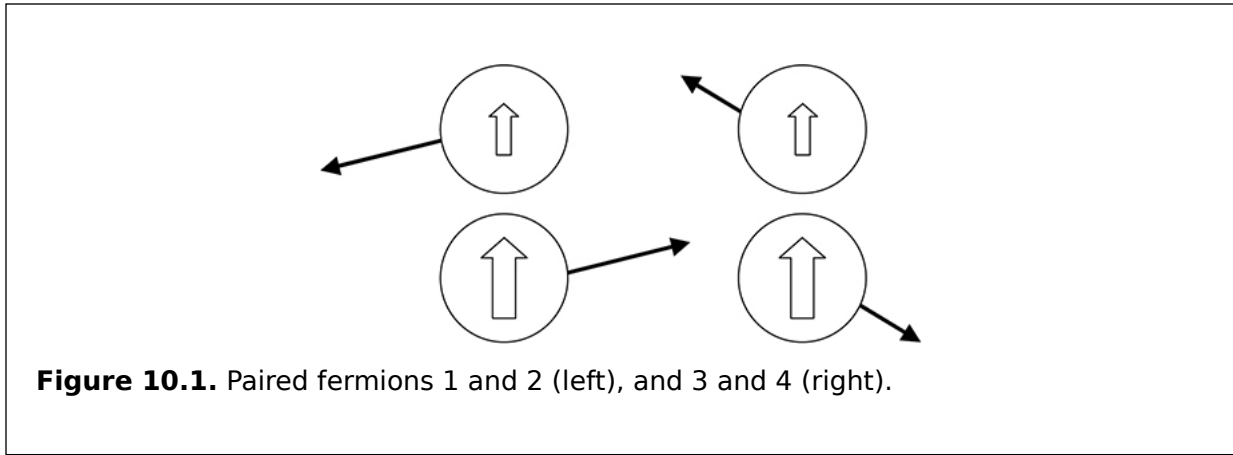
In the form of the final equality of the penultimate equation, the permutation weight is exactly the dimer symmetrization weight written in momentum permutation loop form for two boson molecules located at $\mathbf{q}_{12}$ and $\mathbf{q}_{34}$ with momenta $\mathbf{p}_1$ and $\mathbf{p}_3$, respectively (see section 7.4, or Attard (2018, 2021)). This is a non-local expression since it does not depend upon the distance between the pairs. Since the result depends on the size (i.e. internal separation) of the two Cooper pairs, each pair is like a bosonic dipolar molecule rather than an atom. If there exists an attractive potential so that the size of a pair is small, then the fluctuations in this term due to the range of momentum differences are also small. There are infinitely more pairs of fermions with macroscopic separations Q_{13} than there are with microscopic separations, and for these the fermionic terms fluctuate infinitely more rapidly than the bosonic terms. The former average to zero; the latter average close to unity. The Cooper pair formulation removes the macroscopic separation between the center of masses, Q_{13}. This is what makes the permutation of Cooper pairs non-local and creates the analogy with Bose–Einstein condensation and superfluidity (chapters 8 and 9).

10.2.2.2 Generalised fermion pairs with nonzero momentum

In section 9.2, which is concerned with the nature of Bose–Einstein condensation in liquid helium and superfluidity, it was argued that there is nothing particularly special about the momentum ground state, and that condensation actually occurred into multiple low-lying momentum states. In so far as the understanding of superconductivity is largely based on an analogy with superfluidity, it would appear that a similar situation arises here. That is, the definition of a Cooper pair should be generalised to two fermions with different spin

and nonzero total momentum. It appears neither necessary nor likely that the bosonic entities that condense to give superconductivity are restricted to the ground momentum state.

In practice consider the following arrangement of momenta for the two pairs above:

$$\mathbf{p}_1 = \frac{1}{2}\mathbf{P} + \boldsymbol{\delta}_1 \quad \text{and} \quad \mathbf{p}_2 = \frac{1}{2}\mathbf{P} - \boldsymbol{\delta}_1,$$

(10.10)

and

$$\mathbf{p}_3 = \frac{1}{2}\mathbf{P} + \boldsymbol{\delta}_3 \quad \text{and} \quad \mathbf{p}_4 = \frac{1}{2}\mathbf{P} - \boldsymbol{\delta}_3.$$

(10.11)

Recall that $s_1 = s_3$ and $s_2 = s_4$. The two pairs have the same total momentum $\mathbf{P}$, which need not be zero, so that the total momentum difference remains zero $\mathbf{P}_{13} = 0$. However, the fermions with the same spin should have different momenta, $\boldsymbol{\delta}_{13} \neq 0$, so that they do not occupy the same momentum state. The non-trivial part of the bosonic exponent in this more general case is

$$
\begin{aligned}
\mathbf{q}_{13} \cdot \mathbf{p}_{13} + \mathbf{q}_{24} \cdot \mathbf{p}_{24} &= \mathbf{Q}_{13} \cdot \mathbf{P}_{13} + \frac{1}{2}(\mathbf{q}_{12} - \mathbf{q}_{34}) \cdot (\mathbf{p}_{13} - \mathbf{p}_{24}) \\
&= \frac{1}{2}(\mathbf{q}_{12} - \mathbf{q}_{34}) \cdot (\boldsymbol{\delta}_{13} + \boldsymbol{\delta}_{13}) \\
&= \mathbf{q}_{12} \cdot \mathbf{p}_{13} + \mathbf{q}_{34} \cdot \mathbf{p}_{31}.
\end{aligned}
$$

(10.12)

This is a non-local expression in position that involves only the sizes of the bound pairs. It also involves the difference in the momentum excess of the two pairs, which cannot be zero because that would mean fermions occupying the same momentum state. Presumably one could integrate these over the momentum continuum. In any case the permutation entropy associated with non-locality can be very large; in the case of ^{4}He it is what drives the λ-transition and superfluidity.

A similar result holds for an l-loop composed of fermion pairs all with the same total momentum. For l pairs of fermions $\{2j-1, 2j\}$, $j = 1, \ldots, l$, let the center of mass of a pair be $\mathbf{Q}_{2j-1} = [\mathbf{q}_{2j-1} + \mathbf{q}_{2j}]/2$, and the momenta be $\mathbf{p}_{2j-1} = \mathbf{P}/2 + \boldsymbol{\delta}_{2j-1}$, and $\mathbf{p}_{2j} = \mathbf{P}/2 - \boldsymbol{\delta}_{2j-1}$. In position loop form, the essence of the bosonic exponent is (mod l)

$$
\begin{aligned}
\sum_{j=1}^{l} \{\mathbf{q}_{2j-1,2j+1} \cdot \mathbf{p}_{2j-1} + \mathbf{q}_{2j,2j+2} \cdot \mathbf{p}_{2j}\} &= \sum_{j=1}^{l} \{(2\mathbf{Q}_{2j-1} - \mathbf{q}_{2j} - 2\mathbf{Q}_{2j+1} + \mathbf{q}_{2j+2}) \cdot (\mathbf{P}/2 \\
&\quad + (2\mathbf{Q}_{2j-1} - \mathbf{q}_{2j-1} - 2\mathbf{Q}_{2j+1} + \mathbf{q}_{2j+1}) \cdot (\mathbf{P}/2 \\
&= \sum_{j=1}^{l} \{(-\mathbf{q}_{2j} + \mathbf{q}_{2j+2}) \cdot \boldsymbol{\delta}_{2j-1} - (-\mathbf{q}_{2j-1} + \mathbf{q}_{2j+} \\
&= \frac{1}{2}\sum_{j=1}^{l} (\mathbf{q}_{2j-1,2j} - \mathbf{q}_{2j+1,2j+2}) \cdot \mathbf{p}_{2j-1,2j}.
\end{aligned}
$$

(10.13)

This is a non-local expression that involves only the internal size and momentum departure of each pair.

Although ultimately an accurate description of the system on the far-side of the superconducting transition will demand these generalized fermion pairs with nonzero total momentum, in the present chapter the focus is on the approach to, and the location of, the transition. In this case, as a first approximation, defining a Cooper pair to be fermions with equal and opposite momentum, should be as adequate, just as it was in chapter 8 for ^{4}He approaching the λ-transition. Therefore, the analysis in this chapter is restricted to this first approximation with the conventional definition of a Cooper pair having zero total momentum.

10.2.2.3 Rotation of bound fermion pair

There is a need for a correlation between fluid structure and a bound fermion pair in order for the latter to be long-lived and therefore statistically relevant. This holds in both the binary division approximation and more generally.

In the BCS mechanical theory of superconductivity, the binding potentials are the lattice vibrations, which have a characteristic frequency. It is sometimes pictured as one electron distorting the solid lattice ions leaving an excess positive charge as it passes through, which attracts a second electron. Since the two fermions travel in opposite directions along the lattice wave, their lifetime as a pair is limited.

A different idea is to suppose that the bound fermions orbit about their center of mass, as sketched in figure 10.2. Like moths attracted to a flame the fermions circle, bound by their mutual attraction and the cavity field. In a dense fluid it is likely that such rotational motion can only take place in a cavity, which, as a collective fluctuation, likely has a relatively long lifetime. In the binary division approximation the center of mass would have zero momentum. The commutation function contribution reduces the relative momentum parallel to the pair axis (section 8.6.3), which is consistent with the orbital motion.

Figure 10.2. Bound fermion pair rotating in a fluid cavity.
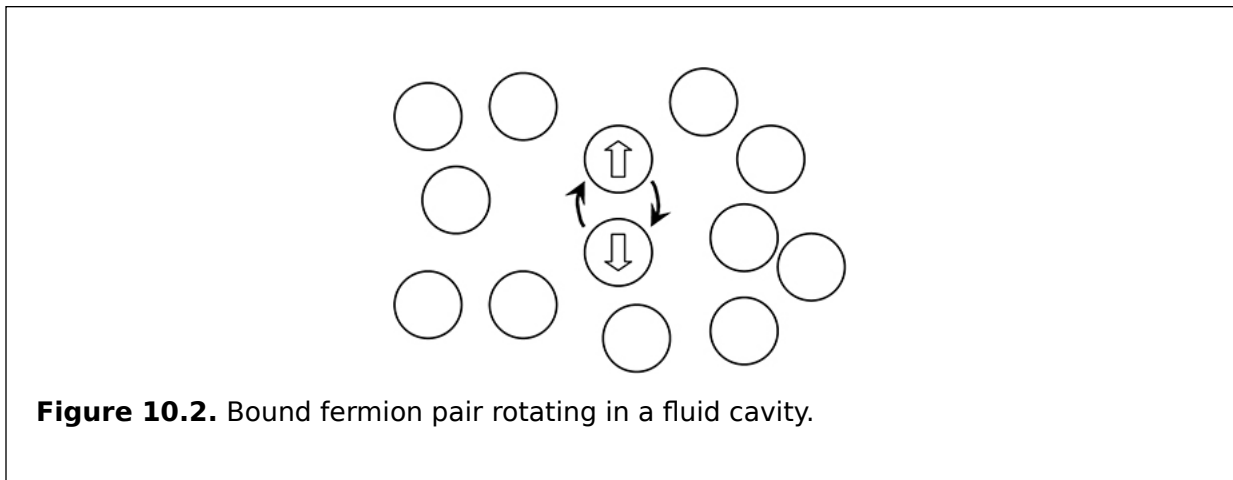
It might seem unlikely that such a relatively large cavity would spontaneously form, that it would have a significant lifetime, and that two fermions would find themselves bound in orbit within it. (A specific mechanism for the binding of electrons is discussed in section 10.5.) However, the large entropy available from the non-local permutations of the effective bosons formed from such bound orbiting fermion pairs suggests that such configurations are much more probable than classical considerations imply.

One can crudely estimate the cavity probability, and incorporate the cost of the cavity into the fugacity per fermion for the bound pair, z_0. (The fugacity weight for $N_{0,s\prime s\prime\prime}$ $s\prime s\prime\prime$-pairs is $z_0^{2N_{0,s\prime s\prime\prime}}$.) For an incompressible fluid one has approximately

$$z_0^2 \equiv z^2 (z_0^{\text{ex}})^2 \approx z^2 \mathrm{e}^{-\beta p v_0^{\text{ex}}}.$$

(10. 14)

The excess cavity volume v_0^{ex} is the additional volume required to accommodate the rotating bound pair over the volume that they would occupy as unpaired fermions. Roughly speaking $v_0^{\text{ex}} \approx 4\pi \bar{q}^3/3 - 2/\rho$; undoubtedly more refined estimates are possible. The excess fugacity goes alongside the bound pair internal weight (see below).

Obviously this rotational mechanism for bound fermion pairs is a little speculative. Also, it is challenging to obtain a more reliable quantitative estimate of the fugacity of the bound pair. For these reasons results will be presented below with $z_0 = z$ as well as with $z_0 = z z_0^{\text{ex}}$ using the above excess cavity volume v_0^{ex}.

10.2.2.4 Pair weight for bound Cooper pairs

One can demonstrate the idea that a Cooper pair is an effective boson molecule by performing the classical momentum integral for the four fermions comprising the above pair dimer. The momentum integral is

$$
\begin{aligned}
I_4 &\equiv \Delta_p^{-6} \int \mathrm{d}\mathbf{p}^4 \, \delta(\mathbf{p}_1 + \mathbf{p}_2) \, \delta(\mathbf{p}_3 + \mathbf{p}_4) \, \mathrm{e}^{-\beta \mathcal{K}(\mathbf{p}^4)} \mathrm{e}^{-\mathbf{q}_{12}\cdot\mathbf{p}_{13}/i\hbar} \mathrm{e}^{-\mathbf{q}_{34}\cdot\mathbf{p}_{31}/i\hbar} \\
&= \Delta_p^{-6} \int \mathrm{d}\mathbf{p}_1 \, \mathrm{d}\mathbf{p}_3 \, \mathrm{e}^{-2\beta p_1^2/2m} \mathrm{e}^{-2\beta p_3^2/2m} \mathrm{e}^{-[\mathbf{q}_{12}-\mathbf{q}_{34}]\cdot\mathbf{p}_1/i\hbar} \mathrm{e}^{-[\mathbf{q}_{34}-\mathbf{q}_{12}]\cdot\mathbf{p}_3/i\hbar} \\
&= \left\{ 2^{-3/2} V \Lambda^{-3} \mathrm{e}^{-\pi[\mathbf{q}_{12}-\mathbf{q}_{34}]^2/2\Lambda^2} \right\}^2 \\
&\approx 2^{-3/2} V \Lambda^{-3} \frac{\Lambda \mathrm{e}^{-\pi q_{12}^2/\Lambda^2}}{q_{12}\sqrt{2\pi}} \, \sinh^{1/2}(2\pi q_{12}^2/\Lambda^2) \\
&\quad \times 2^{-3/2} V \Lambda^{-3} \frac{\Lambda \mathrm{e}^{-\pi q_{34}^2/\Lambda^2}}{q_{34}\sqrt{2\pi}} \, \sinh^{1/2}(2\pi q_{34}^2/\Lambda^2).
\end{aligned}
$$

(10. 15)

Notice that the third equality would give zero weight if the sizes (internal separations) of the Cooper pairs were large and different (in size or orientation). Since the permutations of Cooper pairs are non-local, the vast majority are at macroscopic separations and are therefore uncorrelated. The final equality follows by averaging the coupling exponential over the alignment angle of the bosonic 'dipole',

$$
\begin{aligned}
\frac{1}{2} \int_{-1}^{1} \mathrm{d}x \, \mathrm{e}^{2\pi q_{12} q_{34} x/\Lambda^2} &= \frac{\Lambda^2}{2\pi q_{12} q_{34}} \, \sinh\left(\frac{2\pi q_{12} q_{34}}{\Lambda^2} \right) \\
&\approx \frac{\Lambda}{q_{12}\sqrt{2\pi}} \, \sinh^{1/2}(2\pi q_{12}^2/\Lambda^2) \\
&\quad \times \frac{\Lambda}{q_{34}\sqrt{2\pi}} \, \sinh^{1/2}(2\pi q_{34}^2/\Lambda^2).
\end{aligned}
$$

(10. 16)

The second equality is valid if $q_{12} \approx q_{34}$, which is the case when the pairs are bound at the minimum in the pair potential of mean force $\bar{q}$.

In this formulation the weight of the pair dimer factorizes into the product of weights of each pair. A mean field approximation (fix the neighbors in the most likely parallel

configuration and average the intervening pair successively around the permutation loop) shows that a similar factorization holds for the pair trimer, etc. Therefore, each Cooper pair is a boson molecule with average internal weight due to the transposition

$$\nu_{\mathrm{mf}} \equiv \left\langle \mathrm{e}^{-\pi q_{12}^2/\Lambda^2} \frac{\Lambda}{q_{12}\sqrt{2\pi}} \sinh^{1/2}(2\pi q_{12}^2/\Lambda^2) \right\rangle_{\mathrm{bnd}}$$

$$\approx \frac{\Lambda \mathrm{e}^{-\pi\bar{q}^2/\Lambda^2}}{\bar{q}\sqrt{2\pi}} \sinh^{1/2}(2\pi\bar{q}^2/\Lambda^2).$$

(10.17)

This holds for a bound Cooper pair in which the pair potential of mean force has a relatively narrow minimum at $\bar{q}$. This expression for the internal weight for the bound Cooper pair effective boson is less than unity; the expression itself is likely an overestimate due to the mean field approximation, as is shown shortly. The internal weight goes to zero as the size of Cooper pair goes to infinity. The remaining factor of $2^{-3/2}\Lambda^{-3}V$ from the momentum integral holds for all Cooper pairs and will be included explicitly below.

The factorization is valid if the sizes of the bound Cooper pairs are all at the minimum of the pair potential of mean force $\bar{q}$. This means that the departure Δ_q from the position of the minimum must be small,

$$\Delta_q \ll \bar{q}/2 \quad \text{and} \quad \Delta_q\bar{q} \ll \Lambda^2/2\pi.$$

(10.18)

The width Δ_q defines bound Cooper pairs: the smaller it is, the fewer bound Copper pairs exist for condensation, but if it exceeds these criteria, then the calculated mean field internal weight is not valid. In fact the internal weight goes to zero exponentially with increasing size (internal separation) of the Cooper pair. If the curvature of the pair potential of mean force is large enough such that $\beta\bar{w}''\Delta_q^2/2 \gg 1$, then one has a natural upper bound for the bound Cooper pairs.

Figure 10.3 shows that the exact internal weight for a bound Cooper pair of size $\bar{q}$ decreases slowly with increasing loop size from the mean field result, equation (10.17), which is the exact result at $l = 2$. It appears to reach a plateau for large loop size, which value would dominate the symmetrization function. This plateau is confirmed by the exact asymptotic expression,

$$\nu = \mathrm{e}^{-\zeta^2}\left\{1 + \frac{l+2}{6l}\zeta^4 + \mathcal{O}(\zeta^8)\right\}, \quad \zeta \to 0,$$

(10.19)

where $\zeta \equiv \sqrt{\pi}\,\bar{q}/\Lambda$. Small ζ is the relevant regime for Cooper pair condensation and superconductivity.
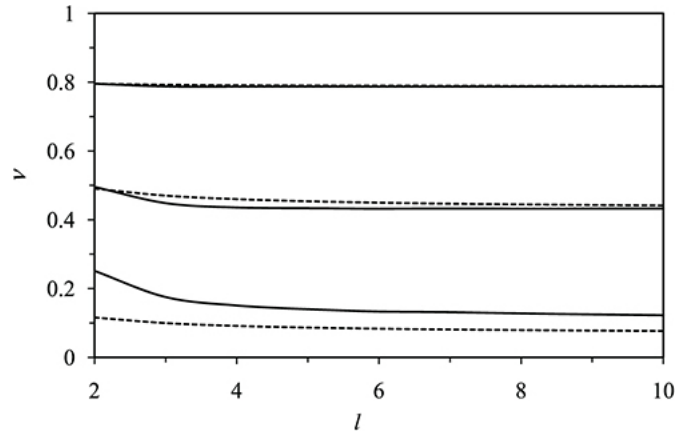
Figure 10.3. Internal weight per bound Cooper pair of size $\bar{q}$ as a function of permutation loop length for $\zeta \equiv \sqrt{\pi}\,\bar{q}/\Lambda = 0.5$ (top), 1.0 (middle) and 2.0 (bottom). The solid curves are the exact numerical result, and the dashed curves are the small z asymptote. The mean field value is the exact value at $l = 2$.

The internal weight of a Cooper pair reflects the formation of bosonic position permutation loops, which is only effective when the members of the pair are bound closer than about the thermal wavelength. This weight represents a correction to the bosonic position permutation loop entropy that will eventually be included in the partition function.

10.2.2.5 Bound Cooper pairs

Only bound Cooper pairs will be considered here. In these the fermions must be separated by less than $\bar{q} + \Delta_q$. At any instant other fermions might form a non-local $s\prime s\prime\prime$-pair, but these do not count. It is assumed that no fermion may belong to more than one $s\prime s\prime\prime$-pair at a time. Let there be $N_{0,s\prime s\prime\prime}$ bound Cooper $s\prime s\prime\prime$-pairs.

For future use, the radial distribution function may be used to define the bound volume,

$$
v_{\mathrm{bnd}} \equiv 4\pi \int_0^{\bar{q}+\Delta_q} \mathrm{d}q \; q^2 g(q)
$$

$$
\approx 4\pi\bar{q}^2 \mathrm{e}^{-\beta\bar{w}}\sqrt{2\pi/\beta\bar{w}\prime\prime}.
$$

(10.20)

The second equality assumes that the radial distribution function, $g(q) = \mathrm{e}^{-\beta w(q)}$, is sufficiently sharply peaked about the minimum in the pair potential of mean force to enable a second order expansion and evaluation of the Gaussian integral. It is not essential to make the Gaussian approximation to the integral, since it can be evaluated numerically by fixing Δ_q to satisfy the internal weight criterion given above. The Gaussian results do enable a transparent analysis and a physical interpretation of the nature of the bound Cooper pairs.

The symmetrization function for Cooper pairs is

$$
\eta_{0,s\prime s\prime\prime}^{+} \approx \nu^{N_{0,s\prime s\prime\prime}}(N_{0,s\prime s\prime\prime}!).
$$

The superscript + indicates that these are treated as effective bosons. The result follows (20) because the permutations of the bound Cooper pairs are non-local and each carries the (21) internal weight of the effective boson, ν.

10.2.3 Grand potential with paired fermions

10.2.3.1 Continuum momentum limit

In the treatment of Bose–Einstein condensation in chapter 8, the continuum momentum limit was taken and the discrete momentum ground state was added explicitly to the continuum integral over the excited momentum states. This is the binary division approximation originally introduced by F London (1938) in his ideal boson analysis of the λ-transition. As discussed in sections 8.6.6 and 9.2, this binary division is qualitatively useful, particularly on the high temperature side of the transition, and for locating the transition, even though more sophisticated analysis recognizes that condensation actually occurs into multiple muliply-occupied low-lying momentum states, not just the ground momentum state.

One can perform a similar trick for Cooper pairs. For fermions 1 and 2 write the discrete sum over momentum states as

$$
\sum_{\mathbf{p}_1,\mathbf{p}_2} = \sum_{\mathbf{p}_1}\left\{\delta_{\mathbf{p}_2,-\mathbf{p}_1} + \sum_{\mathbf{p}_2}^{(\mathbf{p}_2 \neq -\mathbf{p}_1)}\right\}
$$
$$
= \sum_{\mathbf{p}_1}\left\{\delta_{\mathbf{p}_2,-\mathbf{p}_1} + \Delta_p^{-3}\int\mathrm{d}\mathbf{p}_2\right\}
$$
$$
= \frac{1}{\Delta_p^6}\int\mathrm{d}\mathbf{p}_1\,\mathrm{d}\mathbf{p}_2\left\{\Delta_p^3\delta(\mathbf{p}_2+\mathbf{p}_1)+1\right\}.
$$

(10. 22)

In transforming to the continuum integral, it is assumed that the point $\mathbf{p}_2 = -\mathbf{p}_1$ is a set of measure zero and so this formulation it is not really double counting. In this form the integral covers the possible states of the two fermions as both paired and unpaired. A similar procedure for the product of momentum sums for *N* fermions leads to a binomial expansion, the terms of which consist of a certain number of paired and a certain number of unpaired fermions, as is now derived.

10.2.3.2 Numbers of paired and unpaired fermions

Let N_s be the total number of spin-*s* fermions, and let $N_{1,s}$ be the number of unpaired spin-*s* fermions. Let $N_{0,ss\prime}$ be the number of bound paired fermions with spin pair $ss\prime$. To count pairs uniquely, $s < s\prime$; for convenience define $N_{0,ss\prime} = N_{0,s\prime s}$ and $N_{0,ss} = 0$. Obviously

$$
N_s = N_{1,s} + \sum_{s\prime=-S}^{S} N_{0,ss\prime}.
$$

(10. 23)

Also, $N = \sum_{s=-S}^{S} N_s$. Recall that a fermion may not belong to more than one bound pair at a time, and that unbound fermion pairs (i.e. those separated by more than $\bar{q} + \Delta_q$) are counted as unpaired.

In the first instance take the Hamiltonian to be independent of spin, in which case $\bar{N}_s = N/(2S+1)$. This requirement is relaxed when a spin-dependent potential is

included.

Treat the paired and unpaired fermions as $(S+1)(2S+1)$ different species, and only allow permutations within each species. This is similar to the no mixing approximation that has been used in the treatment of the λ-transition, section 8.4.2. For the occupancy $\{\underline{N}_0, \underline{N}_1\} \equiv \{N_{0,ss'}, N_{1,s}\}$ the total number of permutations restricted to each species is

$$M = \prod_{s=-S}^{S} N_{1,s}! \prod_{s'=-S}^{S} \prod_{s''=s'+1}^{S} N_{0,s's''}!$$
$$= \prod_{s,s'<s''} N_{1,s}!\, N_{0,s's''}!.$$

(10. 24)

The second equality defines the abbreviated notation that will be used. This permutation number M goes directly into the denominator in conjunction with the symmetrization factor formalism for the partition function that treats the paired and unpaired fermions as different species.

One can also see this from the binomial expansion that arises from the transformation to the continuum mentioned above. The usual multinomial factor that arises in the expansion is

$$\prod_{s,\, s'<\, s''} \frac{N_s!}{N_{1,s}!\, N_{0,s's''}!}.$$

(10. 25)

Inserting this into the grand partition function, equation (10.6), the numerator here cancels with the denominator there, leaving M as the denominator in that equation.

10.2.3.3 Grand partition function

In view of either of these two results, the grand partition function for paired and unpaired fermions is

$$\Xi^- = \sum_{\underline{N}_0, \underline{N}_1} \frac{z^N (z_0^{\mathrm{ex}})^{2N_0}}{\prod_{s,s'<s''} N_{1,s}!\, N_{0,s's''}!}$$
$$\times \sum_{\mathbf{p}} \sum_{\hat{P}} (-1)^{p_{1,s}} \left\langle \phi_{\mathbf{p}}(\mathbf{q}) | e^{-\beta \hat{\mathcal{H}}} | \phi_{\mathbf{p}}\left(\hat{P}\mathbf{q}\right) \right\rangle$$
$$\approx \sum_{\underline{N}_0, \underline{N}_1} \frac{z^N (z_0^{\mathrm{ex}})^{2N_0} V^{-N}}{\prod_{s,s'<s''} N_{1,s}!\, N_{0,s's''}!} \sum_{\mathbf{p}} \int d\mathbf{q}\; e^{-\beta \mathcal{H}(\mathbf{q},\mathbf{p})} \eta^-(\mathbf{q},\mathbf{p},\mathbf{s}).$$

(10. 26)

The commutation function has been neglected in the final equality. The discrete momentum sums can be replaced directly by continuum momentum integrals since the paired and unpaired fermions have been explicitly identified. The phase space point of the ss' paired fermions is denoted $\Gamma^{2N_{0,ss'}} \equiv \{\mathbf{q}^{2N_{0,ss'}}, \mathbf{p}^{2N_{0,ss'}}\}$, and the phase space point of unpaired s fermions is denoted $\Gamma^{N_{1,s}} \equiv \{\mathbf{q}^{N_{1,s}}, \mathbf{p}^{N_{1,s}}\}$. The total permutator is $\hat{P} = \prod_{s,s'<s''} \hat{P}_{0,s's''} \hat{P}_{1,s}$, which factors commute.

Because permutations are confined to amongst fermions with the same spin, and because of the no mixing approximation (i.e. permutations between paired and unpaired fermions may be neglected), the symmetrization function fully factorizes

$$\eta^-(\mathbf{q}, \mathbf{p}, \mathbf{s}) = \prod_{s, s\prime < s\prime\prime} \eta^+_{0, s\prime s\prime\prime}(\mathbf{\Gamma}^{2N_{0,s\prime s\prime\prime}}) \, \eta^-_{1,s}(\mathbf{\Gamma}^{N_{1,s}}).$$

<div align="right">(10.27)</div>

The symmetrization function for paired fermions, $\eta^+_{0, s\prime s\prime\prime}$, was expressed above in terms of the number of bound Cooper pairs, equation (10.21). It is independent of the point in phase space and can be taken outside of the integrals for the partition function.

The symmetrization function for unpaired fermions is

$$\eta^-_{1,s}(\mathbf{\Gamma}^{N_{1,s}}) = \sum_{\hat{P}_{1,s}}^{N_{1,s}!} (-1)^{p_{1,s}} e^{-[\mathbf{p}^{N_{1,s}} - \mathbf{p}\prime^{N_{1,s}}] \cdot \mathbf{q}^{N_{1,s}}/i\hbar}.$$

<div align="right">(10.28)</div>

Here $\mathbf{p}\prime^{N_{1,s}} \equiv \hat{P}_{1,s} \mathbf{p}^{N_{1,s}}$, and $p_{1,s}$ is the parity of the permutation $\hat{P}_{1,s}$.

In the general classical phase space formulation of quantum statistical mechanics, the permutation loop expansion, which consists of products of loops, can be written as the exponential of a series of single loops, sections 7.5 and 7.7 (Attard 2018, 2021). It follows that the grand potential is the sum of loop potentials, each of which is the classical average of a sum over single permutation loops of a given size. This holds only for the unpaired fermions, whose symmetrization function has $(2S + 1)$ factors. Each factor gives a corresponding loop grand potential; for $l \geqslant 2$ these can all be treated independently.

10.2.3.4 Configuration integral containing bound fermion pairs

Shortly the monomer or classical grand potential will be given, and this depends upon the classical configurational integral containing $N_0 = \sum_{s\prime < s\prime\prime} N_{0,s\prime s\prime\prime}$ bound fermion pairs. This is different to the classical configurational integral $Q(N, V, T)$ unconstrained by such pairing, where $N = N_1 + 2N_0$. The two are related as

$$\begin{aligned}
Q(N, V, T \mid \underline{N}_0) &= \int_V d\mathbf{q}^{N_1} d\mathbf{q}^{N_0} \int_{\Delta_V} d\mathbf{q}\prime^{N_0} \, e^{-\beta U(\mathbf{q}^N)} \\
&= Q(N, V, T) \left\langle \prod_{j=1}^{N_0} \Theta(\bar{q} + \Delta_q - q_{jj\prime}) \right\rangle_{N,V,T} \\
&= Q(N, V, T) \left\{ \frac{4\pi}{V} \int_0^{\bar{q} + \Delta_q} dq \, q^2 g(q) \right\}^{N_0} \\
&\equiv Q(N, V, T) \prod_{s\prime < s\prime\prime} \left(\frac{v_{\text{bnd}}}{V} \right)^{N_{0,s\prime s\prime\prime}}.
\end{aligned}$$

<div align="right">(10.29)</div>

Here $j\prime$ is the fermion bound to j, $\Theta(q)$ is the Heaviside step function, and v_{bnd} is the bound volume, equation (10.20).

10.2.3.5 Monomer grand potential

Using this result, the monomer or classical grand potential is

$$\mathrm{e}^{-\beta\Omega^{(1)}(z,V,T)} = \Xi_{\mathrm{cl}}(z,V,T)$$

$$= \sum_{\underline{N}_0,\underline{N}_1} \frac{z^N (z_0^{\mathrm{ex}})^{2N_0} \prod_{s\prime < s\prime\prime} \nu^{N_{0,s\prime s\prime\prime}} N_{0,s\prime s\prime\prime}!}{V^N \prod_{s,s\prime < s\prime\prime} N_{1,s}! N_{0,s\prime s\prime\prime}!}$$

$$\times \prod_{s,s\prime < s\prime\prime} \frac{\Delta_p^{-3N_{1,s}}}{\Delta_p^{3N_{0,s\prime s\prime\prime}}} \int \mathrm{d}\mathbf{p}^{N_{0,s\prime s\prime\prime}} \, \mathrm{d}\mathbf{p}^{N_{1,s}} \mathrm{e}^{-\beta\mathcal{K}(\mathbf{p}^N)}$$

$$\times \int_V \mathrm{d}\mathbf{q}^{N_1} \mathrm{d}\mathbf{q}^{N_0} \int_{\Delta_V} \mathrm{d}\mathbf{q}\prime^{N_0} \, \mathrm{e}^{-\beta U(\mathbf{q}^N)}$$

$$= \sum_{\underline{N}_0,\underline{N}_1} \frac{z^N \prod_{s\prime < s\prime\prime} (z_0^{\mathrm{ex}\,2}\nu)^{N_{0,s\prime s\prime\prime}}}{V^N \prod_s N_{1,s}!} \prod_{s,s\prime < s\prime\prime} \frac{(V/\Lambda^3)^{N_{1,s}}}{(2^{3/2}\Lambda^3/V)^{N_{0,s\prime s\prime\prime}}} \, Q(N,V,T \mid \underline{N}_{\,}$$

$$= \sum_{\underline{N}_0,\underline{N}_1} \frac{z^N}{V^N} Q(N,V,T) \prod_{s\prime < s\prime\prime} \left(\frac{z_0^{\mathrm{ex}\,2}\nu v_{\mathrm{bnd}}}{2^{3/2}\Lambda^3} \right)^{N_{0,s\prime s\prime\prime}} \prod_s \frac{V^{N_{1,s}}}{\Lambda^{3N_{1,s}} N_{1,s}!}.$$

(10. 30)

Here the continuum momentum limit has been taken. The classical configuration integral, $Q(N,V,T)$, is independent of the spin state of the fermions (unless a spin-dependent potential is added). As above, $N = N_1 + 2N_0 = \sum_s N_{1,s} + 2\sum_{s\prime < s\prime\prime} N_{0,s\prime s\prime\prime}$.

10.2.3.6 Unpaired grand potential

For paired fermions, the symmetrization function was given above as equation (10.21). This is independent of the point in phase space. It was taken outside of the phase space integrals for the partition function and appears above in the monomer grand potential as the factors involving $N_{0,s\prime s\prime\prime}$.

For unpaired fermions the loop grand potential is straightforward to derive as, apart from a factor of $(-1)^{l-1}$, it is identical to that for bosons (see sections 7.5, 7.7, and 8.4.2) (Attard 2018, 2021).

The unpaired loop grand potentials $l \geqslant 2$ are classical averages (Attard 2021 section 5.3), which can be taken canonically

$$-\beta\Omega_{1,s}^{-,(l)} = \left\langle \eta_{1,s}^{-,(l)}(\mathbf{p}^{N_s},\mathbf{q}^{N_s}) \right\rangle_{\underline{N}_0,\underline{N}_1}^{\mathrm{cl}}$$

$$= (-1)^{l-1} \left\langle G^{(l)}(\mathbf{q}^{N_{1,s}}) \right\rangle_{\underline{N}_0,\underline{N}_1}^{\mathrm{cl}}$$

$$= (-1)^{l-1} \left(\frac{N_{1,s}}{N} \right)^l \left\langle G^{(l)}(\mathbf{q}^N) \right\rangle_N^{\mathrm{cl}}$$

$$\equiv N_{1,s}(-1)^{l-1} \left(\frac{N_{1,s}}{N} \right)^{l-1} g^{(l)}.$$

(10. 31)

The fermionic anti-symmetrization factor $(-1)^{l-1}$ appears explicitly. Apart from it, all factors are positive. The average has been transformed from the mixed $\{\underline{N}_0, \underline{N}_1\}$ system to the classical configurational system of N fermions that does not distinguish their state (Attard 2021). The factor of $(N_{s,1}/N)^l$ is the uncorrelated probability that l fermions

chosen at random in the original mixed system are all unpaired and in the same state. The Gaussian position loop function is

$$G^{(l)}(\mathbf{q}^N) = \sum_{j_1,\ldots,j_l} {}^{\prime\prime} \mathrm{e}^{-\pi q_{j_l,j_1}^2/\Lambda^2} \prod_{k=1}^{l-1} \mathrm{e}^{-\pi q_{j_k,j_{k+1}}^2/\Lambda^2}.$$

(10.32)

The double prime indicates that no two indeces may be equal and that distinct loops must be counted once only. There are $N!/(N-l)!l$ distinct l-loops here, the overwhelming number of which are negligible upon averaging. Since the pure excited momentum state permutation loops are compact in configuration space, one can define an intensive form of the average loop Gaussian, $g^{(l)} \equiv \langle G^{(l)}(\mathbf{q}^N)\rangle_N^{\mathrm{cl}}/N$. This is convenient because it does not depend upon $N_{1,s}$.

10.2.4 Maximum entropy for paired fermions

The constrained grand potential is given by

$$-\beta\Omega^-(\underline{N}_0, \underline{N}_1 \mid z, V, T) = -\beta\Omega^{(1)}(\underline{N}_0, \underline{N}_1 \mid z, V, T) - \beta \sum_{s=-S}^{S} \sum_{l=2}^{\infty} \Omega_{1,s}^{-,(l)}(\underline{N}_0, \underline{N}_1).$$

(10.33)

The monomer and paired loop potentials on the right-hand side are given above.

This is the constrained total entropy and the optimum number of paired and unpaired fermions is determined by maximizing it. It is most convenient to take the derivatives at constant number N.

In the absence of a spin-dependent potential, all spins are equal, $N_s = N/(2S+1)$. Similarly the pair number is independent of the spin pair $s's''$, $N_{0,s's''} = [1 - \delta_{s',s''}]\tilde{N}_0$, so that $N_s = N_{1,s} + \sum_{s'} N_{0,ss'} = N_{1,s} + 2S\tilde{N}_0$. One can ensure the derivative at constant N by taking $\mathrm{d}N_{1,s'} = \mathrm{d}N_{1,s''} = -\mathrm{d}N_{0,s's''}$. Setting $N_{1,s'} = N_{1,s''}$ after the differentiation one has

$$\frac{\mathrm{d}(-\beta\Omega)}{\mathrm{d}N_{1,s'}} = -2\ln\frac{\Lambda^3 N_{1,s'}}{V} - \ln\frac{z_0^{\mathrm{ex}\,2}\nu v_{\mathrm{bnd}}}{2^{3/2}\Lambda^3}$$

$$+ 2\sum_{l=2}^{\infty} (-1)^{l-1} l \frac{N_{1,s'}^{l-1}}{N^{l-1}} g^{(l)}.$$

(10.34)

The fraction of unpaired spin-s fermions is $f_{1,s} = N_{1,s}/N \leqslant 1/(2S+1)$. If the right-hand side is positive, then $N_{1,s'}$ should be increased. Setting the derivative to zero determines $\bar{N}_{1,s'}$ for fixed N and V.

At high temperatures, the intensive Gaussian loop integrals are negligible, $g^{(l)} \approx 0$. Neglecting them and focussing only on the logarithmic terms, the derivative vanishes when

$$\bar{\rho}_{1,s'}\Lambda^3 = \sqrt{\frac{2^{3/2}\Lambda^3}{z_0^{\mathrm{ex}\,2}\nu v_{\mathrm{bnd}}}}.$$

(10.35)

Of course one must have $\bar{\rho}_{1,s\prime} \leqslant \rho/(2S+1)$. At high temperatures one expects the binding volume to be small, $v_{\text{bnd}} \lesssim \Lambda^3$, as it is proportional to the number of neighbors within $\bar{q} + \Delta_q$ of a fermion. In this case the right-hand side is larger than $\rho\Lambda^3/(2S+1)$, which is the upper limit on the density of fermions with a given spin. As the temperature is decreased, the binding volume increases as the depth of the minimum in the potential of mean force generally increases. So the right-hand side decreases, and the left-hand side increases with decreasing temperature. At some temperature this will give a solution $\bar{\rho}_{1,s\prime} \leqslant \rho/(2S+1)$. This is the point at which bound fermion pairs first form. Of course this result may be perturbed by the contribution from the intensive gaussian loop integrals.

This argument sets the excess fugacity for the pair to unity, $z_0^{\text{ex}\,2} = 1$. It will be shown below that for ^{3}He the excess fugacity decreases with decreasing temperature. If in consequence the right-hand side grows faster than the left-hand side with decreasing temperature, then condensation will not occur (unless the loop contributions become significant and have the right sign).

10.2.4.1 Spin-dependent external potential

Add an external spin-dependent one-body potential, $U(\sigma) = -B\sum_j \sigma_j$. The parameter B will be called the magnetic field strength. Assume that there is no pair or many-body potential that depends upon the spins. The classical configuration integral becomes

$$Q(N,V,T,B) = \prod_{s,s\prime<s\prime\prime} e^{s\beta BN_{1,s}}e^{(s\prime+s\prime\prime)\beta BN_{0,s\prime s\prime\prime}}Q(N,V,T).$$

(10.36)

To the monomer grand potential $-\beta\Omega^{(1)}$ should be added

$$\beta B\sum_s sN_{1,s} + \beta B\sum_{s\prime<s\prime\prime}(s\prime+s\prime\prime)N_{0,s\prime s\prime\prime}.$$

(10.37)

In general the derivatives are messy because the $N_{1,s}$ and $N_{0,s\prime s\prime\prime}$ vary with spin. But for electrons, $S = 1/2$, one can carry out the derivatives at constant N in two ways to give two equations for two unknowns. The first way is to take $\mathrm{d}N_{1,\uparrow} = \mathrm{d}N_{1,\downarrow} = -\mathrm{d}N_{0,\uparrow\downarrow}$. This gives

$$\frac{\mathrm{d}(-\beta\Omega)}{\mathrm{d}N_{1,\uparrow}} = \ln\frac{V}{\Lambda^3 N_{1,\uparrow}} + \ln\frac{V}{\Lambda^3 N_{1,\downarrow}} - \ln\frac{z_0^{\text{ex}\,2}\nu v_{\text{bnd}}}{2^{3/2}\Lambda^3}$$
$$+ \sum_{l=2}^{\infty}(-1)^{l-1}l\frac{N_{1,\uparrow}^{l-1}+N_{1,\downarrow}^{l-1}}{N^{l-1}}g^{(l)}.$$

(10.38)

The second way is to take $\mathrm{d}N_{1,\uparrow} = -\mathrm{d}N_{1,\downarrow}$ and $\mathrm{d}N_{0,\uparrow\downarrow} = 0$, which gives

$$\frac{\mathrm{d}(-\beta\Omega)}{\mathrm{d}N_{1,\uparrow}} = \ln\frac{V}{\Lambda^3 N_{1,\uparrow}} - \ln\frac{V}{\Lambda^3 N_{1,\downarrow}} + \beta B$$
$$+ \sum_{l=2}^{\infty}(-1)^{l-1}l\frac{N_{1,\uparrow}^{l-1}-N_{1,\downarrow}^{l-1}}{N^{l-1}}g^{(l)}.$$

(10.39)

Adding these together one has for the optimum fraction of up-spins, $f_{1,\uparrow} = N_{1,\uparrow}/N$,

$$0 = \beta B - \ln \frac{z_0^{\text{ex } 2} \nu v_{\text{bnd}}}{2^{3/2}\Lambda^3} - 2 \ln[\rho\Lambda^3 \bar{f}_{1,\uparrow}] + 2\sum_{l=2}^{\infty} (-1)^{l-1} l \bar{f}_{1,\uparrow}^{l-1} g^{(l)},$$
(10.40)

and subtracting the second from the first one has for down-spins,

$$0 = -\beta B - \ln \frac{z_0^{\text{ex } 2} \nu v_{\text{bnd}}}{2^{3/2}\Lambda^3} - 2 \ln[\rho\Lambda^3 \bar{f}_{1,\downarrow}] + 2\sum_{l=2}^{\infty} (-1)^{l-1} l \bar{f}_{1,\downarrow}^{l-1} g^{(l)}.$$
(10.41)

Neglecting the intensive Gaussian loop contributions, these give

$$\bar{f}_{1,\uparrow} \approx \frac{2^{3/4}\text{e}^{\beta B/2}}{\rho\Lambda^3 \sqrt{z_0^{\text{ex } 2} \nu v_{\text{bnd}}\Lambda^{-3}}} \quad \text{and} \quad \bar{f}_{1,\downarrow} \approx \frac{2^{3/4}\text{e}^{-\beta B/2}}{\rho\Lambda^3 \sqrt{z_0^{\text{ex } 2} \nu v_{\text{bnd}}\Lambda^{-3}}}.$$
(10.42)

Since $\bar{f}_{1,\uparrow} + \bar{f}_{1,\downarrow} \leqslant 1$, this gives an upper limit on the external field beyond which the condensation into fermion pairs is disrupted,

$$\cosh \frac{\beta B_{\text{crit}}}{2} \approx 2^{-7/4}\rho\Lambda^3 \sqrt{z_0^{\text{ex } 2} \nu v_{\text{bnd}}\Lambda^{-3}}.$$
(10.43)

Of course this is qualitatively rather than quantitatively accurate because the loop Gaussians are likely to make a significant contribution in any regime in which condensation occurs.

10.3 Results for ³He

Table 10.1 gives the intensive loop Gaussians for Lennard-Jones helium-3. The Lennard-Jones parameters model helium, $\varepsilon_{\text{He}} = 10.22k_{\text{B}}$ J and $\sigma_{\text{He}} = 0.255\,6$ nm (van Sciver 2012). The relation between the Lennard-Jones dimensionless temperature and the actual temperature is $T[\text{K}] = 10.22T^*$. The ³He atom is a spin-half fermion because it comprises a pair of protons and a pair of electrons, each with opposite spin and occupying the same single-particle state, and a single spin-half neutron. The Monte Carlo simulation algorithm used to obtain the results is detailed in Attard (2021 section 5.3.2). In general, the loop Gaussians increase with decreasing temperature. They also decrease with increasing loop size, except at the lowest temperatures studied. This suggests that terminating the loop series with $l^{\text{max}} = 5$ at these temperatures will give reliable results since each term is weighted by the corresponding power of $N_{1,s}/N \leqslant 1/(2S+1)$ (in the absence of an external magnetic field).

Table 10.1. Intensive loop Gaussian for saturated liquid Lennard-Jones He³ at various temperatures, $T^* \equiv k_{\text{B}}T/\varepsilon$.
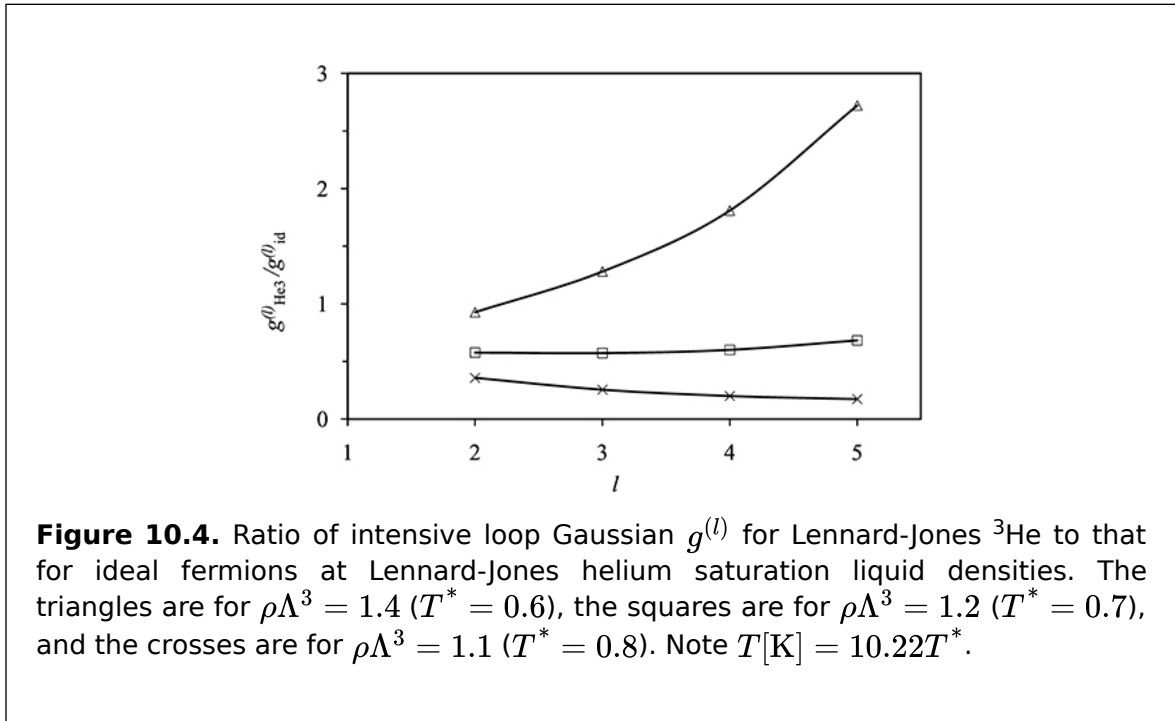
T^*	$g^{(2)}$	$g^{(3)}$	$g^{(4)}$	$g^{(5)}$	$g^{(6)}$
1.2	5.43E−03	2.22E−04	1.67E−05	1.77E−06	

T^*	$g^{(2)}$	$g^{(3)}$	$g^{(4)}$	$g^{(5)}$	$g^{(6)}$
1.1	1.21E−02	7.84E−04	4.69E−05	3.60E−05	
1.0	2.09E−02	2.40E−03	4.58E−04	1.14E−04	
0.9	3.80E−02	6.85E−03	1.92E−03	6.93E−04	
0.8	6.84E−02	1.93E−02	8.05E−03	4.32E−03	2.57E−03
0.7	1.25E−01	5.54E−02	3.49E−02	2.80E−02	2.46E−02
0.6	2.31E−01	1.64E−01	1.59E−01	1.94E−01	
0.5	4.23E−01	4.80E−01	7.32E−01	1.39E+00	

Figure 10.4 compares the intensive loop Gaussians $g^{(l)}$ for Lennard-Jones ^{3}He and for ideal fermions by plotting their ratio for three temperatures. The intensive loop Gaussian for ideal particles is (cf equation (7.119), or section 7.6, or Attard (2021 section 4.1))

$$
\begin{aligned}
g_{\mathrm{id}}^{(l)} &= \frac{1}{N}\big\langle G^{(l)}(\mathbf{q}^N)\big\rangle_{N,\mathrm{cl,id}} \\
&= \frac{1}{N}\frac{N!}{(N-l)!l}\frac{1}{V^l}\int \mathrm{d}\mathbf{q}^l\; \mathrm{e}^{-\pi q_{l,1}^2/\Lambda^2}\prod_{k=1}^{l-1}\mathrm{e}^{-\pi q_{k,k+1}^2/\Lambda^2} \\
&= \frac{\rho^{l-1}}{l}\Lambda^{3(l-1)}l^{-3/2}.
\end{aligned}
$$

(10.44)

Figure 10.4. Ratio of intensive loop Gaussian $g^{(l)}$ for Lennard-Jones ^{3}He to that for ideal fermions at Lennard-Jones helium saturation liquid densities. The triangles are for $\rho\Lambda^3 = 1.4$ ($T^* = 0.6$), the squares are for $\rho\Lambda^3 = 1.2$ ($T^* = 0.7$), and the crosses are for $\rho\Lambda^3 = 1.1$ ($T^* = 0.8$). Note $T[\mathrm{K}] = 10.22T^*$.

It can be seen that at the lowest temperature, the intensive loop Gaussians for the Lennard-Jones model is larger than for ideal fermions, increasingly so as the order of the

loop increases. The presence of an attractive potential of mean force strongly affects the particle correlation functions and hence the values of the intensive loop Gaussians that determine the number of Cooper pairs.

Figure 10.5 compares the simulated heat capacity for Lennard-Jones ³He with that for Lennard-Jones ⁴He. It can be seen that the heat capacity for the fermions is less than that for the bosons. This is due of course to the alternating signs of the terms in the former series. The divergence in the ⁴He heat capacity that is becoming apparent at the lowest temperatures shown is a manifestation of the λ-transition. Compared to ⁴He, ³He has a smaller mass giving a larger thermal wavelength, and therefore the loop series with a fixed number of terms becomes unreliable at a higher temperature. It is noticeable that terminating the loop series for the fermion ³He at an even number of terms (equal number of positive and negative coefficients) gives a lower heat capacity than that terminated at an odd number of terms. The growth in heat capacity that marks the λ-transition in Lennard-Jones ⁴He at these temperatures does not occur in Lennard-Jones ³He. The decrease in heat capacity with decreasing temperature in the latter, which is apparent at the lowest temperatures shown, is likely a manifestation of the well-known fact that only fermions near the Fermi surface contribute. The loop series is probably rather inefficient for fermions at low temperatures as in essence it requires an infinite number of terms to ensure rigorously that no two fermions with the same spin occupy the same momentum state.
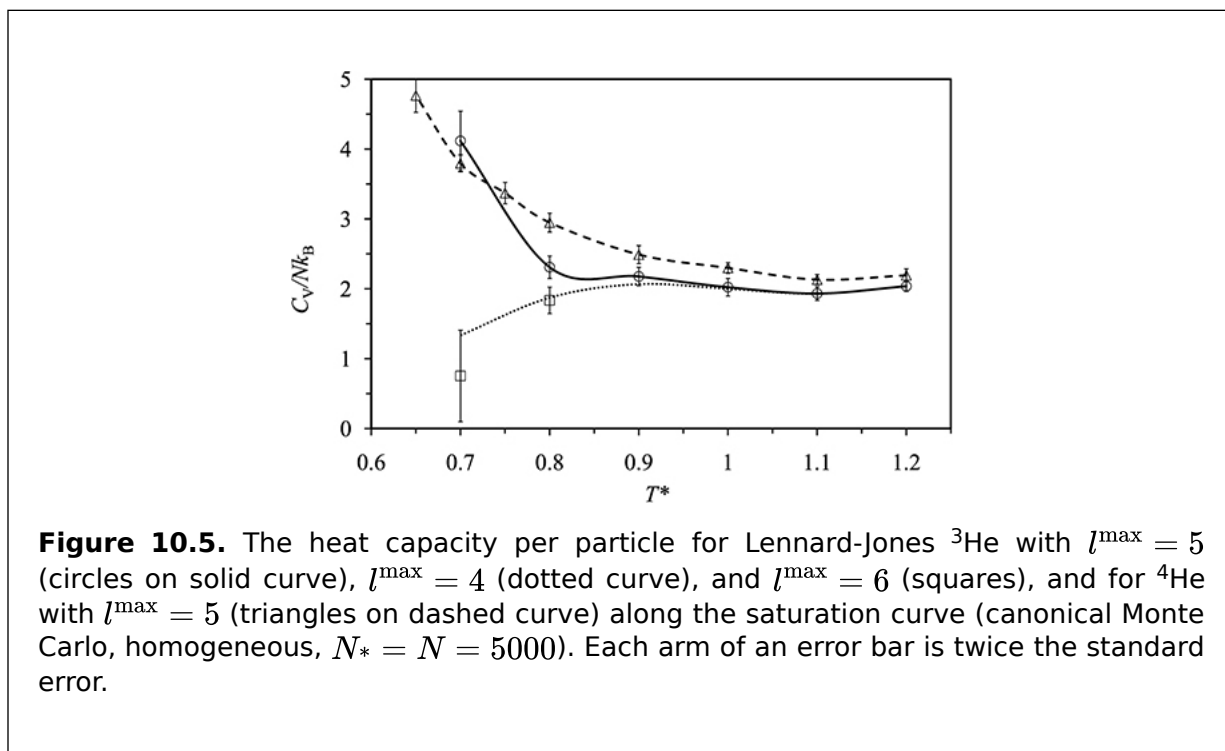

Figure 10.5. The heat capacity per particle for Lennard-Jones ³He with $l^{\max} = 5$ (circles on solid curve), $l^{\max} = 4$ (dotted curve), and $l^{\max} = 6$ (squares), and for ⁴He with $l^{\max} = 5$ (triangles on dashed curve) along the saturation curve (canonical Monte Carlo, homogeneous, $N_* = N = 5000$). Each arm of an error bar is twice the standard error.

The results for both ³He and ⁴He in figure 10.5 were obtained with no condensed particles, $N_* = N$. Hence in the case of fermions the results are independent of Cooper pairing and the effective boson approximation.

Table 10.2 gives the values of various parameters obtained by classical canonical simulations of 5000 Lennard-Jones helium-3 atoms in a homogeneous system with periodic boundary conditions. The density is the liquid saturation density, which was obtained along the saturation curve for a liquid drop. It can be seen that the minimum in the pair potential of mean force gets deeper, and its curvature increases, with decreasing temperature. The

internal weight ν and the bound volume $\mathcal{v}_{\text{bnd}}$ are surprisingly insensitive to temperature. The excess fugacity z_0^{ex} for paired fermions, equation (10.14), decreases with decreasing temperature at lower temperatures, primarily due to the increase in βp along the liquid–vapor saturation curve. (The value greater than unity for $T^* = 0.9$ arises from a small negative pressure, $\beta p/\rho = -0.024\,3(98)$, which may be due to a minor inaccuracy in the liquid density.) The parameter $\sqrt{8}\,\bar{q}/\Lambda^2\sqrt{\beta\bar{w}\prime\prime}$ should be less than about unity in order for the present numerical approximations to be valid (cf equation (10.18)). It is derived from the requirement that $\sqrt{2/\beta\bar{w}\prime\prime} \ll \Delta_q \ll \Lambda^2/2\bar{q}$, with the left-hand bound being necessary for the validity of the Gaussian approximation, and the right-hand bound being necessary for the validity of the mean field internal weight. The relation between the Lennard-Jones dimensionless temperature and the actual temperature is $T[\text{K}] = 10.22T^*$.
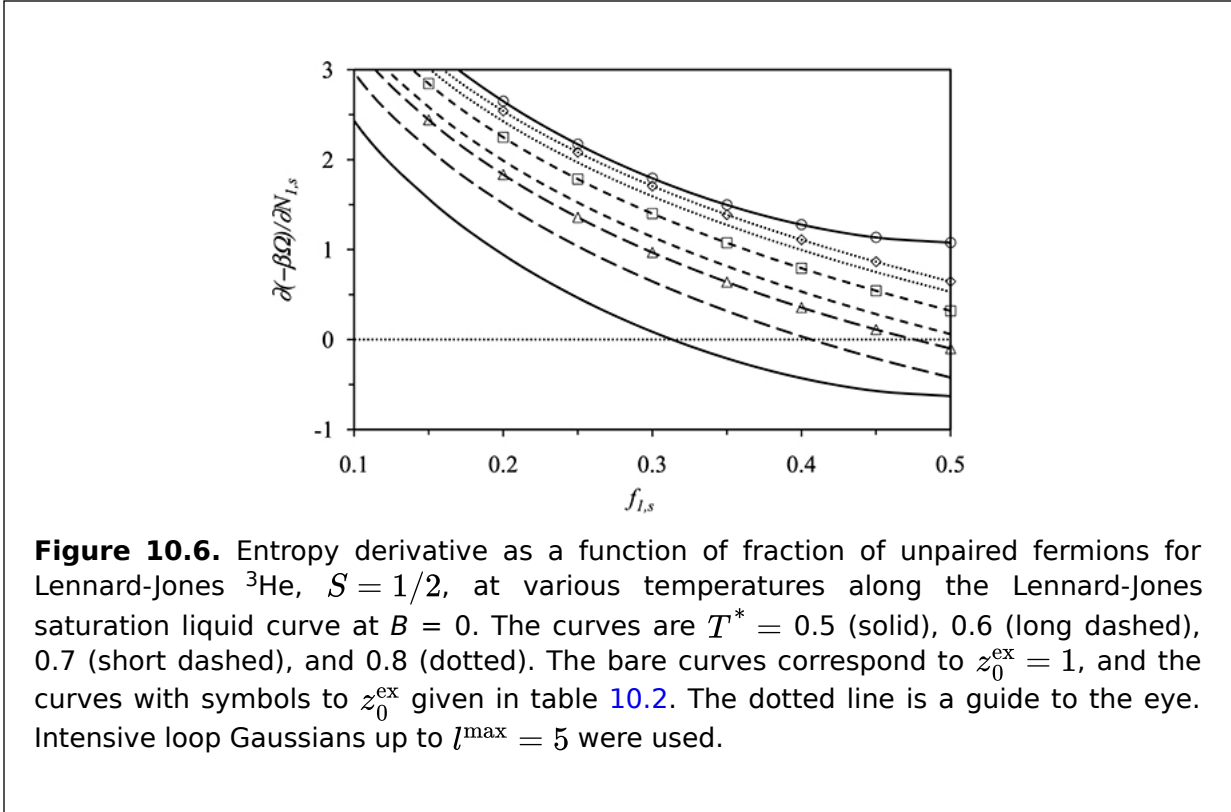
Table 10.2. Parameters for saturated liquid Lennard-Jones He^3 at various temperatures, $T^* \equiv k_{\text{B}}T/\varepsilon$. The Lennard-Jones diameter σ is the unit of length.

T^*	1.2	1.1	1.0	0.9	0.8	0.7	0.6	0.5
ρ	0.53	0.63	0.70	0.75	0.80	0.85	0.89	0.93
Λ	1.13	1.18	1.23	1.30	1.38	1.47	1.59	1.74
$\rho\Lambda^3$	0.75	1.03	1.32	1.65	2.10	2.71	3.58	4.92
$\bar{q}$	1.100	1.097	1.093	1.091	1.090	1.089	1.088	1.086
$\beta\bar{w}$	-0.74	-0.81	-0.88	-0.95	-1.03	-1.12	-1.20	-1.32
$\beta\bar{w}\prime\prime$	64.49	87.17	87.39	100.78	135.60	151.93	176.07	209.80
ν	0.32	0.33	0.35	0.37	0.39	0.42	0.45	0.49
$z_{0,\text{ex}}^2$	0.95	0.96	0.95	1.05	0.89	0.77	0.72	0.18
$\mathcal{v}_{\text{bnd}}$	9.92	9.12	9.68	9.63	9.03	9.25	9.38	9.58
$\dfrac{\sqrt{8}\bar{q}/\Lambda^2}{\sqrt{\beta\bar{w}\prime\prime}}$	0.31	0.24	0.22	0.18	0.14	0.11	0.09	0.07

The bound volume in table 10.2 was calculated using the Gaussian approximation, the second equality in equation (10.20). For $T^* = 0.5$ this is $\mathcal{v}_{\text{bnd}} = 9.58$. Evaluating the actual integral instead, using $\Delta_q = \Lambda^2/2\pi\bar{q} = 0.44$ gives $\mathcal{v}_{\text{bnd}} = 14.10$, $\Delta_q = \Lambda^2/5\pi\bar{q} = 0.18$ gives $\mathcal{v}_{\text{bnd}} = 10.39$, and $\Delta_q = \Lambda^2/10\pi\bar{q} = 0.089$ gives $\mathcal{v}_{\text{bnd}} = 7.63$. (These expressions ensure the validity of the internal weight ν.) The expression $\Delta_q = \sqrt{2/\beta\bar{w}\prime\prime} = 0.098$ gives $\mathcal{v}_{\text{bnd}} = 8.42$. These are all reasonably close to the value estimated with the Gaussian approximation. All of the results below use the value of $\mathcal{v}_{\text{bnd}}$ given in the table.

Figure 10.6 shows the thermodynamic force acting on the fraction of unpaired spin-half fermions, with positive values driving an increase. The curves that don't pass thorough zero have a positive derivative for the whole domain, which means that the system in its

optimum thermodynamic state is composed entirely of unpaired fermions, $\bar{f}_{1,s} = 1/(2S+1)$.

Figure 10.6. Entropy derivative as a function of fraction of unpaired fermions for Lennard-Jones ^{3}He, $S = 1/2$, at various temperatures along the Lennard-Jones saturation liquid curve at $B = 0$. The curves are $T^* = 0.5$ (solid), 0.6 (long dashed), 0.7 (short dashed), and 0.8 (dotted). The bare curves correspond to $z_0^{\text{ex}} = 1$, and the curves with symbols to z_0^{ex} given in table 10.2. The dotted line is a guide to the eye. Intensive loop Gaussians up to $l^{\max} = 5$ were used.

First the case $z_0^{\text{ex}} = 1$, which neglects the cost of forming the cavity to accommodate the rotating bound fermion pair. For $T^* \lesssim 0.7$ the derivative passes through zero, which corresponds to a maximum in the total entropy. It can be seen that the zero is a stable solution, because positive values of the derivative to the left of the zero give a thermodynamic force that increases the number of unpaired fermions, and negative values to the right of the zero give a thermodynamic force that decreases their number. It can be seen that the optimum fraction of unpaired fermions decreases with decreasing temperature below the transition, which is the same as saying that the fraction of paired fermions increases.

Including the excess pair fugacity, equation (10.14), with the values in table 10.2, it can be seen that at all temperatures shown the force is more positive than the corresponding result with $z_0^{\text{ex}} = 1$. This means that the cost of the accommodating cavity suppresses the formation of the rotating bound fermion pair. For most temperatures the force does not pass through zero, which means that there is no condensation transition. For $T^* = 0.6$ the force becomes slightly negative, and gives an optimum fraction of unpaired fermions of $\bar{N}_{1,s}/N \approx 0.47$, which is practically the maximum value. The uncertainty in the value of the excess fugacity makes it reasonable to say that even in this case there is no condensation transition. The increase in force for $T^* = 0.5$ in going from $z_0^{\text{ex}} = 1$ to $z_0^{\text{ex}} = 0.18$ is noticeably larger than for the other temperatures. For $T^* = 0.5$ the pressure is given by $\beta p/\rho = 0.567(33)$, and for $T^* = 0.6$ it is $\beta p/\rho = 0.115(10)$. Presumably at lower temperatures the excess fugacity is even smaller, with the Lennard-Jones ^{3}He turning solid.

The results in figure 10.6 were obtained along the line $f_{1,s} = f_{1,s''}$, as follows by taking $dN_{1,s'} = dN_{1,s''} = -dN_{0,s's''}$. For $z_0^{\text{ex}} = 1$, there is a second zero of the thermodynamic force that occurs for $\bar{f}_{1,s} \neq \bar{f}_{1,s''}$, which can be obtained by taking $dN_{1,s'} = -dN_{1,s''}$. (One can see the beginning of a parabola at the extremity of the range for the curve $T^* = 0.5$ in figure 10.6.) This second zero corresponds to an entropy minimum and is therefore unstable. In practice this does not represent a thermodynamic barrier to condensation because the system can follow the first path starting from the fully unpaired state, $f_{1,s} = f_{1,s''} = 1/(2S+1)$, which symmetry arguments demand in the absence of a magnetic field.

If one accepts that condensation implies individual bound fermion pairs are long-lived, and that these necessarily rotate, and that these therefore require an accommodating cavity, then the calculations with the excess fugacity being less than unity are the more realistic, even if the quantitative values are derived from a rather simple expression. If so, then the results suggest that there is no condensation transition in Lennard-Jones ^{3}He down to $T^* = 0.5$ ($T = 5.1$ K). By extrapolation, the excess fugacity only gets smaller with decreasing temperature. This suggests that condensation will never occur in Lennard-Jones ^{3}He, at least not by this mechanism. The measured condensation transition in real ^{3}He occurs at about 2.5 mK.

For Lennard-Jones ^{4}He in the unmixed approximation, section 8.4.2, the condensation transition, as measured by the first nonzero occupancy of the momentum ground state, occurred at about $T^* = 0.9$ and $\rho\Lambda^3 = 1.04$ (table 8.2). In the present calculations for Lennard-Jones ^{3}He with $z_0^{\text{ex}} = 1$, it occurs at about $T^* = 0.7$ and $\rho\Lambda^3 = 2.71$. Comparing these two models with identical interaction potential and identical level of approximation (but different masses) shows that condensation is more difficult for fermions than for bosons. And of course this does not take into account the cost of the putative cavity required to accommodate the paired fermions.

Figure 10.7 plots against temperature the dimensionless parameter that combines the arguments of the logarithms in the entropy derivative, equation (10.34). Taking $\rho_{1,s} = \rho/(2S+1)$ is equivalent to setting the fraction of paired fermions to zero. If the series of loop Gaussians were negligible, then the point at which this curve passes through unity would be the condensation transition. It can be seen that for $z_0^{\text{ex}} = 1$ this occurs at about $T^* \approx 0.63$, which is somewhat less than the value $T^* \approx 0.7$ estimated from the previous figure that includes the loop series. One can conclude from this that the permutation loops have a small influence on the condensation transition. Including the excess fugacity $z_0^{\text{ex}} \lesssim 1$, table 10.2, shows that the argument of the logarithm never reaches unity.
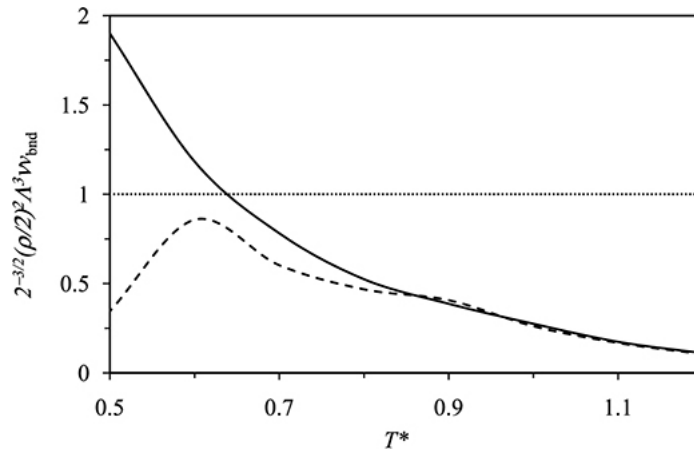
Figure 10.7. The argument of the logarithms in the entropy derivative, equation (10.34) with $f_{1,s} = \rho/(2S+1)$, for saturated liquid Lennard-Jones ^{3}He, $S = 1/2$, as a function of temperature. The solid curve uses $z_0^{\mathrm{ex}} = 1$, and the dashed curve uses z_0^{ex} from table 10.2. The dotted line is a guide to the eye.

Figure 10.8 shows the effect of an applied magnetic field on the optimum fractions of spin-half fermions at $T^* = 0.5$. The excess fugacity has been neglected, $z_0^{\mathrm{ex}} = 1$. With increasing field strength, the number of aligned spins increases and the number of anti-aligned spins decreases, which are to be expected. The number of Cooper pairs decreases, albeit quite slowly, which is no doubt a non-linear affect due to the Boltzmann exponential weight of the magnetic energy. A critical field strength is reached at this temperature at $\beta B_{\mathrm{crit}} = 0.62$. At this point the Cooper pairing is disrupted and the condensation is destroyed.
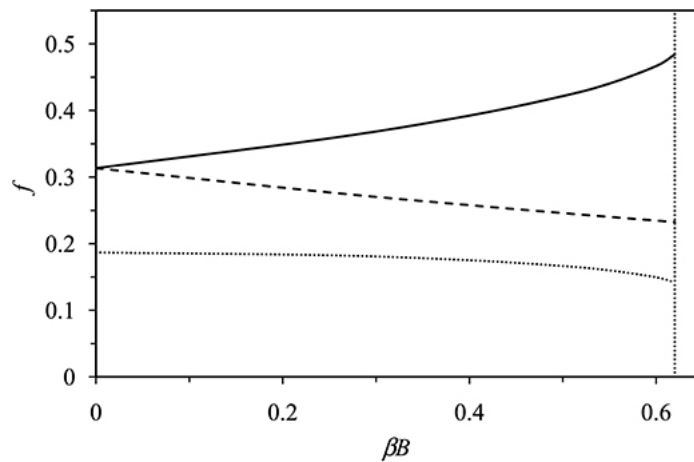

Figure 10.8. Optimum fraction of spin-up, unpaired (solid curve), spin-down unpaired (dashed curve), and paired (dotted curve) spin-1/2 fermions (saturated liquid Lennard-Jones ^{3}He) as a function of magnetic field strength at $T^* = 0.5$ with $z_0^{\mathrm{ex}} = 1$. The

dotted line indicates the critical magnetic field strength, $\beta B_{\text{crit}} = 0.62$, at this temperature.

Figure 10.9 shows the sum of the entropy derivatives as a function of the fraction of unpaired spin-up fermions, equation (10.40). Again the excess fugacity has been neglected, $z_0^{\text{ex}} = 1$. The full equation gives a parabolic curve that crosses zero in two places. The lower fraction is the stable solution and corresponds to the entropy maximum. As the magnetic field is increased, the parabola moves upward (results not shown). Its tip touches zero when the critical magnetic field is reached. For higher fields than this there is no entropy maximum for a physical value of the fraction of unpaired spin-up fermions. This the same as saying condensation into bound fermion pairs does not occur.
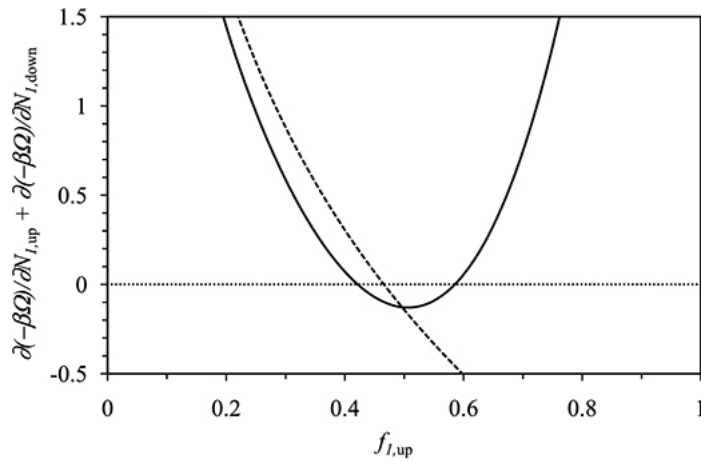

Figure 10.9. Sum of entropy derivatives as a function of the non-optimum fraction of unpaired spin-up fermions (saturated liquid Lennard-Jones ^{3}He, $S = 1/2$) at temperature $T^* = 0.5$ and magnetic field strength $\beta B = 0.5$, with $z_0^{\text{ex}} = 1$. The solid curve is the full equation (10.40), whereas the dashed curve neglects the loop series contribution. The dotted line is a guide to the eye.

The figure also shows the entropy derivative neglecting the contribution from the permutation loop series. The point at which this curve crosses zero is reasonably close to the stable solution predicted by the full equation. However, the simpler equation predicts a monotonic curve, which appears to have a zero for all magnetic field strengths. Hence the criterion for the critical magnetic field neglecting the loop series, equation (10.43), which is based upon $\bar{f}_{1,\uparrow} + \bar{f}_{1,\downarrow} > 1$, is qualitatively different from the criterion used for the full equation, namely the vanishing of the entropy maximum.

This point is confirmed by the data in the penultimate figure, which includes the loop series, where the critical field strength at $T^* = 0.5$ was $\beta B_{\text{crit}} = 0.62$. This compares to the estimate $\beta B_{\text{crit}} = 1.69$ from equation (10.43), which neglects the loop series. The large difference between these two estimates appears to be due to the qualitatively different criterion used rather than to the contribution of the loop series itself.

10.4 Comparison of fermion pairs in BCS and the present theory

10.4.1 Lennard-Jones helium-3

The present calculations for ^{3}He have several quantitative limitations. Obviously the Lennard-Jones pair potential is a crude approximation to the actual interactions in helium. It has limited accuracy for low temperature, high density states because it neglects the exact short-range interactions and many-body contributions. Also, the present analysis neglected the short-ranged commutation function on the grounds that Bose–Einstein condensation is dominated by non-local permutations. Although this works well for bosons, the present results for fermions show the essential role played by the potential of mean force in binding the Cooper pair. Since the minimum in this occurs at short ranges, where the function is rapidly varying, neglecting the commutation function in this case is questionable. The discussion in section 8.6.3 concluded that the commutation function will reduce the curvature of the potential of mean force at its minimum, and reduce the component of momentum parallel to the axis of the pair. The present analysis also suggested that a cavity was necessary to accommodate a long-lived rotating bound fermion pair, but only a simple expression for the excess fugacity associated with this was used, equation (10.14). Finally, the present results are dependent on the binary division approximation, the no-mixing approximation, and retaining only five terms in the loop series for the grand potential.

The simple transition criterion, equation (10.35), which neglects permutation loop series contribution to the grand potential, is

$$\frac{z_0^{\text{ex}\ 2}\nu\rho^2 v_{\text{bnd}}\Lambda^3}{2^{3/2}(2S+1)^2} = 1.$$

(10.45)

Here the unpaired density has been set equal to the total density for that spin, which is correct on the high temperature side of the transition. Of the various parameters that appear, the thermal wavelength and the excess fugacity are generally the most sensitive to temperature. At high temperatures the left-hand side is smaller than unity. At low temperatures the excess fugacity appears to decrease faster with decreasing temperature than the thermal wavelength increases. Of course there is a deal of extrapolation here, but it does suggest that either the condensation transition is non-existent in Lennard-Jones ^{3}He, or else that it occurs at temperatures much lower than those studied here. The measured superfluid transition in real ^{3}He occurs below 3 mK.

10.4.2 Fermion condensation

The present theory for condensation of fermions has generic similarities and differences with BCS theory. What it has in common with BCS theory is the notion of paired fermions with different spin and with net zero momentum. The present theory does go a little beyond this in pointing out the possibility that bosonic permutations can also occur between pairs with the same nonzero net momentum.

In the BCS theory of superconductivity the size of the Cooper pairs is hundreds or thousands of times larger than in the present theory. Since ν, the internal weight of the fermion pair given by equation (10.17), goes to zero for large separations, the above criterion cannot be satisfied by BCS theory. The two theories appear qualitatively different and likely apply in different regimes. The fact that BCS theory is a quantum mechanical rather than a quantum statistical mechanical theory is consistent with it applying quantitatively to low temperature superconductors.

Another significant difference is the mechanism for binding the fermion pair. In BCS theory it is lattice vibrations, and there is a correlation between the frequency and wavelength of the vibration and the speed of the fermions in the pair. In the present theory the pair are bound by a short-ranged minimum in the static pair potential of mean force. In addition it was suggested in section 10.2.2.3, that to satisfy the requirement for the net momentum of the pair, it likely rotates about its center of mass in a cavity in the fluid. The potentials from the fermions that surround the rotating pair help bind them. The cost of forming the cavity, which is reflected in the excess fugacity z_0^{ex}, is unfavorable to pair formation. The cost will be lower for a compressible fluid than is calculated here for incompressible Lennard-Jones ^{3}He.

The present quantum statistical theory arguably applies to high temperature superconductors. As will be shown in the following section 10.5, the pair potential of mean force following the monotonic-oscillatory transition in a highly coupled charged system fits this condensation criterion of possessing a small separation minimum.

There are many similarities between the present treatment of the fermionic superconducting transition and the treatment of the superfluid λ-transition in chapter 8. Ultimately both phenomena are a manifestation of Bose–Einstein condensation, with the paired fermions in the present case being effective bosons. For bosons, condensation into the low-lying momentum states and the consequent superfluidity is driven by the increase in permutation entropy due to the non-local correlations of pure momentum permutation loops. For paired fermions, the permutations between the zero-momentum pairs are similarly non-local. Superfluidity in a bosonic fluid persists because momentum changing collisions can only occur for the state as a whole, and hence they have to be of macroscopic, not of molecular, size (chapter 9). Similarly for superconductivity, because the zero-momentum pairs (or the more general nonzero momentum pairs, section 10.2.2.2) occupy a single state. In both cases it is the permutation entropy that ensures flow without dissipation.

10.5 Monotonic-oscillatory transition in charge fluids

As mentioned in the introduction to this chapter, the long standing BCS (Bardeen *et al* 1957) theory of superconductivity invokes an attractive potential due to lattice vibrations to form the Cooper pairs. This quantum mechanical approach works well for ordinary low temperature superconductivity, but is known to fail for the more recently discovered high temperature superconductors. None of the suggestions for an alternative physical origin for an attractive potential have proved convincing, which tends to suggest that the problem lies with the conventional quantum mechanical formulation of the BCS approach to the superconducting transition. As also pointed out in the introduction, the very name 'high temperature superconductor' suggests that entropy is essential, and that the theory should be formulated in quantum statistical mechanical terms rather than quantum mechanical terms. This in fact is the motivation for, and goal of, the statistical analysis of Bose–Einstein condensation for fermions in the preceding sections in this chapter.

The aim of the present section is to analyse a possible physical mechanism for the attractive potential of mean force that must exist to bind the electrons in the quantum statistical mechanical theory of fermionic condensation developed above. The mechanism was originally suggested by Attard (2022), and the following analysis closely follows that paper.

It is suggested that the attraction between electrons responsible for Cooper pairs in high temperature superconductors is due to the oscillatory pair static correlation function that occurs at high coupling. This has long been established for like-charged particles in the one-component plasma and in primitive model electrolytes (Brush *et al* 1966, Stillinger and Lovett 1968, Fisher and Widom 1969, Stell *et al* 1976, Outhwaite 1978, Parrinello and Tosi 1979, Attard 1993, Ennis *et al* 1995). In the present section the high temperature

superconductor is modeled as a one-component plasma *in media*, with the relatively few electrons in the Fermi mist comprising the fluid charges, and the fixed nuclei and majority immobile electrons in the Fermi sea forming the neutralizing background and static relative permittivity (dielectric constant). The predicted temperatures for the monotonic-oscillatory transition in this model encompass the measured transition temperatures for high temperature superconductors for physically reasonable values of the total electron density and static relative permittivity.

10.5.1 Model and analysis

The solid conductor is modeled as a one-component plasma (mobile electrons in a uniform counter-charge background), together with a finite relative permittivity $\epsilon_{\mathrm{r}} = \mathcal{O}(10^2)$ that results from the remaining immobile but polarizable electrons (i.e. those deep in the Fermi sea). The mobile electrons in the Fermi mist have number density $\rho_{\mathrm{m}}(T)$ derived below. Arguments for and against this model are canvassed in section 10.5.3.

The analysis and specific examples begin with the restricted primitive model electrolyte for three reasons: First, there are a wealth of analytic, numeric, and experimental results known for electrolytes. Second, it shows the generality of the monotonic-oscillatory transition in charge systems. And third, it gives a specific value for the width of the accessible energy states, which is required to determine the electron density of the Fermi mist.

In the restricted primitive model electrolyte (ions of equal hard-sphere diameter in a dielectric continuum), the pair distribution function undergoes an oscillatory transition when (Attard 1993)

$$\kappa_{\mathrm{D}} d \geqslant \sqrt{2},$$

(10.46)

where d is the hard core diameter of the ions. The inverse Debye screening length for the binary symmetric electrolyte is $\kappa_{\mathrm{D}} = \sqrt{(4\pi\beta/\epsilon)2\rho_{\mathrm{F}}q^2}$, where q is the ionic charge (in this case the electron charge), and ρ_{F} is the number density of each type of ion. Here $\beta = 1/k_{\mathrm{B}}T$ is the inverse temperature, and $\epsilon = 4\pi\epsilon_0\epsilon_{\mathrm{r}}$ is the total permittivity of the medium, ϵ_0 being the permittivity of free space (SI units). This result is based on the Debye–Hückel form for the pair distribution function combined with the exact Stillinger–Lovett second moment condition. More accurate analytic and numeric approximations exist (Attard 1993), but this is sufficient for the present purposes.

To make the connection with the one-component plasma, which does not impose a hard core diameter, the distance of closest approach of the electrons can be set as the point at which the Coulomb potential *in media* reaches several times the thermal energy, $u(d) = \alpha k_{\mathrm{B}}T$, or $d = \beta e^2/\epsilon\alpha$. With these the oscillatory transition in the symmetric electrolyte occurs when

$$2 \leqslant \frac{4\pi\beta 2\rho_{\mathrm{F}} e^2}{\epsilon} \frac{\beta^2 e^4}{\epsilon^2 \alpha^2} = \frac{6}{\alpha^2} \Gamma^3.$$

(10.47)

Here the plasma coupling parameter with finite relative permittivity is $\Gamma \equiv \beta e^2/[\epsilon(3/4\pi\rho_{\mathrm{F}})^{1/3}]$.

As mentioned, the parameter α is the multiple of the thermal energy which bounds the accessible states. Choosing $\alpha = \sqrt{24} \approx 4.9$, the transition criterion becomes

$$\Gamma \geqslant 2.$$

(10.48)

With this value of α, the value of the coupling constant at the transition given for the restricted primitive model electrolyte agrees with that found by Monte Carlo simulations of the one-component plasma (Brush *et al* 1966).

Now estimate the density of the electrons in the Fermi mist by modeling them as a non-interacting ideal gas, also known as the free electron model. The Fermi momentum and the Fermi energy for ideal fermions are (Pathria 1972 section 8.1)

$$p_{\mathrm{F}} = 2\pi\hbar \left(\frac{3\rho}{8\pi} \right)^{1/3}, \quad \text{and} \quad \epsilon_{\mathrm{F}} = \frac{(2\pi\hbar)^2}{2m} \left(\frac{3\rho}{8\pi} \right)^{2/3},$$

(10.49)

where $\rho = N/V$ is the total electron number density. The thermal wavelength is $\Lambda = \sqrt{2\pi\beta\hbar^2/m}$, and $\beta\epsilon_{\mathrm{F}} = 2\pi(3\rho\Lambda^3/8\pi)^{2/3}/2$, which is much larger than unity.

With momentum state spacing being $\Delta_p = 2\pi\hbar/L$ (Messiah 1961, Merzbacher 1970), where the volume is $V = L^3$, the number in the Fermi mist is

$$N_{\mathrm{m}} = 2\Delta_p^{-3} \int_{\epsilon_{\mathrm{F}} - \alpha/\beta}^{\epsilon_{\mathrm{F}} + \alpha/\beta} \mathrm{d}\epsilon \, 4\pi m \sqrt{2m\epsilon} \frac{\mathrm{e}^{-\beta(\epsilon - \epsilon_{\mathrm{F}})}}{1 + \mathrm{e}^{-\beta(\epsilon - \epsilon_{\mathrm{F}})}}$$

$$\approx 4\alpha V \Lambda^{-3} (3\rho\Lambda^3/8\pi)^{1/3}.$$

(10.50)

An expansion to leading order for large $\beta\epsilon_{\mathrm{F}}$ has been made to obtain the final equality. That is

$$\rho_{\mathrm{m}} \Lambda^3 = 4\alpha \left(\frac{3\rho\Lambda^3}{8\pi} \right)^{1/3}.$$

(10.51)

The total excitable electron density, ρ_{m}, is significantly less than the total electron density, ρ. It is proportional to the number of accessible energy states about the Fermi energy, which is fixed by equating the results of the restricted primitive model to those of the one-component plasma, $\alpha = \sqrt{24}$.

The idea that only the electrons at the Fermi surface contribute to screening also underlies the Thomas–Fermi model of the electron gas (Kittel 1976). This idea is taken a little further here by modeling the remaining immobile electrons as being polarisable and contributing to the residual dielectric constant.

10.5.2 Numerical results

The predicted oscillatory-monotonic transition temperature is now explored for a range of the two free parameters: the total electron density ρ and the residual static relative permittivity ϵ_{r}. For these a guide is provided by values for ceramic materials. The total electron density of zirconia ZrO_2 is $\rho = 1.65 \times 10^{30}$ m^{-3}. The relative permittivity of typical ceramic insulators is on the order of $\epsilon_{\mathrm{r}} = 10^1$–$10^2$ (www.ceramtec.com/ceramic-materials/dielectric/).

Figure 10.10 shows the coion pair potential of mean force above and below the oscillatory transition close to contact. The results were obtained with the hypernetted chain approximation for the restricted primitive model binary electrolyte (Attard 1993). The qualitative difference that the transition makes is apparent. Increasing coupling corresponds to decreasing temperature at constant concentration, or increasing concentration at constant temperature. The latter is the case shown in the figure.
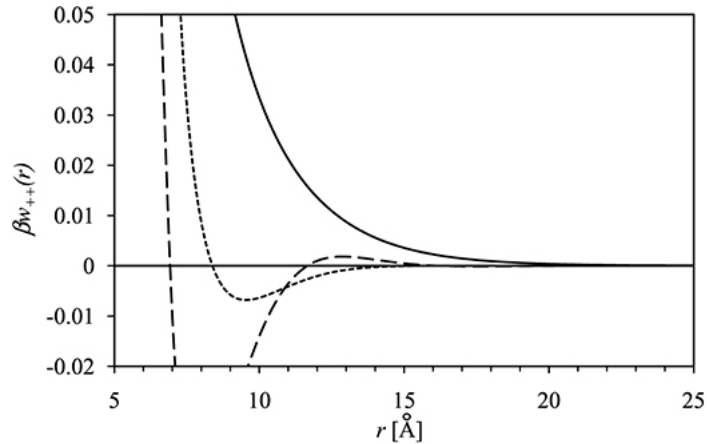
Figure 10.10. Coion pair potential of mean force as a function of separation in the symmetric binary monovalent electrolyte $(d = 3.41\,\text{Å},\ \ \epsilon_r = 100,\ \ T\ =\ 100\ \text{K},$ hypernetted chain approximation). The solid curve is 0.5 M, $(\kappa_D^2 d^2 = 1.5,\ \Gamma = 1.8)$, the short-dash curve is 1.0 M, $(\kappa_D^2 d^2 = 2.9,\ \Gamma = 2.3)$, the long-dash curve is for 2.0 M $(\kappa_D^2 d^2 = 5.9,\ \Gamma = 2.9)$. The solid line is an eye guide.

At the lowest concentration shown 0.5 M, the potential of mean force is monotonic repulsive and exponentially decaying. In this case the coupling is below the transition value predicted for the restricted primitive model and also below that predicted for the one-component plasma. At the intermediate concentration in the figure, 1.0 M, a shallow primary minimum appears with width on the order of 10^3 Å. At the highest concentration shown, 2.0 M, the primary minimum has become relatively deep and narrow, and the potential of mean force is clearly oscillatory with a noticeable barrier to the primary minimum. For the present parameters, the location of the transition predicted by the Debye–Hückel–Stillinger–Lovett approximation for the primitive model electrolyte, equation (10.45), is more or less equal to that observed in the one-component plasma, equation (10.47).

The molecular interpretation of the attraction is that it arises at high coupling from over-charging by counterions combined with packing constraints (Attard 1993). In the one-component plasma, where charge oscillations were first found (Brush *et al* 1966), the oscillation and primary minimum arise from a type of hard-sphere packing effect due to the combination of high density and short-range repulsion.

Figure 10.11 shows the value of the plasma coupling parameter as a function of the relative permittivity for fixed total electron density. Values $\Gamma > 2$ mark oscillatory pair correlation functions. One sees that this occurs for low temperatures. For the lowest relative permittivities shown, $\epsilon_r = 100$ and 75, the transition temperatures are greater than those measured for high temperature superconductors (Bednorz and Möller 1986, Wu *et al* 1987).
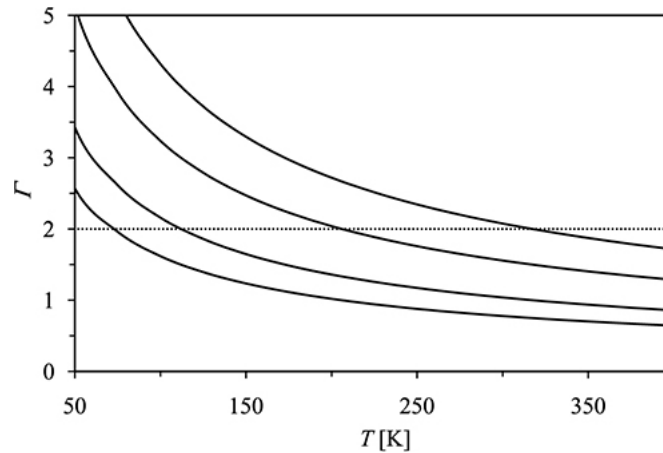
Figure 10.11. Plasma coupling parameter *in media* using electron density $\rho_F(T)$, equation (10.51), and, from bottom to top, relative permittivity $\epsilon_r = 200, 150, 100$, and 75. The dotted line marks the transition to oscillatory behavior. In all cases $\alpha = \sqrt{24}$ and $\rho = 10^{30}$ m^{-3}.

It is arguable that the appearance of the first minimum in the oscillatory pair potential of mean force following the transition is necessary to form bound Cooper pairs (cf the internal weight, equation (10.17)). Whether it is also sufficient is unclear. This raises the question of whether the monotonic-oscillatory transition signifies the actual superconducting transition, or whether it represents an lower bound for the coupling necessary for the superconducting transition (i.e. is it necessary and sufficient?, or is it only necessary?). In addition, one should be aware that the monotonic-oscillatory transition in a layered dielectric material where the coupling is enhanced may occur at higher temperatures than in a uniform dielectric continuum. Figure 10.12 shows the transition temperature as a function of relative permittivity for several values of the electron density. The transition temperature increases with decreasing relative permittivity and with increasing electron density. It is more sensitive to changes in the permittivity than the electron density. The range of calculated transition temperatures encompasses those measured for high temperature superconductors. From the figure one can conclude that a high superconducting transition temperature requires both a high electron density and a low static relative permittivity. Since the latter can be expected to increase with increasing electron density, there is an obvious competition between these two requirements. This undoubtedly restricts the structure and composition of suitable candidate materials for high temperature superconductivity.
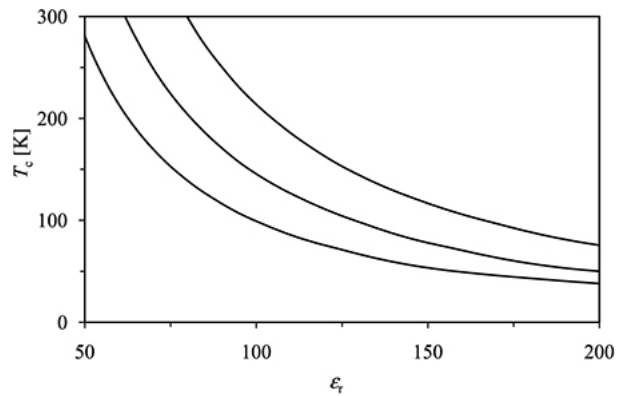
Figure 10.12. Transition temperature as a function of the relative permittivity. From left to right the curves are for a total electron density of $\rho = 0.1$, 1, and 10 $\times 10^{30}$ m^{-3}.

10.5.3 Discussion of the statistical binding of electron pairs

One unusual aspect of the proposed model is the invocation of a finite relative permittivity for the solid. The perceived wisdom is that a conductor has infinite permittivity. Conversely, the one-component plasma is usually modeled *in vacuo*, $\epsilon_r = 1$. Using a relative permittivity in between these two extremes has several justifications. First, one has to distinguish between an experimental challenge and a fundamental limitation of nature. There is no doubt that a macroscopic measurement of the dielectric constant of a conductor yields an infinite dielectric constant in the zero frequency limit. Arguably, however, all this says is that the conductivity dominates the measurement; it does not say that the residual relative permittivity is either unity or infinity. Second, fixed atoms surrounded by the electrons deep in the Fermi sea, which include the immobile inner shell electrons, remain polarizable, and therefore they must contribute to a finite relative permittivity. Third, if the relative permittivity were truly infinite at the molecular level, then the immobile electrons could not interact via the Coulomb repulsion, and they would be utterly transparent to each other, which is obvious nonsense. Fourth, and finally, insulators that are close in chemical composition and physical structure to specific high temperature superconductors, have a measured finite relative permittivity greater than unity. For example, cuprate superconductors are insulators if the doping fraction is less than 0.1 holes per CuO$_2$ (http://hyperphysics.phy-astr.gsu.edu/hbase/Solids/hitc.html). It seems plausible that the small changes in composition that turn these into conductors do not much change the residual static relative permittivity. Instead, the simpler interpretation is that the infinite conduction permittivity swamps any attempt to measure macroscopically the finite residual static relative permittivity.

The most reliable way to measure the residual static dielectric permittivity of a high temperature superconductor may turn out to be by extrapolation from an insulator of similar chemical and structural composition.

As mentioned in connection with figure 10.12, a high monotonic-oscillatory transition temperature relies upon the competing requirements of high electron density and low relative permittivity. One might speculate that one of the reasons for the prevalence of layered structures amongst high temperature superconductors is that they combine a high electron density within the conducting planes together with a low electron density and hence low polarisability in the interlayer space. At the level of dielectric continuua, the

consequent image charges increase the coupling between the electrons in the dense layer (Attard 1988) and increase the temperature of the oscillatory transition compared to a uniform dielectric medium.

The quantitative accuracy of modeling a crystalline layered conductor as a homogeneous charge fluid is a little unclear. On the one hand the universality of the monotonic-oscillatory transition for the like-charge pair potential of mean force at high coupling is robust and undeniable. On the other hand, the quantitative prediction of the transition temperature and the depth of the primary minimum cannot be taken too literally for electrons in the layered crystalline solids that are of interest in high temperature superconductivity.

In the preceding sections on ^{3}He, it was suggested that long-lived bound fermion pairs rotate in cavities spontaneously formed in the fluid, section 10.2.2.3. The cost of cavity formation was calculated for an incompressible fluid, equation (10.14), and it appears to be directly responsible for suppressing condensation in Lennard-Jones ^{3}He, figure 10.6.

In the present case of high temperature superconductors, the density of excitable electrons in the Fermi mist, ρ_{m}, is relatively low, and so the charge fluid that they comprise may be more compressible than ^{3}He. Hence the cost of cavity formation may not be prohibitive to condensation in this case. It is unclear whether a cavity is essential at low densities, or, if so, whether it would inhibit superconducting flow of the contained paired fermions. In the one-component plasma model with uniform background charge, a fluctuation that locally depletes the electron density provides an excess positive charge center about which a binary pair of electrons might orbit.

Alternatively, one might consider that the continuum model of the atoms and immobile electrons is reasonably accurate, with the layers of the crystal being modeled as a net uniform positive charge density. Such an infinite plane of uniform positive charge gives rise to a linearly decreasing electrostatic potential beyond it. Driven by entropy some or most of the fluid electrons spill over into the interlayer space, creating a one-component plasma in a region of low relative permittivity, and a diffuse electric double layer at the interface. These are the conditions for the oscillatory transition to occur. As argued here, this leads to the subsequent formation of electron pairs bound by the primary minimum in the potential of mean force. Since the interlayer space is dilute or devoid of atoms and fixed electrons, the rotation of the pair is relatively unimpeded and there is no need to invoke a cavity.

On the basis of the results in this section it is plausible that the oscillatory potential of mean force that occurs at high coupling is responsible for Cooper pair formation in high temperature superconductors. The BCS theory is predicated upon an attractive potential, not an attractive potential of mean force. The difference between the pair potential and the pair potential of mean force is fundamentally the difference between quantum mechanics and quantum statistical mechanics. Whereas the minimum in the binding potential occurs at hundreds of nanometers separation in BCS theory, the minimum in the potential of mean force must occur at separations on the order of a nanometer in the present statistical theory. Quantum statistical mechanics, not quantum mechanics, is the appropriate theory for condensed matter. In the preceding sections of this chapter the quantum statistical mechanical theory for superconductivity was developed. It shows explicitly how the potential of mean force creates Cooper pairs and how they subsequently condense. On the basis of those and the present results one can rationally argue that the first minimum in an oscillatory electron–electron potential of mean force is sufficient to form bound fermion pairs. It follows that the monotonic-oscillatory transition that occurs generally in charge fluids signifies the superconducting transition in high temperature superconductors.

Summary

- Fermions may form Cooper pairs, the members of which have different spin and equal and opposite momentum, making an effective boson in the ground momentum state.

More general nonzero momentum pairs may be considered that lead to condensation into multiple low-lying momentum states.

- In quantum statistical mechanics, the effective boson has an internal weight that is close to unity if the pair potential of mean force has a highly curved minimum at small separations. The bound fermion pair rotates about it center of mass, with each fermion having equal and opposite momentum in the binary division picture, in a spontaneously formed cavity that helps to stabilize the pair. Bose–Einstein condensation is determined by the balance between the cost of cavity formation and the gain in non-local permutation entropy.
- The loop expansion for quantum statistical mechanics in classical phase space appears less efficient for fermions than for bosons.
- High temperature superconductors should be treated with quantum statistical mechanics rather than with quantum mechanics. The monotonic-oscillatory transition in highly coupled charge fluids provides an attractive candidate for the binding potential of mean force for Cooper pair formation. For realistic values of the electron density and relative permittivity the predicted temperature for the transition is similar to the measured superconducting transition temperature.

References

Anderson P 1987 The resonating valence bond state in La_2CuO_4 and superconductivity *Science* **235** 1196

Attard P, Mitchell D J and Ninham B W 1988 Beyond Poisson–Boltzmann: images and correlations in the electric double layer. II. Symmetric electrolyte *J. Chem. Phys.* **89** 4358

Attard P 1993 Asymptotic analysis of primitive model electrolytes and the electrical double layer *Phys. Rev.* E **48** 3604

Attard P 2018 Quantum statistical mechanics in classical phase space. Expressions for the multi-particle density, the average energy, and the virial pressure arXiv:1811.00730

Attard P 2021 *Quantum Statistical Mechanics in Classical Phase Space* (Bristol: IOP Publishing)

Attard P 2022 Attraction between electron pairs in high temperature superconductors arXiv:2203.02598

Bardeen J, Cooper L N and Schrieffer J R 1957 Theory of superconductivity *Phys. Rev.* **108** 1175

Bednorz J G and Möller K A 1986 Possible high T_c superconductivity in the Ba–La–Cu–O system *Z. Phys.* B **64** 189

Bickers N E, Scalapino D J and Scalettar R T 1987 CDW and SDW mediated pairing interactions *Int. J. Mod. Phys.* B **1** 687

Brush S G, Sahlin H L and Teller E 1966 Monte Carlo study of a one-component plasma. I. *J. Chem. Phys.* **45** 2102

Cooper L N 1956 Bound electron pairs in a degenerate Fermi gas *Phys. Rev.* **104** 1189

Ennis J, Kjellander R and Mitchell D J 1995 Dressed ion theory for bulk symmetric electrolytes in the restricted primitive model *J. Chem. Phys.* **102** 975

Fisher M E and Widom B 1969 Decay of correlations in linear systems *J. Chem. Phys.* **50** 3756

Gros C, Poilblanc D, Rice T M and Zhang F C 1988 Superconductivity in correlated wavefunctions *Physica* C **153-155** 543

Inui M, Doniach S, Hirschfeld P J, Ruckenstein A E, Zhao Z, Yang Q, Ni Y and Liu G 1988 Coexistence of antiferromagnetism and superconductivity in a mean-field theory of high-Tc superconductors *Phys. Rev.* B **37** 5182

Kipling R 1918 *Many Inventions* (London: Macmillan)

Kittel C 1976 *Introduction to Solid State Physics* 5th edn (New York: Wiley)

Kotliar G and Liu J 1988 Superexchange mechanism and d-wave superconductivity *Phys. Rev.* B **38** 5142

London F 1938 The λ-phenomenon of liquid helium and the Bose-Einstein degeneracy *Nature* **141** 643

Mann A 2011 High-temperature superconductivity at 25: Still in suspense *Nature* **475** 280

Maxwell E 1950 Isotope effect in the superconductivity of mercury *Phys. Rev.* **78** 477

Messiah A 1961 *Quantum Mechanics* **vols I and II** (Amsterdam: North-Holland)

Merzbacher E 1970 *Quantum Mechanics* 2nd edn (New York: Wiley)

Monthoux P, Balatsky A V and Pines D 1991 Toward a theory of high-temperature superconductivity in the antiferromagnetically correlated cuprate oxides *Phys. Rev. Lett.* **67** 3448

Outhwaite C W 1978 Modified Poisson–Boltzmann equation in electric double layer theory based on the Bogoliubov–Born–Green–Yvon integral equations *J. Chem. Soc. Faraday Trans. 2* **74** 1214

Parrinello M and Tosi M P 1979 Structure and dynamics of simple ionic liquids *Rev. Nuovo Cimento* **2** 1

Pathria R K 1972 *Statistical Mechanics* (Oxford: Pergamon)

Reynolds C A, Serin B, Wright W H and Nesbitt L B 1950 Superconductivity of isotopes of mercury *Phys. Rev.* **78** 487

Stell G, Wu K C and Larsen B 1976 Critical point in a fluid of charged hard spheres *Phys. Rev. Lett.* **37** 1369

Stillinger F H and Lovett R 1968 Ion-pair theory of concentrated electrolytes. I. Basic concepts *J. Chem. Phys.* **48** 3858

van Sciver S W 2012 *Helium Cryogenics* 2nd edn (New York: Springer)

Wu M K, Ashburn J R, Torng C J, Hor P H, Meng R L, Gao L, Huang Z J, Wang Y Q and Chu C W 1987 Superconductivity at 93 K in a new mixed-phase Y–Ba–Cu–O compound system at ambient pressure *Phys. Rev. Lett.* **58** 908