Tesis Doctoral

2014

The Theory of Coarse-Graining Without Projection Operators

Marc Meléndez Schofield

Máster Universitario en Física de Sistemas Complejos

Programa de Doctorado en Ciencias

Director: Pep Español Garrigós

Doctoral Dissertation at the Department of Fundamental Physics in the Faculty of Science, UNED.

Title: The Theory of Coarse-Graining Without Projection Operators.
Author: Marc Meléndez Schofield, M.Sc. in Physics of Complex Systems.
Director: Dr Pep Español Garrigós.

Escuela de Doctorado de la UNED. Tesis doctoral presentada en el Departamento de Física Fundamental, Facultad de Ciencias.
Título: The Theory of Coarse-Graining Without Projection Operators (La teoría del coarse-graining sin operadores de proyección).
Autor: Marc Meléndez Schofield, Máster Universitario en Física de Sistemas Complejos.
Director: Dr. Pep Español Garrigós.

Marc Meléndez Schofield (2014).

Acknowledgements

A thesis rarely comes into being without the help of many people. First, I would like to thank my advisor, Pep Español, for his patience. I do not think he realises how much the discussion we had in our first meeting shaped all the work that I have carried out since. I would also like to thank my colleagues and friends at the university. They often lent me a hand, especially with keeping spirits high. William G. Hoover read some sections in Chapter 4 and made many helpful comments, for which I am very grateful. Finally, I must thank my family, to whom I dedicate these pages. My mother braved through the whole text hunting for typographical errors and suggesting many improvements. Thank you very much, mother. Foremost, I would like to say thank you to Emma, who patiently accepted that I had little time to play with her during the summer, and thank you to Ana: without your constant support and encouragement I would have never written this book.

To my family, especially Ana,
Emma and Eric.—This is not poetry,
but it does praise beauty.

Abstract

The theory of coarse-graining, or nonequilibrium statistical mechanics, developed by Onsager, Green, Zwanzig and many others faces three challenges, which we will refer to as the problems of *sampling*, *projected dynamics* and *memory*. This thesis dissertation addresses all three and suggests new approaches to tackle them.

Here we present the theory of coarse-graining in the general framework of Markov processes. Assuming that we do not have access to the exact state of the stochastic process and that we can only determine the values of several functions of the state, the theory of coarse-graining infers the equations for the time evolution of the average values of these functions. We refer to the averages as the macroscopic variables. A given set of values of these variables determines a macrostate, which will in general correspond to many possible states in the Markov process. Even with memoryless reversible laws, as soon as we group the states into macrostates and consider transitions among them, we find that the resulting process often displays memory effects and irreversibility.

The coarse-grained equations for the nonequilibrium evolution of macroscopic variables include coefficients that are usually calculated from time correlation functions obtained from molecular dynamics simulations, which sample the microstates accessible to a system by following its trajectory in phase space. To make the sampled microstates representative, the numerical trajectories must cover sufficiently large time intervals. Because the coefficients depend on the macroscopic state, we must carry out simulations for each possible combination of the macroscopic variables. Therefore, the

required numerical work often lies beyond the reach of present day computing power. While we do not solve this problem completely, we propose the variational principle of maximum relative entropy as a distinct way to derive ensembles that can extend the results of simulations to nearby values of the macroscopic variables (Chapter 2).

In Zwanzig's equations, unfortunately, the time evolution operator appearing in the time correlation functions corresponds to a projection of the dynamics onto a relevant subset of phase space, instead of the real evolution defined by the microscopic dynamics, which forces us to use approximations to the projected trajectoriess in terms of the real dynamics. Chapter 3 explains how to derive exact equations for the macroscopic variables that depend directly on the real dynamics. This simplifies the analytical work and the connection with simulation results, and it constitutes one of the central results of this thesis.

Coarse-grained equations of motion become memoryless partial differential equations when there exists a separation between the characteristic times of molecular motion and significant changes in the relevant variables. When no such separation can be found, the description must incorporate the memory kernels. The final chapter discusses how memory effects relate to time scales. When we contemplate time extending indefinitely into the future, memory effects eventually become irrelevant. However, the properties we derive from that perspective may not adequately portray the behaviour of a system over the time scales of interest. In this context, we argue that the recently proposed logarithmic thermostat is inefficient because its memoryless thermodynamic properties apply to time scales much larger than the intervals over which it is supposed to thermostat a system. We end with some comments on how to estimate the relevance of memory effects.

Resumen

La teoría del *coarse-graining*, o mecánica estadística fuera del equilibrio, desarrollada por Onsager, Green, Zwanzig y muchos otros se enfrenta a tres retos, que denominaremos los problemas del *muestreo*, de la *dinámica proyectada* y de la *memoria*.

Aquí presentamos la teoría del coarse-graining desde el marco general de los procesos de Markov. Suponiendo que no tengamos acceso al estado concreto de un proceso estocástico y que solo podemos determinar los valores de varias funciones del estado, la teoría del coarse-graining infiere las ecuaciones para la evolución temporal de los valores medios de estas funciones. Llamamos a los valores medios variables macroscópicas. Un conjunto de valores de estas variables determina un macroestado, que en general es compatible con muchos estados del proceso de Markov. Incluso con leyes reversibles y sin memoria, en cuanto agrupamos los estados por macroestados y consideramos las transiciones entre estos, encontramos que el proceso resultante a menudo exhibe efectos de memoria e irreversibilidad.

Las ecuaciones que describen la evolución fuera del equilibrio de las variables macroscópicas incluyen coeficientes que normalmente se calculan a partir de correlaciones temporales obtenidas a partir de simulaciones de dinámica molecular, que muestrea los estados accesibles al sistema siguiendo su trayectoria en el espacio de fases. Para que los microestados conformen una muestra representativa, las trayectorias numéricas deben cubrir intervalos temporales suficientemente largos. Como los coeficientes dependen del estado macroscópico, debemos llevar a cabo simulaciones para cada combinación posible de las variables macroscópicas. Por lo tanto, el trabajo

numérico frecuentemente queda más allá de nuestras potencia de cálculo actual. Aunque no resolvemos este problema completamente, proponemos el principio variacional de máxima entropía relativa como una forma alternativa de derivar colectividades, que permiten extender los resultados de una simulación a valores cercanos de las variables macroscópicas (Capítulo 2).

En las ecuaciones de Zwanzig desgraciadamente aparece un operador de evolución temporal que corresponde a la proyección de la dinámica sobre un subconjunto relevante del espacio de fases, en lugar de la evolución real definida por la dinámica microscópica. Este hecho obliga a utilizar aproximaciones para describir la dinámica proyectada en términos de la real. El tercer capítulo explica cómo derivar ecuaciones exactas para las variables macroscópicas que dependen directamente de la dinámica real. Esto simplifica el trabajo analítico y la conexión con los resultados de las simulaciones numéricas, y constituye uno de los resultados centrales de esta tesis.

Las ecuaciones para las variables macroscópicas se convierten en ecuaciones diferenciales en derivadas parciales sin memoria cuando existe una separación entre los tiempos característicos del movimiento de las moléculas y los cambios en las variables relevantes. Cuando no se puede encontrar tal separación, la descripción debe incorporar núcleos de memoria. El capítulo final discute la conexión entre los efectos de memoria y las escalas temporales. Cuando contemplamos intervalos temporales que se extienden indefinidamente, los efectos de memoria se acaban volviendo irrelevantes. Sin embargo, las propiedades que calculamos desde esta perspectiva no siempre retratan el comportamiento del sistema en las escalas de tiempo que nos interesan. En este contexto, argumentamos que el termostato logarítmico propuesto recientemente es ineficiente debido a que sus propiedades termodinámicas se aplican a escalas temporales mucho más grandes que aquellas sobre las que se supone que debe actuar como termostato. Terminamos con comentarios sobre cómo estimar la importancia de los efectos de memoria.

Contents

Introduction

This thesis presents the theory of nonequilibrium statistical mechanics, or coarse-graining. We will see how to derive dynamical equations for the evolution of macroscopic variables from the underlying microscopic laws. Our results differ from Zwanzig's in that our exact equations depend directly on the dynamical trajectories, instead of their projection onto a relevant subspace.

We present the theory of coarse-graining from the general framework of Markov processes, explaining how memory and irreversibility emerge from memoryless reversible laws. We propose a way to measure the irreversibility of a process and show that perfectly reversible simulations give rise to irreversible effects. We have also included comments on how to extend the results of sampling and how to estimate the relevance of memory effects in terms of dissipation.

About a decade ago, I watched Àlex Pastor's short film *La Ruta Natural* (*The Natural Route*), which imagined life in reverse motion. The palindromic title mirrored the structure of the story. Divad lay dying in the opening scene and we heard him narrate his life all the way back to his birth in an upside-down but meaningful biography. The fascinating and rather sad reverse account did indeed resemble the tale in the opposite direction, but what a strange world it presented! Divad loved wars, with their enormous power to bring people to life and quickly build cities, and he hated the slow mindless destruction he carried out in his job; he liked

rubbish, the source of many useful things, and disliked the money he got in exchange for giving his possessions away.

Inverting the sequence of events transforms familiar occurrences into alien phenomena. As a storytelling device, it has appeared in music videos, like Coldplay's *The Scientist* (2002), in films, such as *Irreversible* (2002), *Memento* (2000) and *Happy End* (1966), as well as in short stories and novels: Damon Knight's *This way to the Regress* (1956) or H. G. Wells's famous *The time machine* (1895), for example.

The asymmetric relation between the past and the future caught the attention of great 19th century scientists before it was picked up by artists. The major breakthrough in the comprehension of irreversibility came from the works of Rudolf Clausius, Josiah Willard Gibbs and Ludwig Boltzmann between 1862 and 1877. The first stories dealing with an inverted time flow were published about ten years later. Among them, we find a chapter in Lewis Carroll's *Sylvie and Bruno* [1], where an outlandish watch could make the events of the next hour happen in reverse order when the "reversal peg" was pushed in.

Impossible as this may seem, we do not really need any supernatural manipulation of the fundamental laws of nature to observe a reverse sequence of events. Hamiltonian equations behave like a palindrome with respect to time. Flipping the direction of time does not have any effect on them. In theory, we only need to reverse all the velocities to make a process retrace its steps.

Loschmidt pointed precisely to this symmetry in his friendly criticism of Boltzmann's H theorem [2]. Boltzmann had applied statistical methods to the analysis of collisions between gas molecules in order to infer the second law of thermodynamics [3], but Loschmidt argued that any trajectory along which the entropy increased could be reversed to yield a new process with decreasing entropy. The laws of mechanics accept both motions as valid possibilities.

Boltzmann responded by noting that there are many more high-entropy states in phase space than low-entropy states. Therefore, if we pick an initial state at random, the overwhelming majority of possibilities will lead us along a path of increasing entropy [4], no matter what direction of time

we choose. A convincing explanation, although, as Lebowitz remarked, Boltzmann was left with the "nagging problem" of how the universe had gotten into a state of such low entropy in the past [5].

Apart from introducing the notation and principal assumptions, the first chapter of this dissertation is devoted to general aspects of the theory of coarse-graining. Concerning Loschmidt's paradox, it explains how macroscopic irreversibility and memory effects emerge from memoryless reversible laws. Our presentation simply restates Boltzmann's main point in Ref. [4], but we concentrate on how to quantify irreversibility in terms of the probabilities of direct and reverse trajectories connecting macroscopic states. In the section on molecular dynamics, I develop my own version of Carroll's "reversal peg": a memoryless symplectic bit-reversible integration scheme. With it, we demonstrate that genuine Markovian time-reversible laws (a sequence of canonical transformations) give rise to asymmetries between the past and the future. I came up with the algorithm inspired by similar work by Levesque and Verlet, who found a similar bit-reversible recipe which was neither symplectic nor memoryless.

Chapter 1 also contains a section on Jaynes's maximum entropy formalism. We cannot measure the exact states of all the microscopic components of a system with infinite precision, so we need some way to quantify our ignorance about them. By maximising entropy we can establish, from macroscopic information, the likelihood of different atomic states. Chapter 2 examines an alternative variational principle from which to infer probability distributions: *the relative entropy route*. Our contribution to this area of research consists in proving that under certain conditions the latter method yields the same answer as Jaynes's. We claim that relative entropy simplifies some calculations, in addition to improving sampling results when the maximum entropy approach performs poorly due to nonergodic behaviour.

Sampling constitutes one of the most pressing problems in present day nonequilibrium statistical mechanics. The transport coefficients appearing in the equations depend on the whole set of parameters that determines the thermodynamic state. Viewed on the molecular time scale, thermodynamic processes proceed too slowly to simulate with molecular dynamics. In some cases, we can use equilibrium molecular dynamics or Monte Carlo

algorithms to calculate the desired coefficients, but we have to carry out a numerical experiment for each possible combination of the parameters, leading to an exponential explosion in the number of simulations for just a few parameters. In response to this problem we propose a way to extend the results of one simulation to nearby values of the parameters.

Furthermore, the second chapter explains how to derive a few generalised theorems in thermodynamics and nonequilibrium statistical mechanics from the measure of irreversibility introduced in Chapter 1. We will prove the Jarzynski equality, the work theorem and the relation between dissipated work and relative entropy. We will also come across a very simple proof of the Second law of thermodynamics based on S. Vaikuntanathan and C. Jarzynski's link between relative entropy, and dissipation and lag. Although these results are not new, they appear as relatively simple consequences of our assumptions and definitions. We also present some generalisations.

The first two chapters deal with transitions between macrostates when the probability densities correspond to relevant ensembles, without paying attention to how probability distributions change in time. Chapter 3 explains how to deduce the irreversible laws of nonequilibrium thermodynamics from underlying reversible Hamiltonian or thermostated dynamics. There we rewrite Liouville's equation for the evolution of the probability density, or its generalisation for compressible flows in phase space, in terms of the macroscopic variables. We illustrate the method by working out macroscopic equations for systems equilibrating and conclude with the heat equation as a limiting approximation.

Robert Zwanzig developed a similar programme in the early 1960s with the help of projection operators. Appendix A explains the main idea behind Zwanzig's technique and adds some comments on how to recast the theory presented here in terms of projection operators. While the absence of projected operators might sound interesting from the pedagogical point of view, the main interest of my contribution lies in the absence of projected *dynamics* in the expressions. In contrast to Zwanzig's results, which depend on the trajectory of the system projected onto a relevant subset of phase space, we will discover equations that depend directly on the real dy-

namics. This simplifies the analytical work and the connection to molecular dynamics results, though we have to add terms to the dynamical equations.

The final chapter begins with alternative definitions of temperature in the context of nonequilibrium states. It then explains how to transform the canonical equations of motion into the deterministic time-reversible expressions of thermostated dynamics. After a plausible reconstruction of the thought process that uncovered Nosé's canonical dynamics and an explanation of Nosé-Hoover equations, we plunge into our heated argument with Campisi *et al.*, which ended up on the pages of Physical Review Letters. The discussion concerned the role of the logarithmic oscillator as a Hamiltonian thermostat. Apart from other technical problems, we will show that the exponential growth of time and length scales with the energy makes the logarithmic oscillator very ineffective as a thermostat.

Using heat transfer as a reference point, the last chapter discusses the role of memory in the equations for macroscopic variables. We discovered that Fermi-Pasta-Ulam-Tsingou chains do not approach equilibrium in the same way as the memoryless hard-sphere gas in Chapter 3, because of a subtle form of macroscopic work. We end with some comments on how to estimate the relevance of memory effects.

All chapters contain a brief nontechnical introduction and summaries at the end, for convenience. Appendix B lists the scientific contributions (publications, talks and posters) that I have worked on for the last four years, while I investigated the theory of coarse-graining.

Scientific truth feels very different from revelation. It comes suddenly in flashes of insight and then wanes. In *Out of Oz* [6], Gregory Maguire wrote that

> There is a reason we live in time. We are too small a flask, even as an Elephant, to tolerate too much knowing. Instead, truth must dip through us as through a pipette, to allow only moments of apprehension. Moments diffuse and miniature enough to be survived.

I hope these pages afford the reader some of the moments of appre-

hension that I have felt while doing the research. This is a beautiful and fascinating field, and I will try to keep that it mind as I make a cup of tea and sit down to write.

Chapter 1

Memoryless Dynamics

Herein we propose to measure irreversibility as a ratio of densities of states. We develop a symplectic bit-reversible Markovian numerical integration scheme inspired by Levesque and Verlet's work and use it to show that irreversibility emerges from reversible Hamiltonian laws.

There was once a time, now long gone, when physicists considered it fashionable to begin their treatises with metaphysical postulates. I fear that scientific literature has changed so much that nowadays nothing short of an apology justifies a few paragraphs on philosophy. Nevertheless, I propose to begin this chapter precisely by examining the implications of a widespread metaphysical assumption, on the grounds that leaving it unstated does not make it any less present.

The assumption in question underlies most, if not all, of modern physics, and expresses the idea that history may alter the future only through the traces that it has left in the present. Were it possible to scrub out the effects of past events completely, then they would never have any bearing on the future. They would be neither felt nor remembered, just as if they had never taken place. The same thought may be phrased in words familiar to physicists by saying that the future evolution of a system depends only

on its present state.

Physicists did not establish this memoryless character of the fundamental laws of nature empirically. The concept probably came from finding the time analogue of action at a distance unpalatable. How could a phenomenon fail to affect the near future but then be felt later on? Chains of cause and effect seem *a priori* to be continuous in time. We believe that even when effects lag behind their cause (a phenomenon known as *hysteresis*) some deeper level of detail will reveal memoryless laws that do not refer to the history of the process. Despite its metaphysical soundness, however, our "principle of memorylessness" should be accepted only insofar as it provides good explanations and does not contradict empirical data.

Now, with laws that forget previous states instantly, the history of a process would never affect its subsequent evolution unless memory of the past were stored in the present in some way. Picture a man standing by a river. Yesterday he dived in without thinking but today he hesitates. The setting is seemingly the same, but the outcome is different, though surely we do not need Heraclitus to remind us that both man and river have changed. Our crude description of the scene left out a myriad of tiny details that make yesterday's situation differ from today's. If the man were touched by the exact same emotions experienced yesterday, if he could go through the same thoughts and memories, if he saw before him the same calm and shining waters, would he not feel impelled to dive in once again? Perhaps he would, but how could so many circumstances ever coalesce again?

Similarly, thermodynamic states (or *macrostates*) of a system of interest (the volume, temperature and density, for example) leave out most of the information that would be needed to specify the corresponding *microstates*, that is, the precise positions and velocities of all the component atoms. The unseen details that make up what appears to us as the same final macrostate record the fact that we arrived by a different route, and even slight changes in the atomic configuration can lead to different outcomes later in time. Therefore, *the main problem for nonequilibrium statistical mechanics is that the behaviour of a system depends on its precise microstate, which we can almost never determine in practice.*

1.1 Markov processes

Memoryless stochastic processes give rise to memory effects when we group states into sets of states, which we call *macrostates*, and consider the probabilities of transitions between the macrostates.

Let us rephrase our metaphysical assumption as a mathematical statement. Suppose we are interested in a system passing through microstates $z_1,\ z_2,\ \ldots,\ z_n$ at the corresponding times $t_1,\ t_2,\ \ldots,\ t_n$. Given this information, we would like to know the conditional probability of finding the system in state z_{n+1} at time t_{n+1}, which we will denote $p(z_{n+1},\ t_{n+1} \mid z_n,\ t_n;\ \ldots;\ z_2,\ t_2;\ z_1,\ t_1)$. The "principle of memorylessness" declares that whatever happened before time t_n is irrelevant, so that

$$p(z_{n+1},\ t_{n+1} \mid z_n,\ t_n;\ \ldots;\ z_2,\ t_2;\ z_1,\ t_1) = p(z_{n+1},\ t_{n+1} \mid z_n,\ t_n). \quad (1.1)$$

In other words, once we have found out the transition probability from z_n at t_n to z_{n+1} at t_{n+1}, there is nothing else to know, and specifying the history of the system before time t_n does not alter the probability of finding it at z_{n+1} at time t_{n+1}. In mathematical parlance, a family of trajectories $z_1,\ z_2,\ \ldots,\ z_n$ with probabilities that satisfy equation (1.1) is known as a Markov process. Note that deterministic trajectories such as those in Hamiltonian mechanics count as particular cases of Markov processes in which the conditional probabilities vanish, except when the Hamiltonian evolution transforms z_1 into z_n.

We will find it convenient to denote the conditional probabilities in (1.1) by

$$p_{ji}(t_n) \equiv p(z_j,\ t_{n+1} \mid z_i,\ t_n), \quad (1.2)$$

and we will call $p_{ji}(t_n)$ the *transition probability* from state i to state j at time t_n. If we represent the probability of state z_i at time t_n by $\rho_{t_n}(z_i)$, then the probabilities at time t_{n+1} result from multiplying a matrix of transition probabilities by a vector of microstate probabilities,

$$\rho_{t_{n+1}}(z_j) = \sum_i p_{ji}(t_n)\ \rho_{t_n}(z_i). \quad (1.3)$$

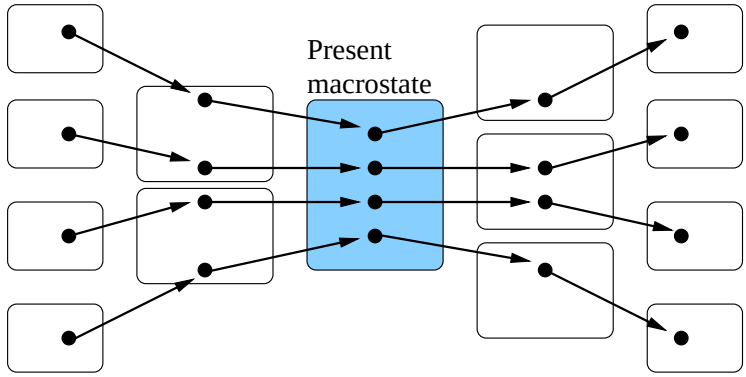

Figure 1.1: Deterministic trajectories. Microstates are represented by dots, transitions by arrows and macrostates by boxes. Even though each microstate determines its sequent completely, we cannot tell which macrostate the system will occupy next from the present macrostate. In the case illustrated here, we would also need the previous two macrostates to infer the next step.

For continuous sets of states, the sum should be replaced by an integral.

Within a Markovian framework, memory effects appear as soon as we start leaving out some of the details. Suppose that we can only give a *coarse-grained* description of a system, that is, we can determine the value of some function F of the microstates but not the exact microstate occupied by the system. We define the macrostate x_f as the set of microstates compatible with $F(z) = f$,

$$x_f = \{z : \ F(z) = f\}. \tag{1.4}$$

Now consider the trajectories sketched in figure 1. Any state determines the next step at the microscopic level. However, knowing the present macrostate will not suffice to predict the next macrostate unless we also have the previous two.

Surprisingly, the probability of a transition from x_f to $x_{f'}$ cannot generally be calculated exclusively from the set of microscopic transition prob-

abilities p_{ij}. To prove this fact, we write the probability of the macrostate x_f at time t in terms of ρ, that is,

$$P_F(f;\ t) = P(x_f;\ t) = \sum_{z_i \in x_f} \rho_t(z_i), \tag{1.5}$$

for discrete sets of microstates, or

$$P_F(f;\ t) = P(x_f;\ t) = \int \rho_t(z)\delta(F(z) - f)\ dz, \tag{1.6}$$

for continuous cases. When the limits of integration are omitted, as above, we assume that integrals extend over the whole domain of their variables. The Dirac delta function δ ensures that the integral is carried out over the set of microstates z that satisfy $F(z) = z$.

Let us first calculate the transition probabilities between macroscopic states for the discrete scenario. The probability of occupying $x_{f'}$ at time t_{n+1} equals

$$P_F(f';\ t_{n+1}) = P_F(f;\ t_n)\ P_F(f';\ t_{n+1} \mid f;\ t_n). \tag{1.7}$$

The probability $P_F(f';\ t_{n+1})$ can also be expressed in terms of the probabilities for the microstates,

$$P_F(f';\ t_{n+1}) = \sum_{z_j \in x_{f'}} \rho_{t_{n+1}}(z_j) = \sum_{z_j \in x_{f'}} \sum_{z_i \in x_f} p_{ji}(t_n)\ \rho_{t_n}(z_i). \tag{1.8}$$

Equations (1.7) and (1.8) imply that the macrostate transition probability,

$$\begin{aligned} P_F(f';\ t_{n+1} \mid f;\ t_n) &= \frac{P_F(f';\ t_{n+1})}{P_F(f;\ t_n)} \\ &= \frac{\sum_{z_j \in x_{f'}} \sum_{z_i \in x_f} p_{ji}(t_n)\ \rho_{t_n}(z_i)}{\sum_{z_i \in x_f} \rho_{t_n}(z_i)} \end{aligned} \tag{1.9}$$

depends on how the probability ρ_{t_n} is distributed among the microstates in x_f, and not only on the transition probabilities p_{ji}.

The argument above can easily be extended to continuous sets by substituting integrals for sums in (1.7) and (1.8) according to

$$\sum_{z \in x_f} \longleftarrow \int \delta(F(z) - f)\, dz. \tag{1.10}$$

Two special cases stand out as exceptions to the dependence on ρ. If the transition probabilities to each microstate j depend only on the previous *macrostate*, so that $p_{ji}(t_n) = P(z_j,\ t_{n+1} \mid x_f,\ t_n)$ then (1.9) reduces to

$$P_F(f';\ t_{n+1} \mid f;\ t_n) = P(z_j,\ t_{n+1} \mid x_f,\ t_n). \tag{1.11}$$

A second very interesting case occurs when ρ depends on the microstate only through F, that is to say, when all the microstates in a given macrostate have the same probability,

$$\rho(z_i) = \frac{P_F(F(z_i);\ t_n)}{\Omega(x_{F(z_i)})} \tag{1.12}$$

where $\Omega(x_f)$ represents the number (or density) of states that corresponds to the macrostate x_f, and is defined either as

$$\Omega(x_f) = \sum_{z_i \in x_f} 1, \tag{1.13}$$

or as

$$\Omega(x_f) = \int \delta(F(z) - f)\, dz, \tag{1.14}$$

depending on whether we are confronted with a discrete or a continuous case, respectively.

Probability distributions that satisfy (1.12) are known as *relevant distributions*, and they also give rise to macroscopic transition probabilities that depend only on the probabilities of the macrostate. Inserting (1.12) into (1.9) leads to

$$P_F(f';\ t_{n+1} \mid f;\ t_n) = \frac{\sum_{z_j \in x_{f'}} \sum_{z_i \in x_f} p_{ji}(t_n)}{\Omega(x_f)}. \tag{1.15}$$

Consequently, coarse-grained descriptions will only be *truly* Markovian when the microscopic transition probabilities satisfy the stern requirement of depending on the present macrostate and not the specific microstate, unless we agree to restrict our attention to relevant distributions. But even then, we must keep in mind that a relevant distribution at time t_n will not generally become a new relevant distribution at time t_{n+1}, as we will see in Chapter 3.

Note, though, that initial relevant probability distributions do allow us in principle to calculate the transition probabilities between the initial macrostates at t_0 and macrostates at any other later time, t_n,

$$P_F(f_n;\ t_n \mid f_0;\ t_0) = \frac{\sum_{z_m \in x_{f_n}} \sum_{z_i \in x_{f_0}} p_{mi}(t_0)}{\Omega(x_f)}, \tag{1.16}$$

with the microscopic transition probabilities p_{mi} calculated according to

$$p_{mi}(t_0) = \sum_l \sum_k \cdots \sum_j p_{ml}(t_n) p_{lk}(t_{n-1}) \cdots p_{ji}(t_1). \tag{1.17}$$

We will encounter descriptions that satisfy (1.11) whenever F represents a dynamical invariant, such as the total energy or the number of particles, for example. Furthermore, we will show that relevant distributions are in a sense the best choice for initial distributions (Section 1.6). But before we do, we must turn to the fascinating link between microscopic reversibility and macroscopic irreversibility.

1.2 Symmetry and irreversibility

> Irreversible stochastic processes defined over macrostates often emerge out of reversible Markov processes.

Most of our everyday life experience is essentially irreversible. We do not need to go far to find examples. Mixing, ageing, cooling and many other phenomena take place irreversibly. I now have a cold cup of tea by the

keyboard, but it would be absurd to wait for a while to see if it warms up again. Does this observed difference between the past and the future imply an asymmetry in the laws of nature with respect to time? Not necessarily. We have shown that coarse-grained descriptions may display memory effects even when the underlying laws possess no memory. Similarly, time irreversibility may emerge from reversible laws.

The set of transition probabilities $p_{ij}(t_n)$ for every i, j and t_n defines a Markov process. To specify the corresponding *reverse* Markov process mathematically, we simply choose new transition probabilities $\tilde{p}_{ij}$ equal to the p_{ij}, but reversing their time sequence. Assume for simplicity that the time steps were equally spaced and that the original Markov process evolved from $t_0 = 0$ to t_N, then

$$\tilde{p}_{ij}(t_n) = p_{ij}(t_N - t_n). \tag{1.18}$$

In a *reversible* process, any trajectory that begins at z_0 and then traverses $z_1, \ldots, z_n$ (for any number of states) should be as likely as a trajectory in the reverse process that starts from z_N and then goes through the same states in reverse sequence, ending at z_0,

$$\tilde{p}(z_{N-1},\ t_1;\ \ldots;\ z_0,\ t_N \mid z_N,\ t_0) = p(z_1,\ t_1;\ \ldots;\ z_N,\ t_N \mid z_0,\ t_0). \tag{1.19}$$

A Markov process is reversible in this sense[1] if, and only if, it satisfies the symmetry condition,

$$p_{ij}(t_n) = p_{ji}(t_n), \tag{1.20}$$

for every pair of states i and j and for every time t_n.

As stated above, reversible Markov processes may underlie irreversible phenomena. A simple example will illustrate the point. We will choose a continuous set of states to demonstrate the straightforward generalisation to probability density functions.

[1]The time dependence of the transition probabilities in (1.19) makes our definition of reversibility different from the customary stationary reversibility implied by Kolmogorov's criteria (See Theorem 1.7 in [7]).

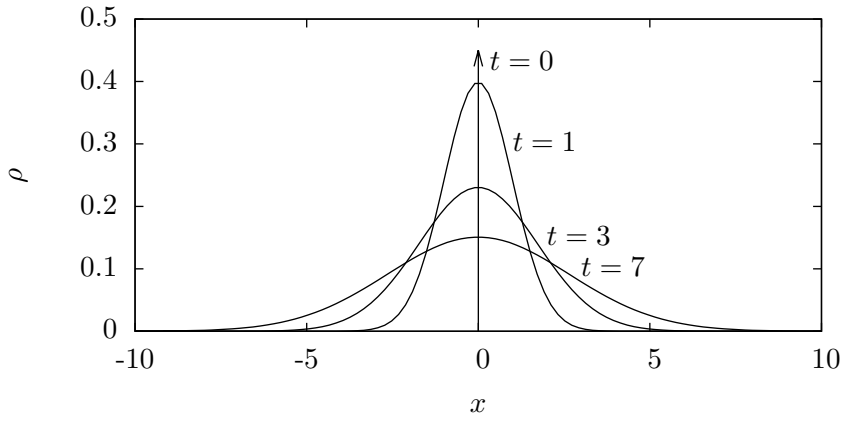

Figure 1.2: Probability density of a random walker at different times. The arrow represents the initial delta function distribution, $\rho_0(x) = \delta(x)$.

Consider a one-dimensional random walk. The probability of stepping from x to x' depends only on the distance between the two points and not on time,

$$p(x' \mid x) = p(x - x') = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x'-x)^2}{2\sigma^2}}. \tag{1.21}$$

Because (1.21) satisfies the symmetry condition (1.20), we have a reversible Markov process, so any trajectory is as likely as the corresponding reverse trajectory. But let us turn to the probability distribution ρ over x and see how it evolves in time. If we are told the shape of ρ at some time t_n, the probability density of x' at time t_{n+1} should equal the sum over x of the probability density times the probability of a transition from x to x'. Hence, the new distribution results from the convolution product of ρ and p.

$$\rho_{t_{n+1}}(x') = (\rho * p)(x') = \int \rho_{t_n}(x,)\ p(x - x')\ dx. \tag{1.22}$$

Beginning at $x = 0$, ρ turns into a normal distribution that widens as time goes by (see Figure 1.2).

The reverse process is identical to the direct process because the transition probabilities do not depend on time. This means that if we start with the final distribution $\rho_{t_N}(x)$, the reverse process will only make it spread out even more!

On the one hand we have reversible trajectories, and on the other we have an irreversible spreading of the probability distribution. We seem to be holding the two main ingredients of a contradiction, though in fact they fit together nicely.

Although we cannot tell from looking at a single trajectory whether we have been given the reverse sequence, a collection of independent random walkers that set out from a given point tend to move away from each other. If we wish to recover the initial state, then we record the positions at each time and run the sequence backwards. It would be hopeless to expect the reverse process to bring all the particles back to their origin, not because the law forbids such a trajectory, but rather because it is extremely improbable.

Therefore, even with reversible trajectories, the probability density function exhibits time irreversibility. We may wish to interpret the increasing width of ρ as a measure of our uncertainty about the position of a random walker initially standing at the origin, or we may prefer to picture an ensemble of independent random walkers, with $\rho(x)$ representing the fraction of walkers at position x. Whichever way we choose to interpret ρ, we will employ the entropy functional [8],

$$S[\rho] = -k_B \int \rho(z) \ln\left(\frac{\rho(z)}{m(z)}\right)\, dz, \tag{1.23}$$

to quantify this irreversibility, because S grows as ρ spreads out and becomes more uniform. The $m(z)$ in the denominator stands for a reference measure, which keeps the results consistent under scaling and other transformations of variables, and k_B stands for Boltzmann's constant.

Let us agree to set $m(x) = 1$ and calculate the entropy for ρ at time t_n.

$$S[\rho_t] = -k_B \int \rho_t(x) \ln(\rho_t(x))\, dx = \frac{1}{2} \ln(2\pi e n \sigma^2). \tag{1.24}$$

From the formula above, we can easily tell that S increases into the future as n grows.

1.3 Shannon and Boltzmann entropies

The difference between the Boltzmann entropies of two macrostates constitutes a way to quantify the irreversibility of a transition between them.

Although we have found irreversibility in the evolution of a probability distribution (or an ensemble of systems), this result does not appear to illuminate the everyday examples with which we began our discussion of irreversibility. In our day-to-day lives, we do not see probability distributions evolve, and we rarely deal with collections of many copies of the same system. Nevertheless, we clearly observe irreversible processes. A single system, my cup of tea, gradually cooled down while I was typing. It went through a sequence of states which it could apparently never have traversed in the opposite direction. Can we not define irreversibility for this single trajectory? Following the cue from the previous section, we might wonder whether *individual trajectories display irreversibility as soon as we decide to use coarse-grained variables instead of microstates.*

For convenience, we will once again assume we have a discrete set of states. The appropriate entropy functional in this context, introduced by Shannon [9], reads

$$S_S[\rho] = -k_B \sum_i \rho(z_i) \ln(\rho(z_i)). \tag{1.25}$$

When the system occupies a given state z_j with probability equal to one, that is, when $\rho(z_i) = \delta_{ij}$ (δ_{ij} stands for the Kronecker delta), the Shannon entropy takes its lowest value,

$$S_S[\rho] = -k_B \ \ln(\rho(z_j)) = -k_B \ \ln(1) = 0. \tag{1.26}$$

But if we only know the system's macrostate x_f then, assuming the microstates $z_i \in x_f$ are all equally probable,

$$\rho_{x_f}(z_i) = \frac{\delta_{F(z_i),f}}{\Omega(x_f)}, \tag{1.27}$$

we can define the Boltzmann entropy[2] of macrostate x_f as

$$S_B(x_f) = S_S[\rho_{x_f}] = k_B \ \ln(\Omega(x_f)) \tag{1.28}$$

(the Kronecker delta $\delta_{F(z_i),f}$ equals one when $F(z_i) = f$ and zero otherwise). We might still harbour some inner reservations about the equiprobability assumption in (1.27). The suitability of such a choice will be proven in Section 1.6 but, for now, we can treat it as a reasonable approximation to the "real" distribution (whatever that means).

From a coarse-grained point of view, a system transitions irreversibly from one macrostate to another whenever the corresponding reverse process is very unlikely. Therefore, let us compare the direct with the reverse transition probabilities. Consider the probability $P_F(f' \mid f)$ of going from x_f to $x_{f'}$ in several steps, $t_1, \ t_2, \ \ldots, \ t_n$. Microscopic reversibility (1.19) guarantees that whenever the transition is possible (that is to say that $P_F(f' \mid f) > 0$), the reverse transition must also be possible. Probability mass functions for x_f and $x_{f'}$ like the one in (1.27) fall into the category of relevant distributions (1.12), so we may use (1.16) to write down the macroscopic transition probabilities between them.

$$\begin{aligned}
\frac{P_F(f' \mid f)}{P_F(f \mid f')} &= \frac{\Omega(x_{f'}) \sum_{z_l \in x_{f'}} \sum_k \cdots \sum_j \sum_{z_i \in x_f} p_{lk}(t_n) \cdots p_{ji}(t_1)}{\Omega(x_f) \sum_{z_i \in x_f} \sum_j \cdots \sum_k \sum_{z_l \in x_{f'}} p_{ij}(t_1) \cdots p_{kl}(t_n)} \\
&= \frac{\Omega(x_{f'}) \sum_{z_l \in x_{f'}} \sum_k \cdots \sum_j \sum_{z_i \in x_f} p_{lk}(t_n) \cdots p_{ji}(t_1)}{\Omega(x_f) \sum_{z_l \in x_{f'}} \sum_k \cdots \sum_j \sum_{z_i \in x_f} p_{ji}(t_1) \cdots p_{lk}(t_n)} \\
&= \frac{\Omega(x_{f'})}{\Omega(x_f)}.
\end{aligned} \tag{1.29}$$

[2]The formula (1.28), carved on Boltzmann's gravestone, should not be confused with Boltzmann's H function, which results from (1.23) when the probability distribution factors into a product of identical single-particle distributions, $\rho = \rho_1^N$ [10].

Changing the order of summation and applying the symmetry condition (1.20) allowed us to cancel the sums in the numerator with those in the denominator. If we now take the logarithm of both sides and multiply by Boltzmann's constant, we are left with a neat expression,

$$k_B \ \ln\left(\frac{P_F(f' \mid f)}{P_F(f \mid f')}\right) = S_B(x_{f'}) - S_B(x_f). \tag{1.30}$$

The term on the left measures irreversibility. When positive, it indicates that the transition has a greater probability of occurring in the direction from x_f to $x_{f'}$ than the other way around. It vanishes when both probabilities are equal and becomes negative when the process tends to go the other way. Equation (1.30) applies to any reversible Markov process.

The theorem has two interesting consequences for nonequilibrium evolution. First, consider irreducible stationary Markov processes. *Irreducibility* means that any state can be reached from any other state (at least one path with non-zero probability connects any pair of states). In *stationary* processes the transition probabilities do not depend on time, $p_{ij}(t) = p_{ij}$ for any of the values of t. If both conditions are met, equation (1.29) implies detailed balance [11]. By choosing an appropriate probability distribution,

$$P_F(f) = \frac{\Omega(x_f)}{\sum_x \Omega(x)} \tag{1.31}$$

where the sum in the denominator extends over all the accessible macrostates, we see that

$$\frac{P_F(f')}{P_F(f)} = \frac{\Omega(x_{f'})}{\Omega(x_f)}, \tag{1.32}$$

and equation (1.29) becomes the condition of detailed balance

$$\frac{P_F(f' \mid f)}{P_F(f \mid f')} = \frac{P_F(f')}{P_F(f)}, \tag{1.33}$$

usually written in the form

$$P_F(f) \ P_F(f' \mid f) = P_F(f') \ P_F(f \mid f'). \tag{1.34}$$

The probabilities P in (1.31) must therefore correspond to the equilibrium distribution because, if we sum (1.34) over the states $x_{f'}$ and use the fact that $\sum_{x_{f'}} P_F(f' \mid f) = 1$, then we obtain the condition for global balance,

$$P_F(f) = \sum_{x_{f'}} P_F(f \mid f')\ P_F(f'), \tag{1.35}$$

so that the macroscopic transition matrix times P equals P again. The foregoing discussion could be summed up in one single (cryptic) statement: *irreducible reversible stationary Markov processes relax to equilibrium irreversibly*[3].

In the case of non-stationary processes, the argument in the preceding paragraphs no longer holds, because the complete set of accessible states might change over time. Still, if $P_F(f;\ t_0)$ and $P_F(f';\ t_N)$ represent initial and final equilibrium distributions, we can rewrite equation (1.30) in terms of P_F in (1.31). We begin by subtracting the logarithm of $P_F(f';\ t_N)/P_F(f;\ t_0)$ from both sides of (1.30)

$$\begin{aligned} k_B\ \ln\left(\frac{P_F(f' \mid f)\ P_F(f;\ t_0)}{P_F(f \mid f')\ P_F(f';\ t_N)}\right) &= -k_B\ \ln\left(\frac{P_F(f';\ t_N)}{\Omega(x_{f'})}\right) \\ &\quad + k_B\ \ln\left(\frac{P_F(f;\ t_0)}{\Omega(x_f)}\right). \end{aligned} \tag{1.36}$$

Next, we average over the joint initial and final distributions, which are independent because the distributions P_F (1.31) depend only on the present set of accessible states and not on what states were accessible at other times.

$$\begin{aligned} \sum_{x_f}\sum_{x_{f'}} P_F(f';\ t_N)\ P_F(f;\ t_0) k_B\ \ln\left(\frac{P_F(f' \mid f)\ P_F(f;\ t_0)}{P_F(f \mid f')\ P_F(f';\ t_N)}\right) \\ = S_S[P_F(f;\ t_N)] - S_S[P_F(f;\ t_0)]. \end{aligned} \tag{1.37}$$

[3]Strictly speaking, we have not shown that any arbitrary initial probability distribution *approaches* P_F asymptotically in the long run, but this *can* be proven rigorously [13].

If the P_F represented equilibrium probability density functions, the same argument entails a continuous equivalent of the previous formula,

$$\int\int P_F(f';\ t_N)\ P_F(f;\ t_0)\ k_B\ \ln\left(\frac{P_F(f'\mid f)\ P_F(f;\ t_0)}{P_F(f\mid f')\ P_F(f';\ t_N)}\right)\ df\ df' \\ = S[P_F(f;\ t_N)] - S[P_F(f;\ t_0)]. \quad (1.38)$$

with the entropy functional (1.23) instead of the Shannon entropy, as in (1.37).

Whereas equation (1.30) quantifies the extent to which transitions between definite macrostates take place irreversibly, (1.37) and (1.38) measure irreversibility when we do not possess information about the exact initial and final macrostates, and instead we only know P_F. In the laboratory, scientists typically control a set of parameters such as pressure, temperature, pH, and so on. While the parameters remain fixed, the system goes from one macrostate to another sampling the stationary equilibrium distribution in the long run. From this point of view, states of equilibrium do not really correspond to definite macrostates but rather to probability distributions determined by the values of the control parameters. In (1.37) and (1.38) we are given the expected value of (1.30) as a result of changes in the parameters. In other words, the difference between the entropies of the initial and final probability distributions equals the expected value of our measure of irreversibility (1.30).

Although we have not yet paid any attention to the actual reversible laws obeyed by physical systems, we can already conclude that manipulating the control parameters is unlikely to bring a system back to a lower entropy equilibrium unless the system belongs to a larger setup that does not decrease its entropy in the process. We now interpret the *second law of thermodynamics* precisely in this statistical sense.

Equation (1.30) qualifies as a *fluctuation theorem*. I have not yet found it stated in this way in the statistical mechanics literature, but I suppose it is probably well known. It most resembles Jarzynski's result for *Hamiltonian*

systems: the detailed fluctuation theorem [14],

$$\frac{P_{S'-S}(s;\ t)}{P_{S'-S}(-s;\ \tilde{t})} = e^{\frac{s}{k_B}}. \tag{1.39}$$

Equation (1.39) compares the probability $P_{S'-S}(s)$ of observing an entropy production equal to s to the probability of the opposite entropy production during the reverse process. If there is a one-to-one correspondence connecting each possible entropy difference to a pair of macrostates, then (1.39) expresses the same fact as (1.30).

Similar results have been derived also for stochastic systems in contact with a heat reservoir[4]. Another popular variant, proven by Crooks relates the probability of observing an entropy production $\dot{S} = s$ during the direct process to the opposite $\dot{S} = -s$ in the reverse process [16].

$$\frac{P_{\dot{S}}(s;\ t)}{P_{\dot{S}}(-s;\ \tilde{t})} = e^{\frac{s}{k_B}}. \tag{1.40}$$

The crucial difference between (1.40) and (1.30) comes from the assumption that the system exchanges energy with a heat reservoir at temperature T. We did not use this assumption. However, if we do assume canonical densities of states, then (1.40) follows along the same lines as the 1999 proof in Ref. [16].

1.4 Hamiltonian reversibility

Trajectories in Hamiltonian mechanics count as Markov processes, but they are not reversible in the sense of equation (1.20). Here we define a different notion of reversibility that allows us to extend the results of the previous section to Hamiltonian dynamics.

In a deterministic Newtonian universe, the laws of motion would have no recollection of the past and would proceed blindly into the future, but the

[4] See Section V in Ref. [15].

present state of the universe, with the whole of its history imprinted on the finest details of its structure, would also contain the seeds of everything to come. Laplace famously remarked in his essay on probabilities [17] that an intelligence capable of comprehending the laws of nature, who could picture and analyse the complete state of the universe (the exact position and velocity of every single particle), would also have before its eyes the past and future, and nothing would be uncertain or lost.

On the contrary, life teaches us that some events are forgotten and irrecoverable. If the tide washed away your footprints, nobody can tell from looking at the sand that you walked across the beach. Time works hard to eliminate the memory of things gone by. For example, when a child puts her hand in a pool, sending ripples across the surface, the water is calm again a little while later. In the absence of external interference, matter tends towards states of equilibrium which depend only on the values of a handful of parameters and not on the previous states, with regard to which it behaves as if it suffered from amnesia. Nonequilibrium statistical mechanics typically involves this kind of loss of information about the history of a system because many different paths could have led to the same final state, so we cannot tell from the outcome which path the system must have followed.

Laplace claimed that this loss of the past was no more than an illusion caused by our ignorance. When nature tries to remove the traces of an event, it creates other traces, just like the criminal in a detective novel, who leaves the front door locked from the inside to confound the police and then jumps from the rooftop to the neighbour's house and flees, without realising that he has left his footprints in a flower bed for Sherlock Holmes to discover. Similarly, we might be unable to tell whether the water in a kettle has cooled down to room temperature, or perhaps Polly just forgot to put the kettle on, but if only we could make all the atoms in the room turn around and backtrack their steps, then we would know for sure. The clues are still there, hidden among all the other minute details of the room's microstate.

Hamiltonian mechanics spring from a variational principle [18]. In this context, microstates represent the complete set of coordinates q and conju-

gate momenta p necessary to determine the state of the system completely, $z = (q,\ p)$. The Hamiltonian function $H(z,\ t) = H(q,\ p,\ t)$ corresponds to the total energy for microstate z at time t. The principle of stationary action declares that the actual trajectory followed by the system in the space of states, known here as *phase space*, makes the action integral A stationary.

$$A = \int_{t_1}^{t_2} (p \cdot \dot{q} - H(q,\ p,\ t))\ dt. \tag{1.41}$$

Dots above a symbol conventionally represent time derivatives, $\dot{q} = \frac{dq}{dt}$. Solving the variational equation $\delta A = 0$ leads to Hamilton's famous canonical equations of motion

$$\left.\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p}, \\ \dot{p} &= -\frac{\partial H}{\partial q}. \end{aligned}\right\} \tag{1.42}$$

Hamiltonian dynamics differ from the laws hitherto considered in two important aspects. Firstly, Hamiltonian systems do not satisfy our definition of reversibility (1.19). When a system follows a trajectory from z to z', the reverse trajectory is impossible.

A very interesting symmetry allows us to extend our results to Hamiltonian laws: simultaneous inversion of p and t leaves the above equations unchanged. Let $\tilde{z}$ represent a microstate like z but with all the velocities reversed, $\tilde{z} = (q,\ -p)$. When Hamilton's equations decree that a trajectory will go from z_t to $z_{t+\tau}$ then, if we flip the velocities, the reverse process will lead from $\tilde{z}_{t+\tau}$ back to $\tilde{z}_t$ [19]. Because the dynamical equations establish a one to one correspondence between microstates z at time t and microstates $T(z;\ t;\ \tau)$ at the later time $t + \tau$, the time-reversal symmetry implies that

$$\tilde{z} = \tilde{T}(\widetilde{T(z;\ t;\ \tau)};\ t + \tau; \tau). \tag{1.43}$$

In the above equation, we have used $\tilde{T}$ to indicate the reverse process transformation.

Suppose that the function F chosen to define the macrostates satisfies the condition $F(\tilde{z}) = F(z)$ (kinetic temperature and species concentration

are two examples of such functions). Then we can find an equation parallel to (1.29) for our Hamiltonian system.

$$\begin{aligned}\frac{P_F(f' \mid f)}{P_F(f \mid f')} &= \frac{\Omega(x_{f'}) \int \delta(F(z) - f)\ \delta(F(T(z;\ t;\ \tau)) - f')\ dz}{\Omega(x_f) \int \delta(F(z') - f')\ \delta(F(\tilde{T}(z';\ t+\tau;\ \tau)) - f)\ dz'} \\ &= \frac{\Omega(x_{f'}) \int \delta(F(z) - f)\ \delta(F(\widetilde{T(z;\ t;\ \tau)}) - f')\ dz}{\Omega(x_f) \int \delta(F(\tilde{z}) - f)\ \delta(F(\widetilde{T(z;\ t;\ \tau)}) - f')\ |J_{\tilde{T}}(z)|\ dz} \\ &= \frac{\Omega(x_{f'}) \int \delta(F(z) - f)\ \delta(F(T(z;\ t;\ \tau)) - f')\ dz}{\Omega(x_f) \int \delta(F(z) - f)\ \delta(F(T(z;\ t;\ \tau)) - f')\ |J_{\tilde{T}}(z)|\ dz}.\end{aligned} \tag{1.44}$$

The change of variable $z' = \widetilde{T(z;\ t;\ \tau)}$ was responsible for the appearance of the Jacobian $|J_{\tilde{T}}(z)|$ in the denominator. In principle, as we shall see below, we can always make this determinant equal to one by inclusion of extra coordinates to complete the description of our system. In that case, we would get an expression analogous to (1.29),

$$\frac{P_F(f' \mid f)}{P_F(f \mid f')} = \frac{\Omega(x_{f'})}{\Omega(x_f)}, \tag{1.45}$$

but remember that equation (1.45) holds when $F(z) = F(\tilde{z})$. Had we chosen a function that changes sign when the velocities flip, $F(z) = -F(\tilde{z})$ (like the centre of mass velocity of a group of particles, for example) then we would have inferred a different formula:

$$\frac{P_F(f' \mid f)}{P_F(-f \mid -f')} = \frac{\Omega(x_{f'})}{\Omega(x_f)}. \tag{1.46}$$

Mathematically, there might be absolutely no relation between $F(z)$ and $F(\tilde{z})$, but in practice most interesting cases are covered by the two conditions mentioned above.

Before we continue, we should make sure that an adequate choice of coordinates renders the Jacobian determinant equal to unity, $|J_{\tilde{T}}(z)| = 1$.

The Hamiltonian equations of motion transform any microstate in phase space into another, so we may imagine the laws of time evolution as a flow through phase space. In this picture, the Jacobian determinant represents the factor by which a volume element expands or shrinks at z. A well-known result in classical mechanics states that autonomous Hamiltonian dynamics conserve the volume in phase flows, that is to say, whenever the *laws* of time evolution do *not* depend on time, the Jacobian determinant equals one, corresponding to the steady flow of an incompressible phase fluid[5].

We now argue as follows. Time-dependent laws appear as a result of external interactions. The coordinates and momenta used in our description do not include the state of other objects that affect the movement of our system. Moving walls, charged plates, heat baths and other similar external devices could in principle be counted as part of the system of interest by adding their microstates to the complete description of the dynamics. Let z stand for the microstate of what we previously regarded as our system, and z_S for the microstate of the surroundings. Assume that we now have an isolated extended system. If necessary, include the whole universe. The Hamiltonian laws that govern our enlarged system should now be autonomous, so the Jacobian determinant will equal one. The densities of states in equations (1.45) and (1.46) will become integrals over the extended phase space,

$$\Omega_E(x_f) = \int \delta(F(z) - f)\ dz_E, \tag{1.47}$$

where $dz_E = dz\ dz_S$, but realise that the additional integrals cancel.

$$\frac{\Omega_E(x_{f'})}{\Omega_E(x_f)} = \frac{\int \delta(F(z) - f')\ dz_E}{\int \delta(F(z) - f)\ dz_E} = \frac{\Omega(x_{f'}) \int dz_S}{\Omega(x_f) \int dz_S} = \frac{\Omega(x_{f'})}{\Omega(x_f)}. \tag{1.48}$$

This means that the ratio of densities of states in the extended phase space equals the ratio for the macrostates of the original system, making equations (1.45) and (1.46) valid also in the case of time-dependent Hamiltonians.

[5]See [18], Chapter VI, 6.

We do not really need to speculate about the source of the time-dependent effects, as an alternative (though considerably more abstract) line of reasoning leads to the same result. Let us extend our original system by including time as a single extra coordinate dependent on a parameter named s[6]. If we use primes to denote derivatives with respect to s, then $\dot{q}$ becomes q'/t' and we can rewrite the action integral (1.41) in terms of s,

$$A = \int_{s_1}^{s_2} (p \cdot q' - H(q,\ p,\ t)t')\ ds. \tag{1.49}$$

We may think of our extended system as possessing a vanishing Hamiltonian $H_E(q,\ p,\ t, p_t) = 0$. The momentum conjugate to the time coordinate coincides with the negative of the original Hamiltonian function, $p_t = -H(q,\ p,\ t)$. Defining $K(q,\ p,\ t,\ p_t) = p_t + H(q,\ p,\ t)$, the problem of finding the equations of motion turns out to be equivalent to working out the stationary value of the action subject to the constraint $K = 0$. The well-known method of Lagrange multipliers states that the solution to our problem comes from solving the following variational equation:

$$\delta \int_{s_1}^{s_2} (p \cdot q + p_t t - \lambda K(q,\ p,\ t,\ p_t))\ ds = 0. \tag{1.50}$$

We can easily make the Lagrange multiplier λ equal to one by scaling the parameter s by a factor of $1/\lambda$, so we will leave λ out of our equations. Solving (1.50) brings us to the equations of motion,

$$\left.\begin{aligned} q' &= \frac{\partial K}{\partial p}, \\ p' &= -\frac{\partial K}{\partial q}, \\ t' &= \frac{\partial K}{\partial p_t}, \\ p_t' &= -\frac{\partial K}{\partial t}. \end{aligned}\right\} \tag{1.51}$$

[6]For a more detailed explanation of the parametric form of canonical equations, see [18], Chapter VI, 10.

Note that the resulting laws are autonomous (i.e., independent of parameter s) and therefore represent a steady incompressible phase flow in the extended phase space for which the Jacobian matrix equals one everywhere. Replacing K above with its definition, we recover Hamilton's equations plus two extra equations for t' and p'_t,

$$\left.\begin{aligned} \dot{q} &= \frac{\partial H}{\partial p}, \\ \dot{p} &= -\frac{\partial H}{\partial q}, \\ t' &= 1, \\ p'_t &= -\frac{\partial H}{\partial t}. \end{aligned}\right\} \tag{1.52}$$

Because $t' = 1$, we discover that $s = t + t_0$ for some constant t_0, so differentiating with respect to s amounts to differentiating with respect to t.

In summary, because we can always make a Hamiltonian system autonomous by adding coordinates and momenta to the description, our equations (1.45) or (1.46) apply also to time-dependent Hamiltonians.

A second contrast between Hamiltonian systems and reversible Markov processes poses a serious threat to the whole project described in this chapter. In Section 1.3 we assumed that our processes were irreducible to prove that they relax to equilibrium. Unfortunately, this assumption is not valid for Hamiltonian systems. To make matters worse, Hamiltonian trajectories are either periodic or nonrecurrent. In the theory of Markov chains, reducible systems allow you to divide your probability distribution into different parts which relax to equilibrium independently, but this cannot be done with periodic or nonrecurrent states [13]. This observation forces us to conclude that Hamiltonian systems do not relax to equilibrium!

In fact, we can easily prove that entropy remains constant under autonomous Hamiltonian time evolution. Because of the deterministic dynamics, probability densities move along trajectories in phase space without

diffusing,

$$\rho_t(z) = \rho_0(\tilde{T}(z;\ 0;\ t)). \tag{1.53}$$

This means that the probability density at time t for z came from the initial probability density at the start of the trajectory that led to z. Therefore, the entropy at time t equals the initial entropy.

$$\begin{aligned} S[\rho_t] &= \int \rho_0(\tilde{T}(z';\ 0;\ t))\ \ln\left(\frac{\rho_0(\tilde{T}(z';\ 0;\ t)}{m(z')}\right)\ dz' \\ &= \int \rho_0(z)\ \ln\left(\frac{\rho_0(z)}{m(z)}\right)\ dz = S[\rho_0]. \end{aligned} \tag{1.54}$$

Equation (1.54) follows from changing variables $z' = T(z;\ t;\ t)$ and the fact that $|J_T(z)| = 1$.

There are two popular responses to this challenge, both in line with Laplace's philosophy. To begin with, Hamiltonian mechanics may well keep probability densities fixed to a trajectory, but we can regard the trajectory itself as stochastic to some extent. When we choose a Hamiltonian function to describe a phenomenon, we typically idealise its interactions with the environment either by neglecting them completely or by modelling them with a deterministic time-varying potential. External interference will only ever be approximated by such idealisations. Real systems react to unpredictable external effects. For example, we describe a liquid in a Dewar flask on the laboratory table as a mechanically and thermally isolated system. Experimental results validate this approximate description, but when we consider what really takes place at the atomic level we realise that energy leaks in and out of the system through interactions with the container walls and electromagnetic radiation. Sound waves and footsteps make the flask vibrate and this transmits momentum to the atoms in our system of interest. Furthermore, atoms have inner degrees of freedom that we usually leave out of the description although they may absorb and emit energy. These processes barely affect the total energy inside the flask, but they do have substantial effects on the atomic trajectories. Since we ignore all these fine details, as Laplace would have said, we can think of them as amounting to

a small probability of shifting from one idealised Hamiltonian trajectory to another[7].

What if we insisted on examining the ideal autonomous Hamiltonian dynamics as a mathematical problem? Would we then have to abandon the concept of equilibrium? Not necessarily. We will still observe equilibration, in a sense, as long as two conditions are met: *mixing* and *finite resolution*. The latter requires us to accept that we cannot specify the state of our system with infinite precision, that our knowledge of positions and momenta involves at least some minimum amount of uncertainty. We imagine phase space divided up into tiny volumes and say that the system's microstate belongs to one or more of them. In the context of transitions between volumes, mixing dynamics (the second condition) spread volumes all over the accessible phase space in the way that a drop of milk mixes with tea. The mathematical definition of (strong) mixing invokes the idea that, in the long run, the probability of finding the microstate in a given volume becomes independent of the volume initially containing the microstate. Let v and v' stand for arbitrary volumes, and choose F so that $F(z) = f$ when $z \in v$ and $F(z) = f'$ when $z \in v'$, then

$$\lim_{t\to\infty} \int \rho_t(z)\delta(F(z) - f')\, dz \\ = \lim_{t\to\infty} \frac{\int \rho_0(z)\delta(F(z) - f)\delta(F(T(z;0;t) - f')\, dz}{\int \rho_0(z)\delta(F(z) - f)\, dz}. \quad (1.55)$$

The right hand side expresses the long term probability of finding the system in v' when it started off in v.

As soon as we face systems with a handful of bounded degrees of freedom and some kind of nonlinear interaction, most of the points in phase space for an autonomous Hamiltonian become nonrecurrent, which is to say that the system visits them at most once, and the dynamics satisfy (1.55). To say that we have mixing dynamics corresponds to claiming that the distribution in the long run will be independent of the current distribution.

[7]This point of view would lead us to wonder about the correct statistical description of these jumps.

What will the long time averages look like? To find out, let me first define the concept of *dynamical invariant*. A function I of the microstates counts as a dynamical invariant if its value does not change as time goes by,

$$I(T(z;\ t;\ \tau)) = I(z), \tag{1.56}$$

for any z, t and τ. If we are given any probability distribution over the values of I, $P_I(i)$, then we can immediately define an invariant probability distribution over the *microstates* (see Section 2.1),

$$\rho^{eq}(z) = \frac{P_I(I(z))}{\Omega(x_{I(z)})}. \tag{1.57}$$

Note that our equilibrium distribution (1.57) belongs to the set of relevant distributions (1.12) and that it does not change in time, $\rho^{eq}(T(z;\ t;\ \tau)) = \rho^{eq}(z)$.

Now suppose we allow an arbitrary initial distribution ρ_0 to evolve under autonomous Hamiltonian laws. As time moves forward, we will generally find the corresponding ρ_t changing shape, but the probability distributions P_I for the dynamical invariants will not. We may use P_I to construct an equilibrium distribution (1.57) with the same probabilities for the dynamical invariants. The mixing condition (1.55) states that averages calculated with ρ_t converge towards the averages calculated with ρ^{eq} as time increases, yet ρ^{eq} remains constant in time, so this means that ρ_t approaches ρ^{eq} as time tends towards infinity. Remember that "approach" in this context refers to average values or, as mathematicians would have it, convergence in the *weak* sense, as more often than not $\lim_{t\to\infty} \rho_t(z) \neq \rho^{eq}(z)$. Be that as it may, having admitted our inability to resolve a distribution's fine details beyond a certain point, we realise that we will eventually become incapable of telling ρ_t and ρ^{eq} apart.

The last few paragraphs were dedicated to autonomous Hamiltonian dynamics, so the argument does not apply to time-dependent laws. The trick of extending the phase space to include time will not work now because trajectories will not be bounded in the direction of the new t coordinate. We could, of course, bring the whole environment into our system once

again and argue that, once we consider the surroundings, equilibrium will eventually set in. I must confess my scepticism about the merits of this point of view. When we include everything that interacts with our system, equilibration often takes place over vast time scales, leaving out the details involved in the process we were interested in. Consider convection in the atmosphere, caused by the Sun's light warming the ground. Focusing on relaxation to equilibrium tells us very little about the behaviour of the weather. Perhaps the Sun will burn out in the long run and reach thermal equilibrium with whatever remains of the Earth but, as John Maynard Keynes once remarked, in the long run we are all dead [20].

1.5 Reversible molecular dynamics simulations

We present a memoryless symplectic bit-reversible algorithm for molecular dynamics inspired by the work of Levesque and Verlet. At the end of this chapter, we will use our scheme to demonstrate that irreversibility emerges from reversible laws.

Scientists must check their theories against reality. When a theory in nonequilibrium statistical mechanics fails to predict experimental results, we can have a hard time explaining what went wrong. To begin with, we infer macroscopic laws from the behaviour of microscopic interactions, which we take for granted. Maybe the error crept in when we adopted a particular model for the microscopic dynamics. Mistakes can also come from incorrect steps in the mathematical derivation of our macroscopic laws or from invalid approximations used to make the resulting equations tractable. When cornered, we can also blame the experimentalists.

Computer simulations help us validate the connection between microscopic and approximate macroscopic laws. Molecular dynamics implement Laplace's idea by calculating the evolution of every single degree of freedom in a system and tracking the transition from one macrostate to the next [21]. When our derived laws predict the outcome of a simulation, we show

that *if* a system behaves according to the dynamics we adopted, *then* our laws approximate its macroscopic evolution.

To integrate the equations of motion numerically, we divide the time axis into discrete steps. The trajectories we calculate do not follow the exact solutions, due partly to the approximation involved in determining the next microstate and partly to roundoff, as we cannot store real numbers with infinite precision. The tiny differences between exact and approximate solutions suffice to induce irreversibility in the numerical solutions. In the following pages, we will see how to construct a simple Hamiltonian reversible integration scheme. Using this algorithm, we know for sure that irreversibility originates in the phenomenon, as opposed to being an artefact of the program.

Our Hamiltonian $H(q,\ p,\ t)$ corresponded to a time evolution transformation T that belongs to the family of *canonical transformations*, defined as one-to-one maps in phase space that do not change the form of Hamilton's canonical equations (1.42). We could think of any motion of the phase fluid itself as a transformation of this kind that varies continuously in time, with the canonical equations representing an infinitesimal canonical transformation from each z to its corresponding $z + \dot{z}\ dt$.

No general method exists to figure out the analytic expression of $T(z;\ t,\ t+\tau)$, though we do have methods to check whether an arbitrary transformation is canonical or not. We will write $\{u,\ v\}$ for the Poisson bracket of u and v,

$$\{u,\ v\} = \sum_i \left(\frac{\partial u}{\partial q_i}\frac{\partial v}{\partial p_i} - \frac{\partial u}{\partial p_i}\frac{\partial v}{\partial q_i} \right). \tag{1.58}$$

Given a rule that maps q and p to Q and P, we can confirm its canonical character if and only if it satisfies

$$\{Q_j,\ Q_k\} = 0,\ \{P_j,\ P_k\} = 0,\ \{Q_j,\ P_k\} = \delta_{jk}, \tag{1.59}$$

for every pair of degrees of freedom j and k [18]. We can easily generate a canonical transformation from an arbitrary differentiable function of the

old coordinates and the new momenta [22], $G(q,\ P)$, by defining

$$\left.\begin{aligned} p &= \frac{\partial G}{\partial q}, \\ Q &= \frac{\partial G}{\partial P}. \end{aligned}\right\} \tag{1.60}$$

We would like to have a canonical transformation that takes the system from its state at time t to the state at time $t+\tau$. Unfortunately, as we mentioned above, we do not possess a general method to calculate the exact transformation, but the generating function $G(q(t),\ p(t+\tau),\ t) = q(t)\ p(t+\tau) + \tau\ H(q(t),\ p(t+\tau),\ t)$ gives us a reasonable approximation for small values of τ,

$$\left.\begin{aligned} q(t+\tau) &= q(t) + \tau\ \frac{\partial H(q(t),\ p(t+\tau),\ t)}{\partial p(t+\tau)}, \\ p(t+\tau) &= p(t) - \tau\ \frac{\partial H(q(t),\ p(t+\tau),\ t)}{\partial q(t)}. \end{aligned}\right\} \tag{1.61}$$

Dividing by τ and leaving only the partial derivatives on the right, we see that (1.61) turns into Hamilton's equations (1.42) as $\tau \to 0$.

Most of the Hamiltonians in molecular dynamics simulations equal the kinetic energy T_k plus the potential energy V, $H(q,\ p,\ t) = T_k(p) + V(q,\ t)$. In that case, the above equations become

$$\left.\begin{aligned} q(t+\tau) &= q(t) + \tau\ \frac{\partial T_k(p(t+\tau))}{\partial p(t+\tau)}, \\ p(t+\tau) &= p(t) - \tau\ \frac{\partial V(q(t),\ t)}{\partial q(t)}. \end{aligned}\right\} \tag{1.62}$$

A straightforward proof of the canonical nature of (1.62) follows from applying the conditions (1.59) to this transformation. Not only do (1.62) approximate the exact trajectories, but we can consider them exact Hamiltonian solutions in their own right. These equations describe the exact time-τ evolution of an associated Hamiltonian system [23], described by

$$H_p = H + \tau H_1 + \tau^2 H_2 + \tau^3 H_3 + \ldots, \tag{1.63}$$

where H stands for the Hamiltonian used to form (1.62) and H_1, H_2, etc. represent combinations of higher derivatives of H with respect to q and p. The time evolution described by (1.62) conserves H_p exactly. Because H_p embodies a slightly perturbed version of our original Hamiltonian function H, the trajectories calculated with our scheme will never wander far from the exact solutions of Hamilton's equations [23]. The error in the total energy never grows beyond order $o(\tau)$[8]. This conclusion extends to the general transformation (1.61) as well [23]. Sadly, the perturbation that gave rise to H_p might completely destroy other dynamical invariants I, in the sense that no modified function I_p exists which the transformations (1.62) conserve exactly [25].

A parallel derivation with the alternative generating function $G(q(t+\tau),\ p(t),\ t) = q(t+\tau)\ p(t) + \tau\ H(q(t+\tau),\ p(t),\ t+\tau)$ and the definitions

$$\left.\begin{aligned} q(t) &= \frac{\partial G(q(t+\tau),\ p(t),\ t)}{\partial p(t)}, \\ p(t+\tau) &= \frac{\partial G(q(t+\tau),\ p(t),\ t)}{\partial q(t+\tau)}, \end{aligned}\right\} \tag{1.64}$$

drives us to a similar canonical integration method,

$$\left.\begin{aligned} q(t+\tau) &= q(t) + \tau\ \frac{\partial T_k(p(t))}{\partial p(t)}, \\ p(t+\tau) &= p(t) - \tau\ \frac{\partial V(q(t+\tau),\ t+\tau)}{\partial q(t+\tau)}. \end{aligned}\right\} \tag{1.65}$$

Equations (1.62) and (1.65) are close relatives. Neither method remains invariant under time inversion, but when we have carried out a simulation with one of them, we can use the other to calculate the reverse process!

Take any initial state $(q_0,\ p_0)$ and apply (1.62) to find the next point

[8] For practical applications, researchers often prefer higher order integrators. For Hamiltonians of the form $H = T_k(p) + V(q)$, explicit schemes exist to construct symplectic integrators [24]. Implicit schemes can approximate arbitrary Hamiltonians, either by the generating function method discussed here or by implicit Runge-Kutta methods [23].

on the trajectory,

$$\left\{ q_1 = q_0 + \tau \frac{\partial T_k(p_1)}{\partial p_1};\ p_1 = p_0 - \tau \frac{\partial V(q_0,\ t)}{\partial q_0} \right\}. \tag{1.66}$$

Now flip the velocities and calculate the next point with (1.65), remembering that the time dependence in the potential V should also be reversed,

$$\begin{aligned} q_0^r &= q_1 + \tau \frac{\partial T_k(-p_1)}{\partial - p_1} \\ &= q_0 + \tau \frac{\partial T_k(p_1)}{\partial p_1} - \tau \frac{\partial T_k(p_1)}{\partial p_1} = q_0, \end{aligned} \tag{1.67}$$

$$\begin{aligned} p_0^r &= p_1 - \tau \frac{\partial V(q_1,\ t-\tau)}{\partial q_1} \\ &= p_0 - \tau \frac{\partial V(q_0,\ t)}{\partial q_0} + \tau \frac{\partial V(q_0,\ t)}{\partial q_0} = p_0. \end{aligned} \tag{1.68}$$

Because the initial state was arbitrary, the identity of $(q_0^r,\ p_0^r)$ and $(q_0,\ p_0)$ applies to any point in phase space, implying that the algorithm (1.65) reverses time with respect to (1.62).

We can decompose the first algorithm (1.62) into two elementary canonical transformations:

$$(q,\ p) \ \rightarrow \ \left(q,\ p - \tau \frac{\partial V(q,\ t)}{\partial q} \right), \text{ and} \tag{1.69}$$

$$(q,\ p) \ \rightarrow \ \left(q + \tau \frac{\partial T_k(p)}{\partial p},\ p \right). \tag{1.70}$$

Canonical transformations do not generally commute. This explains why we need a variant algorithm for the reverse transformation, as applying these elementary transformations in the opposite order yields a different scheme.

When the kinetic energy has the familiar p^2/m form, (1.62) reduces to

$$\left.\begin{aligned} q(t+\tau) - q(t) &= \frac{\tau}{m}\ p(t) - \frac{\tau^2}{m}\ \frac{\partial V(q(t),\ t)}{\partial q(t)}, \\ p(t+\tau) - p(t) &= -\tau\ \frac{\partial V(q(t),\ t)}{\partial q(t)}. \end{aligned}\right\} \tag{1.71}$$

Substituting $t-\tau$ for t above,

$$\left.\begin{aligned} q(t)-q(t-\tau) &= \tau\ \frac{p(t-\tau)}{m} - \frac{\tau^2}{m}\frac{\partial V(q(t-\tau),\ t-\tau)}{\partial q(t-\tau)}, \\ p(t)-p(t-\tau) &= -\tau\ \frac{\partial V(q(t-\tau),\ t-\tau)}{\partial q(t-\tau)}. \end{aligned}\right\} \tag{1.72}$$

Subtracting the expression for $q(t)-q(t-\tau)$ from the equation for $q(t+\tau)-q(t)$ and using the identity for $p(t)-p(t-\tau)$ we get a reversible equation for the coordinates,

$$q(t+\tau)-2\ q(t)+q(t-\tau) = \tau\frac{\partial V(q(t),\ t)}{\partial q(t)}. \tag{1.73}$$

This centred finite-difference numerical integration algorithm translates Newton's second law into the language of discrete computation for forces derived from a potential function V. It was used as early as 1791 by Joseph Delambre to carry out astronomical calculations and was popularised by Loup Verlet in the context of molecular dynamics in 1967 [26].

We can repeat the operations in the previous paragraph with the reverse integration scheme (1.65) to arrive at exactly the same expression (1.73).

In theory, we have procured discrete-time rules to work out the direct and reverse trajectories in a Hamiltonian system. Real programs bring up an extra complication due to the rounding-off of floating point numbers, which we can once again regard as small perturbations to the solutions. The typical instability of the dynamical equations amplifies these perturbations exponentially and, because the roundoff in the direct and reverse process do not necessarily coincide, the reverse trajectory bears no resemblance to its direct counterpart when we lengthen the simulations.

Levesque and Verlet realised how to overcome this problem by paying attention to the rules of roundoff [27]. They proposed the use of very large integers to measure lengths, times and masses. By consistently rounding off the right-hand side of (1.73) to an integer value, they attained a computationally reversible algorithm up to the very last significant bit.

$$q(t+\tau)-2\ q(t)+q(t-\tau) = \mathrm{Int}\left(\tau\frac{\partial V(q(t),\ t)}{\partial q(t)}\right), \tag{1.74}$$

where $\mathrm{Int}(x)$ stands for the integer n closest to x such that $|n| \leq |x|$. Velocities could then be obtained by averaging over position differences divided by time steps. Regrettably, the integration method depends on the states at times t and $t-\tau$, so (1.74) cannot be considered memoryless.

Still and all, we can transfer the same idea to our pair of algorithms (1.62) and (1.65). Instead of using integers, we consistently round off our coordinate and momenta differences to a fixed number of significant figures d, indicated below by square brackets and a subindex d.

$$\left.\begin{aligned} q(t+\tau) - q(t) &= \tau \ \frac{\partial T_k(p(t+\tau))}{\partial p(t+\tau)}, \\ p(t+\tau) - p(t) &= -\left[\tau \ \frac{\partial V(q(t),\ t)}{\partial q(t)}\right]_d, \end{aligned}\right\} \tag{1.75}$$

$$\left.\begin{aligned} q(t+\tau) - q(t) &= \tau \ \frac{\partial T_k(p(t))}{\partial p(t)}, \\ p(t+\tau) - p(t) &= -\left[\tau \ \frac{\partial V(q(t+\tau),\ t+\tau)}{\partial q(t+\tau)}\right]_d. \end{aligned}\right\} \tag{1.76}$$

The direct (1.75) and reverse (1.76) numerical integrations define a deterministic Markov chain in a discrete phase space. Within this set of states autonomous Hamiltonians conserve the "volume" of phase flow exactly. Limiting our precision to a set number of significant figures is equivalent to carving up phase space into tiny volumes. Our algorithms (1.75) and (1.76) also conserve the volume of phase flow because they simply shuffle these volumes around according to a fixed rule, so they will not mix in the sense of (1.55), although we will observe mixing of larger sets of these small volumes. In contrast to (1.74), our scheme depends only on the present microstate and incorporates the momenta in such a way that each integration step amounts to the application of a canonical transformation.

Figure 1.3 shows a very short simulation of stacked disks interacting through a truncated Lennard-Jones potential in a gravitational field,

$$V(r_{ij}) = \begin{cases} 4\epsilon\left(\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right) + \epsilon, & r_{ij} \leq 2^{1/6}\sigma \\ 0, & r_{ij} > 2^{1/6}\sigma \end{cases} \tag{1.77}$$

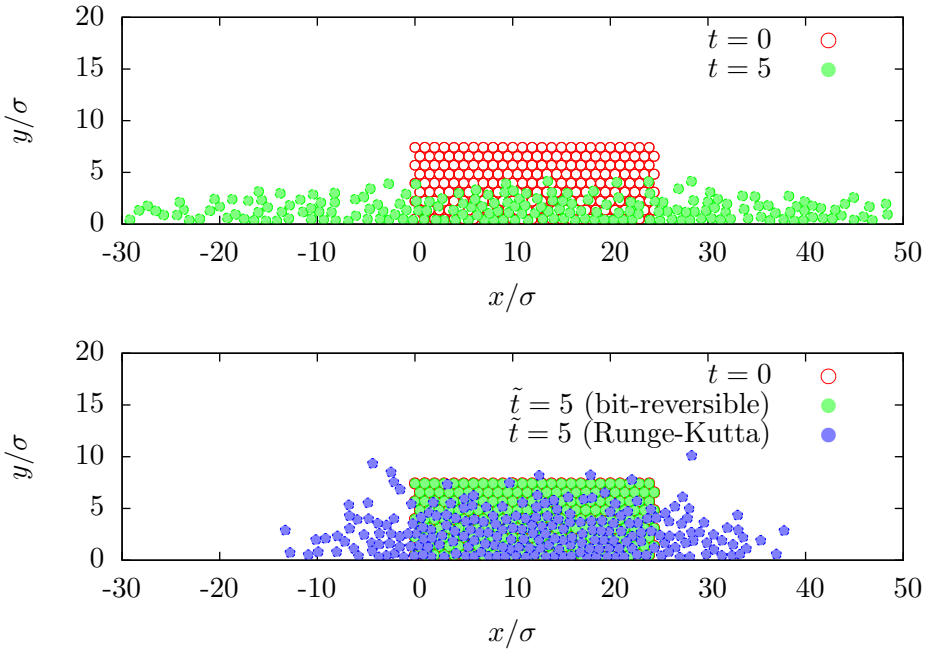

Figure 1.3: Stacked Lennard-Jones disks in a gravitational field. The top figure shows the initial close-packed arrangement at $t = 0$ and the state after 5120 steps with (1.75). Reversing all the momenta and letting the simulation run for another 5120 steps with (1.76) returns the system exactly to the initial state, as seen below. Runge-Kutta integration back from the reversed final state does not bring the system to the close-packed arrangement.

The potential creates a strong repulsion between disks when the distance between their centres, $r_{ij} = \|q_j - q_i\|$, dips below $\sqrt[6]{2}\sigma$. To generate figure 1.3, the disk masses were all set to unity. After a numerical integration with (1.75) for only 5120 steps ($\tau = 2^{-10}$), the momenta were all flipped and the simulation was allowed to retrace its steps with the reverse algorithm (1.76) and with a classic fourth-order Runge-Kutta. Despite the greater accuracy of Runge-Kutta integration[9], it could not recover the initial ordered state, which was in fact exactly where our bit-reversible scheme ended.

However precise the integration algorithm, if some small mismatch leaks in between the forward and the reverse step, then in most cases the dynamics will amplify it exponentially and lead to substantial differences between forward and reverse trajectories. Even numerical rounding causes irreversibility, as illustrated in Figure 1.4. A slightly longer 6144-step forward and reverse integration with a fourth-order Runge-Kutta algorithm led to a state quite different from the starting close-packed arrangement. Delambre's method (1.73), being time-reversible, fares much better, but it cannot recover the exact initial state either. The longer we integrate, the further apart we draw the initial and final reverse states calculated with Delambre's scheme.

1.6 The Maximum Entropy Formalism

> Jaynes's principle of maximum entropy determines the least-biased probability distribution over the microstates compatible with a given set of average values.

Macroscopic information does not determine a unique microstate, it only provides a set of possibilities. Seemingly identical systems from the macroscopic point of view will occupy different microstates, and a given system will often appear not to change while in fact hopping from one microstate to another.

[9] Runge-Kutta integration achieves a local error of order $o(\tau^5)$, to be compared with the $o(\tau^2)$ error of (1.75) and (1.76).

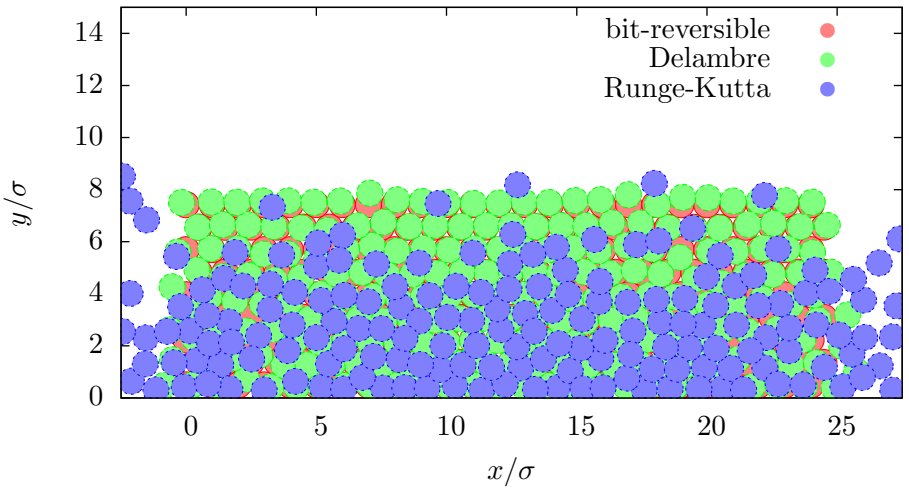

Figure 1.4: A short 6144-step numerical integration of the system shown in Figure 1.3 followed by inversion of momenta and a further 6144-step integration. The fourth order Runge-Kutta algorithm ends up quite far from the initial state. Delambre's time-reversible method (1.73) does recover a state similar (though not exactly equal) to the initial close-packed arrangement.

Not knowing where our system lies in phase space, we choose a probability distribution that represents the likelihood of finding it in different parts. However, we should realise that there is no such thing as the "real" distribution. Our choice depends on the information available[10]. The process that generates our macroscopic knowledge may contain no randomness whatsoever and only appear random *to us* due to our ignorance about the precise trajectories followed by the system.

At the end of Section 1.1, I pledged to justify the privileged status of relevant distributions seen in this chapter, yet without an objective probability density, why should we prefer some types of distributions to others?

In his classic 1957 article on *Information Theory and Statistical Mechanics* [29], E. T. Jaynes described the statistical mechanical problem of picking the best distribution as a search for the least-biased distribution compatible with some prior information. He posed the problem in the following terms: suppose we are told the average values $f_1,\ f_2,\ \ldots,\ f_k$ of a handful of functions $F_1,\ F_2,\ \ldots,\ F_k$ of the microstate z, and are then asked to estimate the average value of a different phase function G. How should we proceed? When viewed in this light, it looks like the problem has no solution. Without any knowledge of how the dynamics sample the microstates, *any* probability distribution compatible with the average values $f_1,\ f_2,\ \ldots,\ f_k$ could be accepted.

Most distributions will favour some microstates at the expense of others, but this bias can often be regarded as completely arbitrary, not justified by the available information. Therefore, argued Jaynes, we should choose the least-biased distribution compatible with the average values provided. Because the entropy functionals (1.23) and (1.25) measure uncertainty about the state of a system (see the explanation of equation (1.23)), the desired least-biased distribution should maximise the appropriate entropy, subject to the constraints on the expected values,

$$\langle F_i \rangle = \mathrm{Tr}[\rho\ F_i] = f_i, \tag{1.78}$$

[10] We can sometimes assume *ergodicity*, equating the probability density at z to the fraction of time spent by the system in a differential volume dv surrounding z, but only at equilibrium and not for every dynamical system [28].

and the fact that ρ should be normalised,

$$\mathrm{Tr}[\rho] = 1. \tag{1.79}$$

The trace Tr represents a sum (or an integral) over the set of microstates.

To determine the maximum entropy compatible with (1.78) and (1.79) we turn to the method of Lagrange multipliers. Thus, we must find the stationary value of a new functional,

$$C[\rho, \lambda, \mu] = -k_B \mathrm{Tr}\left[\rho \ \ln\left(\frac{\rho}{m}\right)\right] - k_B \sum_{i=1}^{k} \lambda_i (\mathrm{Tr}[\rho \ F_i] - f_i) \\ - k_B \mu (\mathrm{Tr}[\rho] - 1). \tag{1.80}$$

Differentiating (1.80) with respect to the Lagrange multipliers λ_i and μ and equating the result to zero, we recover the constraints (1.78) and (1.79). The derivative with respect to ρ (a variational derivative if we are dealing with a continuous set of microstates) yields the following result when it is set equal to zero and solved for ρ:

$$\rho(z) = m(z) \ e^{-\mu - 1 - \sum_{i=1}^{k} \lambda_i F_i(z)}. \tag{1.81}$$

In principle, we could now determine the Lagrange multipliers by plugging the expression for ρ into the constraints (1.78) and (1.79), and solving for λ_i and μ. The latter condition allows us to conclude that

$$e^{\mu + 1} = \mathrm{Tr}\left[m \ e^{-\sum_{i=1}^{k} \lambda_i F_i}\right] = Z. \tag{1.82}$$

The partition function Z, defined above, lies at the heart of equilibrium statistical mechanics. Thermodynamic quantities such as the the pressure and the specific heat can be calculated from Z by differentiation [30].

Jaynes's maximum entropy formalism supplies us with a generalised canonical distribution,

$$\bar{\rho}(z) = \frac{1}{Z} \ e^{-\sum_{i=1}^{k} \lambda_i F_i(z)}, \tag{1.83}$$

from which we can derive equilibrium statistical mechanics. We use the overbar to denote the solution of our variational problem. Note that (1.83) belongs to the class of relevant distributions (1.12), as it depends on z only through the values of the functions F_i. This justifies our focus on relevant distributions.

Putting (1.83) into the entropy functional, we get the entropy for the relevant distribution

$$S[\bar{\rho}] = k_B \sum_{i=1}^{k} \lambda_i f_i + k_B \ \ln(Z), \tag{1.84}$$

and we calculate the variances of the F_i from the partition function with a simple rule [29],

$$\Delta^2 F_i = \mathrm{Tr}[\bar{\rho}\ F_i^2] - f_i^2 = -\frac{\partial^2}{\partial \lambda_i^2} \ln(Z). \tag{1.85}$$

Furthermore, when the Hamiltonian depends on other parameters α_1, α_2, $\ldots$, α_l, then the estimated values of the derivatives of the expected values with respect to these parameters can also be calculated from Z,

$$\mathrm{Tr}\left[\frac{\partial F_i}{\partial \alpha_k}\right] = -\frac{1}{\lambda_i}\frac{\partial}{\partial \alpha_k} \ln(Z). \tag{1.86}$$

We will now use Jaynes's ideas to prove that entropy increases after coarse-graining unless the initial distribution coincides with the relevant coarse-grained distribution, in which case the entropy remains the same. Let us return to the random walk example in Section 1.2, just to have a concrete picture in mind. Imagine the x axis divided into bins b_i ($i = 1,\ 2,\ \ldots$) and suppose that we only know the probability P_i of finding the walker in each bin. We identify the bins by means of functions B_i,

$$B_i(x) = \begin{cases} 1, & x \in b_i, \\ 0, & x \notin b_i. \end{cases} \tag{1.87}$$

The corresponding constraint on the unknown distribution reads

$$\mathrm{Tr}[\rho\ B_i] = P_i. \tag{1.88}$$

If we maximise the entropy subject to this condition, we obtain

$$\rho(x) = e^{-1-\sum_i \mu_i B_i(x)}. \tag{1.89}$$

We can work out the expression for μ_i by inserting ρ into the constraint (1.88),

$$e^{-1-\mu_i} = \frac{P_i}{\mathrm{Tr}[B_i]}. \tag{1.90}$$

Not surprisingly, we finally reach the relevant distribution that assigns equal probabilities to all the states that belong to the same bin,

$$\bar{\rho}(x) = \frac{P_i}{\mathrm{Tr}[B_i]};\ x \in b_i. \tag{1.91}$$

The bar over ρ serves to distinguish (1.91) from the analytic solution of the stochastic process (1.22). The Boltzmann entropy for macrostate b_i follows from choosing $P_i = 1$ and all the other probabilities P_j equal to zero ($j \neq i$). This justifies the equal probability assumption (1.27) carried out in Section 1.3.

Comparing the entropy of (1.22) with that of (1.91), we see that

$$S[\rho] = \sum_i \mathrm{Tr}[\rho\ \ln(\rho)\ B_i] \leq \sum_i \mathrm{Tr}[\bar{\rho}\ \ln(\bar{\rho})\ B_i] = S[\bar{\rho}], \tag{1.92}$$

because the trace of $\rho\ \ln(\rho)\ B_i$ attains its maximum value subject to (1.88) when $\rho = \bar{\rho}$ for every x in b_i.

Sometimes the macrostates x_f form a continuous set with a known probability density function $P_F(f)$. In that case, we have a constraint on ρ for every possible value of f,

$$\mathrm{Tr}[\rho\ \delta(F - f)] = P_F(f). \tag{1.93}$$

Adding these constraints to the entropy creates the functional

$$C = -k_B \ \mathrm{Tr}\left[\rho \ \ln\left(\frac{\rho}{m}\right)\right] - k_B \int \mu(f)(\mathrm{Tr}[\rho \ \delta(F-f)] - P_F(f)) \ df. \quad (1.94)$$

Solving $\delta C = 0$, we encounter the expected relevant distribution [12],

$$\bar{\rho}(z) = \frac{P_F(F(z))}{\Omega(x_{F(z)})}. \quad (1.95)$$

Faced with uncertainty about the actual probabilities from which our information emerges, the relevant distribution assigns honest weights to the set of microstates. This choice of distribution may well be the best we can do, given the circumstances, but is it any good? To answer this question, let us reflect on the process that generated the average values, f_i.

In practice, the f_i result from measurements. In many cases, even single measurements take very long on an atomic time scale, so the result comes out as an average over a huge set of states. The law of large numbers guarantees that sample means converge towards the expected values of distributions as the number of independent samples increases [13]. When we possess high quality measurements, we may reasonably assume that our average values coincide with the expected value of the measured quantity [29]. Consequently, we ask ourselves what set of possible measurements makes the values f_i the most likely outcome [31] or, in other words, which probability distribution renders our measurement results least surprising?

We direct our attention to a system with N states, $z_1, \ z_2, \ \ldots, \ z_N$. M represents a very large number of independent measurements, out of which m_1 found the system in state z_1, m_2 in state z_2, and so on. The m_j remain unknown, though they are constrained by the fact that they must add to M,

$$\sum_j m_j = M, \quad (1.96)$$

and they must determine the measured average values according to

$$\sum_j \frac{m_j}{M} \ F_i(z_j) = f_i. \quad (1.97)$$

Out of the N^M possible sequences measurements, the multinomial distribution gives us the probability of finding the particular values of $m_1,\ m2,\ \ldots,\ m_N$,

$$P(m_1,\ m_2,\ \ldots,\ m_N) = \frac{N!}{m_1!m_2!\cdots m_N!N^M}. \tag{1.98}$$

Instead of maximising (1.98) subject to (1.96) and (1.97), we will find it more convenient to carry out the equivalent calculation by maximising the logarithm of P divided by N. Assisted by Lagrange's method, we solve

$$d\left(\frac{1}{N}\ \ln(P) - \sum_{i=1}^{k}\lambda_i\left(\sum_{j=1}^{N} g_j\ F_i(z_j) - f_i\right) - \mu\left(\sum_{j=1}^{N} g_j - 1\right)\right) = 0, \tag{1.99}$$

where we treat the fractions $g_j = m_j/M$ as if they were continuous variables. We will take advantage of the large N by using Stirling's approximation for the logarithm of P,

$$\begin{aligned}\ln(P) &\approx -\sum_{j=1}^{N} m_j\ \ln(m_j) + (N-M)\ln(N)\\ &= -N\sum_{j=1}^{N} g_j\ \ln(g_j) + -M\ln(N),\end{aligned} \tag{1.100}$$

which makes (1.99) equivalent to maximising the Shannon entropy. We could then consider infinite and continuous sets of states as limiting cases of the problem discussed here.

This second justification of the principle of maximum entropy gives the impression of being objective, in the sense that it depends on the physical behaviour of the system instead of our knowledge. Note, though, that we have had to make an equal probability assumption regarding the microstates in order to write (1.98), so we have not escaped the subjective interpretation of probability completely. Assigning the same weight to every state, in accordance with Laplace's principle of insufficient reason, amounts to maximising the entropy in a situation without any prior information [29].

The usual formulae for ensembles presented in statistical mechanics textbooks follow immediately from the principles laid out in this chapter. Take the constant energy case, for example, with an isolated system characterised by an autonomous Hamiltonian $H(z) = E$. The probability distribution over the values of $H(z)$,

$$P_H(e) = \delta(E - e), \tag{1.101}$$

implies a simple relevant distribution (1.95) in this case, known as the *microcanonical ensemble*,

$$\bar{\rho}(z) = \frac{\delta(E - H(z))}{\Omega(x_{H(z)})}. \tag{1.102}$$

The entropy functional therefore equals the Boltzmann entropy[11],

$$\begin{aligned} S[\bar{\rho}] &= -k_B \int_{x_E} \frac{1}{\Omega(x_E)} \ \ln\left(\frac{1}{\Omega(x_E)}\right) \ dx_E \\ &= k_B \ \ln(\Omega(x_E)) = S_B(x_E), \end{aligned} \tag{1.103}$$

and we define temperature as

$$T = \frac{1}{k_B}\left(\frac{\partial S_B}{\partial E}\right)^{-1}. \tag{1.104}$$

Although our definition of temperature (1.104) looks the same as the classic thermodynamic relation between temperature and entropy, the concepts differ. In the 19th century, Carnot imagined heat transfer by analogy with the flow in an overshot water wheel, with heat dropping from higher to lower temperatures, never moving uphill [32]. Thermal engines were placed between reservoirs at different "heights" and driven by the falling

[11] In (1.103), we carry out the integral only over the states in x_E. If we insert equation (1.102) directly into the definition of the entropy, we face the logarithm of a Dirac delta function. One way out of this problem is to replace the delta function with a normalised Gaussian distribution and then make the standard deviation tend to zero.

heat. Clausius then showed that the spontaneous flow changed the value of a mysterious state function that he later named *entropy* [33]. He claimed that entropy changes were proportional to the amount of heat absorbed Q, and inversely related to the system's temperature T. Heat transferred to a source at temperature T_1 from another at temperature T_2 would produce a total variation equal to

$$\Delta S = \frac{Q}{T_1} - \frac{Q}{T_2}. \tag{1.105}$$

Because heat always flows towards cooler bodies, $\Delta S \geq 0$.

Clausius assigned entropies to *thermodynamic* states, that is to say, to systems in equilibrium. In contrast, information theory allows us to calculate the entropy for *any* description of a system, whether or not the system has reached equilibrium. In the 19th century, temperature was the primitive notion from which entropy was derived, whereas in our presentation entropy plays the leading role, and temperature measures the rate at which the number of accessible states grows with increasing energy.

From the microcanonical ensemble (1.102), we can guess the distribution for a system that *can* exchange energy with its environment, from two simplifying assumptions. First, we describe the system plus its environment (which we will refer to as the heat bath) as a larger isolated compound; and, second, we suppose that our system's microstate, z_S, is independent of the bath's microstate, z_B, so that the probability density for the compound microstate $z = (z_S,\ z_B)$ equals the product $\rho(z) = \rho_S(z_S)\ \rho_B(z_B)$, and its energy equals a sum of two independent Hamiltonians, $H(z) = H_S(z_S) + H_B(z_B)$. The latter assumption amounts to claiming that we can neglect the interaction energy between the system and its environment, $V(z_S,\ z_B) \approx 0$.

The desired distribution ρ_S results from integrating the microcanonical ρ over the accessible z_B,

$$\rho_S(z_S) = \int \rho(z)\ dz_B = \frac{\Omega(y_{E-H_S(z_S)})}{\Omega(x_{H(z)})}. \tag{1.106}$$

The macrostate y_ϵ above stands for the set $\{z_B : H_B(z_B) = \epsilon\}$. Now, for any z that satisfied $H(z) = E$,

$$\rho(z) = \frac{\Omega(y_{E-H_S(z_S)})\ \Omega(y_{E-H_B(z_B)})}{(\Omega(x_{H(z)}))^2} = \frac{1}{\Omega(x_{H(z)})}, \tag{1.107}$$

which implies that the product in the numerator equals $\Omega(x_{H(z)})$. This observation suggests that the densities of states may depend exponentially on the energy,

$$\begin{aligned}\Omega(y_{E-H_S(z_S)})\ \Omega(y_{E-H_B(z_B)}) &\propto e^{\beta(E-H_S(z_S))}e^{\beta(E-H_B(z_B))} \\ &= e^{\beta E} \propto \Omega(x_E),\end{aligned} \tag{1.108}$$

where $\beta = (k_B T)^{-1}$, as we can see from calculating the derivative $\partial S[\rho]/\partial E$ and applying (1.104),

$$\frac{1}{T} = \frac{\partial S[\rho]}{\partial E} = \frac{\partial}{\partial E}(k_B\ \ln(\Omega(x_E))) = k_B\beta. \tag{1.109}$$

Leaving the guesswork to one side, we concentrate on the expected energies E_S and E_B, for the system and the bath, respectively. If only we knew the values of E_S and E_B, we could apply the principle of maximum entropy to the system and bath separately to obtain

$$\rho_S(z_S) = \frac{1}{Z_S}\ e^{-\beta_S H_S(z_S)}, \tag{1.110}$$

$$\rho_B(z_B) = \frac{1}{Z_B}\ e^{-\beta_B H_B(z_B)}, \tag{1.111}$$

where Z_S and Z_B can be thought of as the appropriate normalisation factors. We would then work out the Lagrange multipliers from the constraints on the average energies. But using the fact that, for all positive values of ρ, the probability density for the compound system factors into the product of ρ_S times ρ_B, we write

$$\rho(z) = \rho_S(z_S)\ \rho_B(z_B) = \frac{1}{Z_S Z_B}\ e^{-\beta_S H_S(z_S)-\beta_B H_B(z_B)} = \frac{1}{\Omega(x_E)}. \tag{1.112}$$

The final expression on the right does not depend on z_S or z_B, which means that the exponent $-\beta_S H_S(z_S) - \beta_B H_B(z_B)$ equals some constant. Remember that H_S and H_B could be any Hamiltonian functions, so we must conclude that $\beta_S = \beta_B = \beta$, as

$$-\beta\ H_S(z_S) - \beta\ H_B(z_B) = -\beta\ H(z) = -\beta\ E. \tag{1.113}$$

Once again, the relation between the entropy of the compound system and the total energy E proves that $\beta = (k_B T)^{-1}$.

The foregoing discussion leads us to infer the expression for ρ_S, known as the *canonical ensemble*, which represents a system in equilibrium with a heat reservoir at temperature T.

$$\rho_S(z_S) = \frac{e^{-\frac{1}{k_B T} H_S(z_S)}}{\mathrm{Tr}\left[e^{-\frac{1}{k_B T} H_S(z_S)}\right]}. \tag{1.114}$$

Along the path that brought us to (1.114), at no point did we assume that our system had reached thermal equilibrium. Surely, we cannot apply this result to *nonequilibrium* systems, can we? When we view the problem from the point of view of information theory, we realise that, in fact, we can!

Suppose we have a system in contact with a heat reservoir at temperature T that satisfies the constraints for the average values of several phase functions $F_1,\ F_2,\ \ldots,\ F_k$ (1.78). We assume that the reservoir has a very large heat capacity, so that energy can flow in or out of the system without any detectable change in the temperature of the reservoir. Whether or not the system has reached equilibrium, the principle of maximum entropy implies the relevant distribution below, which generalises (1.114),

$$\rho(z) = \frac{e^{-\frac{1}{k_B T} H(z) - \sum_{i=1}^{k} \lambda_i F_i(z)}}{\mathrm{Tr}\left[e^{-\frac{1}{k_B T} H(z) - \sum_{i=1}^{k} \lambda_i F_i(z)}\right]}. \tag{1.115}$$

We will not prove (1.115) here because we will present a simple derivation

of this expression from the relative entropy functional in the next chapter[12].

1.7 Quantum-mechanical statistics

Jaynes's formalism applies also to quantum mechanics, but in this thesis we concentrate our attention on classical systems.

Up until now, we have deliberately avoided mentioning quantum phenomena. The technical details involved would have only obscured the presentation. In quantum mechanics, "microstates" that share the same values of the dynamical invariants do not necessarily share the same probability density. We must incorporate additional information (namely, the Schrödinger equation). Furthermore, we have to remember that permutations among identical particles do not count as different states. The line of reasoning remains unchanged, though.

The most convenient formulation of quantum statistical mechanics simply interprets ρ in our formulae for discrete sets of states as the density matrix, the trace becomes a summation over diagonal terms, and the Hamiltonian H and relevant variables $F_1, \ldots, F_k$ are read as Hermitian operators.

The values of the dynamical invariants define a pure quantum state completely. Therefore, a system in an energy level E_n with degeneracy g_n would have a microcanonical partition function $Z = g_n$. We would calculate the probability density at q from a set of g_n orthonormal wave functions $\psi_{i,n}(q)$ that satisfy $H\psi_{i,n} = E_n\ \psi_{i,n}$ with

$$P(q) = \frac{1}{g_n} \sum_{i=1}^{g_n} |\psi_{i,n}^2(q)|. \tag{1.116}$$

The tools of maximum entropy reveal that the generalised quantum-mechanical canonical distribution has the same form as in the classical case [35],

$$\rho = \frac{1}{Z} e^{-\sum_{i=1}^{k} \lambda_i F_i}. \tag{1.117}$$

[12] See our article in the *Journal of Statistical Physics* [34] for a derivation of this result from the principle of maximum entropy. For the relative entropy route see page 75.

Both ρ and the exponential should now be interpreted as linear operators. The partition function, defined as the diagonal sum

$$Z = \mathrm{Tr}\left[e^{-\sum_{i=1}^{k}\lambda_i F_i}\right], \tag{1.118}$$

becomes, in the common case where we only have the energy as a relevant variable,

$$Z = \sum_n \int e^{-\frac{1}{k_B T}E_n}\left(\sum_{i=1}^{g_n}|\psi_{i,n}^2(q)|\right)dq. \tag{1.119}$$

1.8 A simple illustration of irreversibility

We end this chapter with two numerical examples. They demonstrate that the application of reversible Hamiltonian laws to a typical microstate leads the system along a path of increasing entropy.

We have only considered very abstract properties of dynamical behaviour so far, but we can already make some concrete predictions about nonequilibrium processes. The paradigmatic thermodynamics textbook example includes a gas cylinder with diathermal walls and a piston on one end [36]. Compression and expansion should occur slowly, with the gas in mechanical and thermal equilibrium at all times. Suddenly plunging the piston inwards would create pressure waves and temperature gradients, and the gas would abandon equilibrium (and the field of classical thermodynamics).

Our investigation encourages us to compare the rapid expansion of the gas with a corresponding compression. First we pull the piston out and then allow the gas to reach the environment temperature. The whole process of expansion and relaxation takes some time t. Now the piston returns in reverse motion to its initial state. At time $2t$ the gas will not yet have equilibrated thermally. This is a simple consequence of equation (1.30). The overwhelming majority of expanded gas macrostates have a greater Boltzmann entropy than their compressed counterparts and, therefore, the reverse process to a lower Boltzmann entropy becomes extremely unlikely.

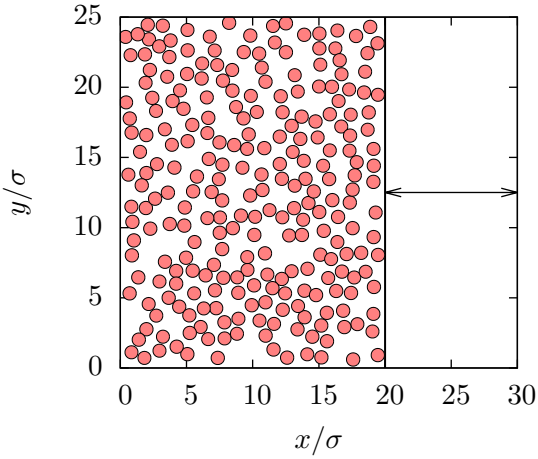

Figure 1.5: 250 disks interacting with each other and the walls through truncated Lennard-Jones potentials (1.77). The right wall moves in and out obeying equation (1.120).

In most realisations, the entropy will not decrease and, because the volume decreases for coordinates, the accessible states in phase space must increase in the directions of momenta, so we expect the gas to warm up. If we want to get the gas back to its initial state we must let it dissipate part of its energy as heat. This discussion suggests that dissipation limits the capacity of a system to reach equilibrium with a heat reservoir. The end of Chapter 2 will confirm this insight.

When we isolate the system of interest, it can no longer return to a lower Boltzmann entropy, unless some cosmic conspiracy brings it into one of the very few microstates on a trajectory of decreasing entropy. A two-dimensional thought experiment will illustrate the point.

We represent the walls and piston with a repulsive potential that isolates the gas particles (Figure 1.5). If the piston moves in and out quickly several times before returning to rest at its original position, the gas will heat up.

We assume that the system reaches the state of equilibrium before the piston stops moving and after it stops. The gas comes from a source at temperature T and we measure the final state temperature T' by bringing the system into contact with a thermometer, so we can represent the initial and final distributions with canonical ensembles (1.114) [34].

In the numerical representation of this setup, I followed the trajectories of 250 isolated unit mass disks with truncated Lennard-Jones potentials (1.77) using the bit-reversible scheme (1.75). The time of the simulation run, t_s, was divided into four parts. During the first and last quarter, the piston remained at rest. For the rest of the simulation, the piston's position x oscillated according to

$$x(t) = 25\sigma + 5\sigma \ \cos\left(\frac{70\pi \ t}{t_s}\right). \tag{1.120}$$

Note that the trajectory $x(t)$ is unaffected by time inversion. As expected, the temperature increased during the simulation run ($k_B\Delta T = 1.5$, see Figure 1.6).

Of course, the reversed trajectory beginning at the final state with all the momenta inverted would also be perfectly acceptable from the point of view of the Hamiltonian laws of motion. In *that* particular process, the system would cool down. The reason why we never observe such cooling is that we choose the initial microstate randomly from the canonical ensemble, and the overwhelming majority of these states lead to a final higher-energy microstate.

In the previous example, the initial and final states have the same appearance in configuration space. The difference lies in the velocity distributions. Sometimes we can see how entropy grows in configuration space. Imagine a dumbwaiter hanging in a gravitational field suspended from a massless inextensible rope attached to an equally massless pulley. We let go of the rope and the dumbwaiter goes into free fall until it reaches a certain speed. From then on, we hold the rope again and the system stops accelerating and lowers at constant speed. Inside the dumbwaiter, we find a system of 250 disks that interact with each other and the walls through

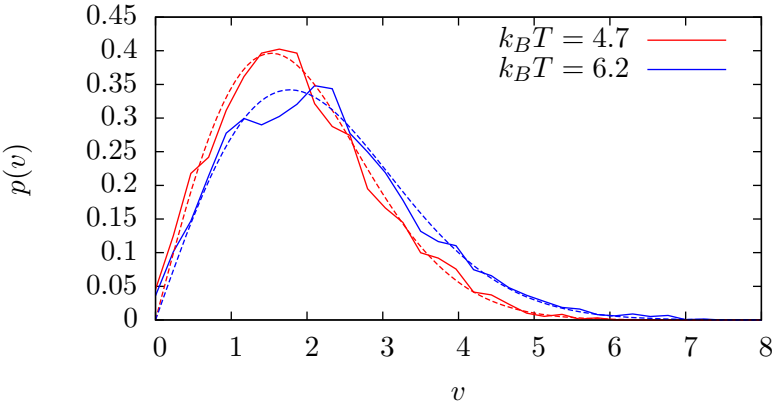

Figure 1.6: Velocity distributions for the disks before (*blue*) and after (*red*) the oscillation of the piston. The dashed lines show the theoretical Maxwell-Boltzmann distributions for the initial and final temperatures.

Lennard-Jones truncated potentials (1.77), $\epsilon = 1$. From the point of view of a reference frame moving with the system, the process described amounts to switching off the external field and then turning it back on again.

In the simulations, shown in Figure 1.7, the box width and height were chosen equal to 25 σ, the masses were set equal to one and the external gravitational field was also given a unit value (which we could interpret as a rescaling of the time axis). The numerical integrations (1.75) ran for 5.12×10^5 time steps of length $\tau = 2^{-8}$. We activated the external field only during the first and last quarters of the run. This means that the dynamical laws and our "macroscopic" action on the system were both time-reversible.

The disks began in a close-packed arrangement and quickly relaxed to a state displaying long-range structure at the bottom of the box, with disorder rising with height. Later, without gravity, the system behaved like a gas, filling the box and mixing. When we switched the field back on, we did not recover the configurations seen at the beginning of the run.

Never mind the forces on the disks, our initial description did not appear

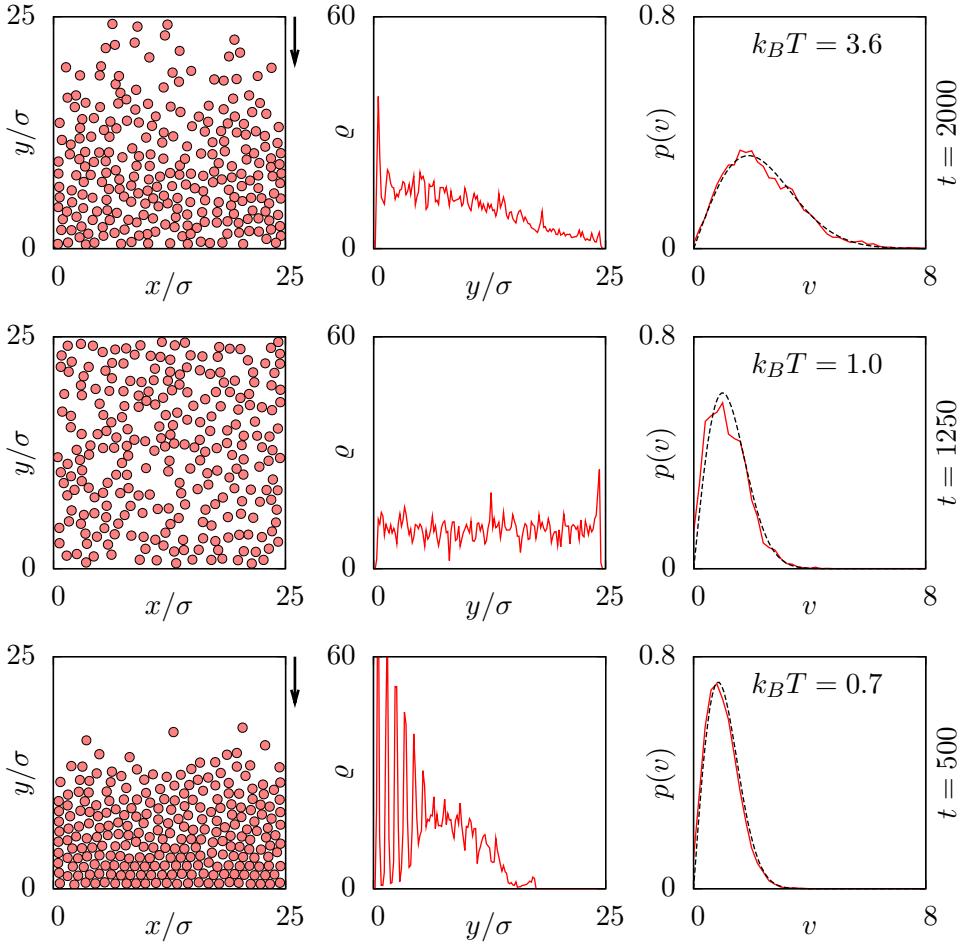

Figure 1.7: Hard disks in the dumbwaiter thought experiment. Rows represent different times. The left column displays the positions of the disks (an arrow indicates the presence of an external field). The central column shows the mass densities ϱ as a function of the y coordinate. The velocity distributions on the right lie reasonably close to the Maxwell-Boltzmann distribution (dashed line) for the corresponding temperatures.

symmetrical. When we dropped the dumbwaiter, some of the gravitational potential energy was transformed into internal energy. On further reflection, we see that we could have set up the experiment the other way round. In this equivalent version, the dumbwaiter moves upwards at constant speed until we launch it into the air and catch it at the top of its trajectory. Then we would say that *kinetic* energy had transformed into internal energy.

Whichever direction we select, an initial "cooler" arrangement of disks will end up in a "warmer" state. Figure 1.7 displays the numerical results. Switching off the external field liberates the Lennard-Jones potential energy of the disks compressed at the bottom of the box, which becomes kinetic energy. This makes the temperature rise slightly, from $k_BT = 0.7$ to $k_BT = 1.0$ (centre row). When the field returns, many disks suddenly acquire a large potential energy, leading to an even higher temperature final state $k_BT = 3.6$ (top row).

In the extremely unlikely case of turning the field on and finding the disks exactly where they were when the field disappeared but with the velocities reversed, the system would proceed backwards to the initial close-packed state.

I find it captivating that reducing irreversibility to a statistical statement about many-body phenomena has the astonishing consequence of rendering the principle of causality symmetrical in time. With reversible laws, cause and effect exchange roles when we change the direction of time, so what grounds do we have for breaking this symmetry between them? If we lived our lives in reverse motion, as in Damon Knight's *This way to the Regress* [37], our idea of natural causal links would differ considerably. With his main character, Sullivan, getting younger and younger, Knight wrote:

> Even in winter, he stood about, watching the freezing ground water rush up the drainspout, or letting the snow form on his head and shoulders, drifting up to the white sky as it did from the ground on which he stood.
>
> Whatever came, he took without question. If his fingers and

nose were bright pink with cold when he went out, the snow would warm them.

1.9 Summary

Statistical laws involve challenging conceptual issues for scientific explanation, especially in quantum mechanics[13], but here we have concentrated on how to infer macroscopic properties from the laws that govern the microscopic constituents of a system, disregarding the problem of how to interpret the probability distributions calculated.

The main difficulty in nonequilibrium statistical mechanics emerges from the huge number of accessible microstates. The future state of a system depends on the exact present microstate, which we can never determine in practice. Even if we could, the vast number of degrees of freedom would prevent us from following the trajectory for very long. In truth, we are not very interested in finding out which microstate the system occupies. We wish to focus on the stable macroscopic properties. A complete list of coordinates and momenta does not establish the temperature, the speed of sound or whether the system has melted, without further calculation. Though I will probably never observe my cup of tea twice in the same microstate, I will still enjoy its scent of bergamot orange, which I always perceive in spite of the constant buzzing and clashing of atoms.

Taking for granted the time reversibility of the fundamental laws, we have shown that coarse-graining predicts the emergence of irreversibility and memory. Furthermore, if a process is symmetrical, stationary and irreducible, then detailed balance between macrostates follows, and the system evolves towards a well-defined macroscopic equilibrium state (which corresponds to the microscopic uniform distribution compatible with the values of the dynamical invariants). Coarse-grained descriptions satisfy the definition of Markov processes when the transition probabilities depend only on the previous macrostate, as in the case where we choose the set of dynam-

[13]See, for example, my book review [39] on Rolleri's *Probabilidad, causalidad y explicación* [38].

ical invariants as relevant variables. However, if we restrict our attention to processes among states described by relevant distributions, then we can define memoryless transition probabilities between macrostates. Relevant distributions represent the least-biased probability assignments compatible with our information about a system [29].

In this chapter we propose to measure irreversibility in a process connecting two macrostates by dividing the probability of observing the direct process divided by the probability of observing the corresponding reverse process, that is, as a function of the difference between the entropies of the end states. Although Hamiltonian systems do not satisfy the symmetry that allows us to define reversible Markov processes, Hamiltonian reversibility extends the validity of our formulae to the realm of classical mechanics.

To test the predictions of a theory, we usually turn first to molecular dynamics simulations. We carry out numerical integration of the equations of motion by discretising time and phase space. Despite this apparent fundamental difference between the discrete computer simulations and the continuous (and usually unknown) analytic solutions of Hamilton's equations, numerical results correctly represent the behaviour of systems. To defend this claim, we have developed an original memoryless symplectic bit-reversible algorithm, inspired by the work of Levesque and Verlet [27], who provided a bit-reversible algorithm that was neither memoryless nor symplectic. Given a Hamiltonian function, our method provides an exact solution to a closely related Hamiltonian problem, obtained by a slight perturbation of the original Hamiltonian function. As the time step tends to zero, the perturbed problem approaches the original problem asymptotically. Moreover, because every single bit in a trajectory can be traced flawlessly back and forth in time, our scheme proves that observed irreversibility need not originate in some asymmetry of the underlying equations of motion.

If macroscopic systems were always adequately characterised by relevant distributions, there would be no need to continue. Knowing the expected values f_i as functions of time, we could calculate all the properties of a system. Simulating a realisation would reduce to picking a random state

for each time from the corresponding relevant distribution until the system reached the equilibrium ensemble. In Chapter 3, we will show that working out the time dependence of the f_i numerically would also constitute a relatively easy task in this case. Sadly, relevant distributions do not generally evolve into other relevant distributions, so we still have plenty of work ahead.

Chapter 2

Relative Entropy

We prove that the maximum relative entropy variational principle coincides with Jaynes's maximum entropy formalism under certain conditions. The distributions obtained with the former approach allow us to extend numerical sampling results to nearby values of the simulation parameters.

Heat and work feel so different to us (compare lying in the sun with swimming across a pool) that it should not surprise us it took so long to understand both these concepts in terms of energy transfer. Paradoxically, when we zoom in on the atomic level, we find it difficult to tell them apart. Any increase in the energy corresponds to the work of some force acting on the system.

Back in the macroscopic world, the first law of thermodynamics states that internal energy changes as a result of work *and* heat flows. In differential form,

$$dE = đQ + đW, \tag{2.1}$$

where $đ$ indicates an infinitesimal quantity that does not correspond to the differential of any function. Is heat a mere illusion arising from our blurred image of the precise microstates?

Within the context of engineering, we find a clear-cut distinction between work and heat. We identify the former with the energy change coming from a variation in the control parameters, such as the volume in a cylinder or the charge in a capacitor. Heat implies energy leaking in or out of the system by other means. Callen's pond analogy illustrates the point clearly[1]: by purchasing flowmeters, a gardener can control how much water he pours into a pond and how much he drains from it, but the water level may also change as a result of rainfall or evaporation. Similarly, we classify energy "poured" into a system in a controlled way as work, and refer to the energy dissipated or absorbed by the system through unknown microscopic interactions with the environment as heat.

For theoretical physicists, control parameters give way to relevant variables, which at least in principle may be as fine-grained as we wish. The energy exchanged through *unknown* microscopic mechanisms depends of course on how detailed we have made our description of the system, so the information theory approach defines heat relative to a choice of relevant variables.

In statistical physics, energy stands out as the prime magnitude determining the set of accessible microstates. Classical Hamiltonian systems trace out well defined trajectories in phase space. If we know how the system energy depends on time, then its evolution is determined in principle by the initial state. The dynamics then carry the initial probability densities along the trajectories. But if we leave some interactions out of the picture, then we must deal with an extra source of uncertainty besides the lack of knowledge about the precise initial state. Heat transfer causes our system to skip from one path in phase space to another at random instants. This implies that we need to quantify the effect of heat flows if we wish to predict how a system will behave away from equilibrium.

[1]See chapter 1, Section 7 of Callen's classic text on thermodynamics [40].

2.1 The sampling problem

In nonequilibrium processes, the macroscopic properties depend on the changing values of the relevant variables. Because we usually determine the macroscopic properties by sampling very long trajectories calculated with molecular dynamics, we generally have to run a simulation for each combination of values of the relevant variables. This usually exceeds our computational power.

Heat transfer depends on the nature of the interaction between a system and its environment, in addition to their precise microstates. We do not usually encounter exactly solvable models of heat exchange, so we often calculate it directly by means of simulations, using several runs to work out average behaviour. This chapter deals with the problem of sampling nonequilibrium distributions. To understand the difficulties involved, consider first the equilibrium ensemble.

Liouville's equation [18] states that the probability distribution ρ for a classical (isolated) Hamiltonian system evolves according to (cf. (1.58))

$$\frac{\partial \rho}{\partial t} = \{H,\ \rho\}, \tag{2.2}$$

and we have already explained that mixing dynamics cause arbitrary distributions to approach equilibrium eventually, in the "volume average" sense already discussed in last chapter's Section 1.4. Let ρ^{eq} represent the final equilibrium ensemble, which must be stationary,

$$\frac{\partial \rho^{eq}}{\partial t} = 0. \tag{2.3}$$

The microstate z changes in time, so making ρ^{eq} stationary involves having it depend on the microstates through functions that *do not* depend on time. Let I represent the set of dynamical invariants, such as the energy and linear momentum, for example. When different microstates z and z' share the same values of the dynamical invariants, $I(z) = I(z')$, then they have

equal probability densities, $\rho(z) = \rho(z')$. This ensures that the probability along a given dynamical trajectory will remain constant. In other words, the equilibrium probabilities should only depend on the values of the dynamical invariants, $\rho^{eq}(z) = \phi(I(z))$. By writing down the probability of $I(z) = i$,

$$P_I(i) = \mathrm{Tr}[\rho^{eq}\delta(I - i)] = \phi(i)\ \mathrm{Tr}[\delta(I - i)], \tag{2.4}$$

we see that we can express ϕ as a function of the probability $P_I(i)$ and the density of states,

$$\phi(i) = \frac{P_I(i)}{\mathrm{Tr}[\delta(I - i)]}. \tag{2.5}$$

The equilibrium ensemble must therefore have the following relevant distribution form [12]:

$$\rho^{eq}(z) = m\ \frac{P(I(z))}{\Omega(x_{I(z)})}. \tag{2.6}$$

We introduce the m factor to redefine Ω as a dimensionless quantity

$$\Omega(x_i) = \int_\Gamma m\ \delta(I - i)\ dz, \tag{2.7}$$

with Γ standing for the accessible region of phase space and m having the same dimensions as dz.

To calculate expected values with molecular dynamics, we simply start with a few random initial states drawn from the equilibrium ensemble and use a discrete version of Hamilton's equations to follow their motion in phase space. Long simulations ensure that the system has enough time to explore the accessible part of phase space, while choosing several initial states at random avoids trajectories trapped in a subregion of phase space due to additional unknown dynamical invariants or an effective lack of ergodicity over the simulated time scales.

Sampling nonequilibrium distributions is a very different game. For starters, we rarely know how a probability density function will move around in time. Hence, the most straightforward approach begins with

a large set of microstates chosen with the initial distribution and then integrates the motion numerically for each of them. Because the expected values will generally change in time, we should calculate averages by taking a single microstate from each trajectory for each instant. This means that we will have to compute many more trajectories than in the equilibrium case. Furthermore, many macroscopic magnitudes of interest change significantly only when we consider very large time scales compared to the molecular collision times that we simulate. In other words, nonequilibrium sampling typically involves many more and much longer simulations, with the necessary computing power greatly exceeding our present capabilities.

Clearly, we need a shortcut. It would be nice to determine nonequilibrium magnitudes from equilibrium sampling. Callen and Welton's 1951 breakthrough fluctuation-dissipation theorem [41] established precisely this connection. With it, they linked the generalised resistance in linear dissipative systems to the spontaneous fluctuation of intensive variables, or "thermodynamic forces". Dissipation was recast in terms of autocorrelation averages of relevant variables. Green [42] and Kubo [43] then worked out the equations of motion for sets of relevant average values. The expressions included the linear dissipation averages as coefficients. Fluctuation-dissipation relations follow most easily from examining the conditions for equilibrium with the Fokker-Planck equation, as we will see in the next chapter (Section 3.7).

Numerical calculations of Green-Kubo coefficients proceed by working out the dissipation as a function of the thermodynamic forces. Equilibrium simulations run for each representative set of relevant variables values within the ranges of interest. The expected values must change slowly over the simulated molecular collision time scales. Otherwise, they will drift away and the simulation results will not count as an average over states with a set value of the thermodynamic forces.

Two problems confront the Green-Kubo *linear* response theory. First, Callen and Welton's proof assumed that an external perturbation brought the system out of equilibrium. Interpreting a nonequilibrium state as the response of a system to a perturbation in the thermodynamic forces fits many cases (think of a particle slowed down by viscous drag, or energy

dissipated in an electrical resistor, for example), but it certainly does not apply to all. Some experiments we set up in such a way that the system starts out in a small region of phase space, from which the trajectories then spread out. Tuning the values of intensive parameters only has the effect of changing the probabilistic weights assigned to microstates by the equilibrium ensemble. To make some microstates impossible, we have to associate an infinite energy with them, but then perturbation theory breaks down. In this sense, pouring cold milk into hot tea does not resemble leaving the cup of milky tea in an external temperature field of some kind.

The other problem arises when we have several relevant variables. With k variables and m representative values per variable, we find that we need to run m^k equilibrium simulations. The number grows very rapidly with increasing m and k and, to use Richard Bellman's expression, "the curse of dimensionality" befalls us [44].

2.2 The wandering king

The wandering king simulation illustrates the fluctuation relation (1.30) in the context of a reversible Markov process.

Before we delve into the issue of sampling, I should mention that not all types of nonequilibrium calculations involve intensive computational power. Fluctuation theorems, first conjectured by Evans, Cohen and Morriss [45], and then demonstrated by Evans and Searles [46], apply even far away from equilibrium. From the probability of observing a spontaneous process, they allow us to calculate, for example, the probability of the corresponding reverse process. The likelihood of these "second law violations" was initially worked out for steady-state shearing fluids and then demonstrated experimentally for a colloidal particle in an optical trap [47].

The fluctuation theorem (1.30) derived in the previous chapter does not require equilibrium probabilities either. In fact, it applies well beyond the realm of physics to any abstract reversible Markov process. We will now discuss a specific example.

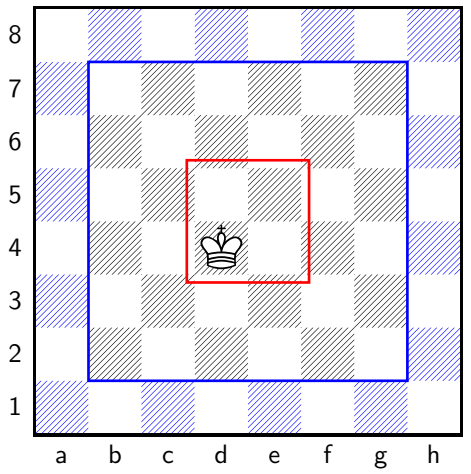

Figure 2.1: The wandering king. Beginning at d4, the king goes for a random walk around the board with a probability $1/9$ of moving to each adjacent square (including diagonals). Figure 2.2 compares the transition probability from the four inner squares (red box) to the edge of the board (marked in blue).

A wandering chess king, with probability $1/9$ of moving to each adjacent square starts a random walk near the centre of the board (I propose to name this game "dull chess", see Figure 2.1). Note that the probability of any string of steps equals that of the reverse sequence conditioned on their corresponding starting squares. We ask ourselves about the probability of a transition from the central four squares (marked with a red box) to the outer edge of the board, and we wish to compare it with the reverse transition probability. According to equation (1.30), the logarithm of the ratio of these quantities equals the difference between their Boltzmann entropies. Let x_r represent the centre of the board and x_b represent the edge.

$$\ln\left(\frac{P(b \mid r)}{P(r \mid b)}\right) = \ln(7). \tag{2.8}$$

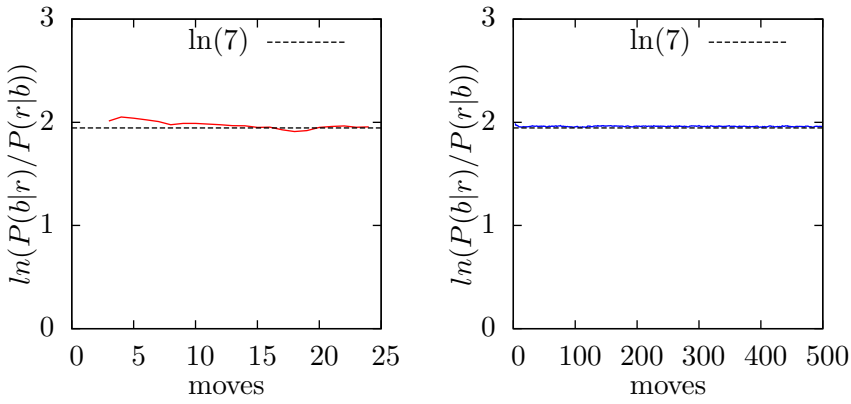

Figure 2.2: Irreversibility in the transition from the inner macrostate x_r to the outer macrostate x_b in the wandering king simulations, versus the number of moves in the process. We measure irreversibility as the logarithm of the direct transition probability, $P(b \mid r)$, divided by the reverse, $P(r \mid b)$. The plot on the left shows an average over ten thousand fifty-step walks. On the right, we have drawn the equilibrium version (ten thousand realisations of a thousand-step walk).

We simulate fifty-step walks. During this time, the probability distribution for the king's location does not have enough time to reach the stationary state (with the most likely square doubling the probability of the least likely spot). Still and all, results do not vary significantly compared to the healthier one thousand steps, which also confirm our ln(7) prediction.

Interestingly, we get the same results when we change the king's initial position and the shape of the initial and final macrostates (although statistical errors seem to have a significant effect on the fifty-step walk for some configurations). In addition, the length of the transition has no significant influence on the result, as long as the king can get from one macrostate to the other in the designated number of steps (see Figure 2.2).

Having confirmed our theorem in a concrete setting, let us now return to the subject of physical phenomena in nature.

2.3 Thermodynamic processes

The fluctuation relation (1.30) implies two well-known results: the work inequality and the Jarzynski equality. Dissipated work can also be expressed as an average logratio of probabilities.

In an infinitesimal process where the expected values change from f_i to $f_i + (df_i/dt)\ dt$ while the phase functions change by $(\partial F_i/\partial t)\ dt$, the variation of the entropy (1.84) equals

$$\begin{aligned} dS &= \sum_{i=1}^{k} \frac{\partial S}{\partial f_i}\frac{\partial f_i}{\partial t}\ dt + \frac{\partial S}{\partial t}\ dt \\ &= k_B \sum_{i=1}^{k} \frac{\partial f_i}{\partial t}\ dt\ \frac{\partial}{\partial t}(\lambda_i f_i) + k_B\ dt\ \frac{\partial}{\partial t}\ln(Z) \\ &= k_B \sum_{i=1}^{k} \lambda_i \left(\frac{\partial f_i}{\partial t} - \left\langle \frac{\partial F_i}{\partial t} \right\rangle \right)\ dt. \end{aligned} \tag{2.9}$$

This allows us to define the heat associated with λ_i as [29]

$$đQ_i = \frac{\partial f_i}{\partial t}\ dt - \left\langle \frac{\partial F_i}{\partial t} \right\rangle\ dt = df_i - \langle dF_i \rangle\,. \tag{2.10}$$

The heat paired with $(k_B T)^{-1}$, which we will write as Q, without a subindex, corresponds to the ordinary meaning of heat. The angle brackets indicate average values,

$$\langle F \rangle = \mathrm{Tr}[\rho\ F]. \tag{2.11}$$

Work indicates an energy difference caused by an alteration of the Hamiltonian function,

$$W = \int \frac{\partial}{\partial t} H(z(t),\ t)\ dt, \tag{2.12}$$

which means that the heat Q (2.10) in the ordinary sense equals the change in the energy minus the expected value of the work.

In thermodynamics, we choose a single relevant variable: the internal energy, which corresponds to the expected value of the Hamiltonian. Consider a system in contact with a heat reservoir at temperature T, described by the canonical ensemble (1.114). An infinitesimal process in which the average energy increases by a small amount $\Delta \langle H \rangle = \Delta E$ would cause the entropy (1.84) to change by

$$S_f - S_i = \frac{\Delta E}{T} + k_B \ \Delta \ln(Z). \tag{2.13}$$

Chapter 1 argued that entropy should remain constant in macroscopically reversible (or *isentropic*) processes, $S_f - S_i = 0$. Applying this condition to the equation above,

$$\Delta E = -k_B T \ \Delta \ln(Z). \tag{2.14}$$

We define the *free energy* as [30]

$$\mathcal{F} = -k_B T \ \ln(Z). \tag{2.15}$$

We will estimate the work done when a system moves from an initial thermodynamic state at temperature T to a final state at the same temperature. We leave the intermediate process unspecified. According to (2.9) and (2.10), the entropy must rise at least by Q/T, where Q stands for the heat transferred from the reservoir to the system. Thus, if S_i and S_f correspond to the entropies of the initial and final distributions, then solving $S_f - S_i$ for the free energy difference $\Delta\mathcal{F}$, we get

$$\Delta\mathcal{F} = \Delta E - T \ (S_f - S_i) \leq \Delta E - Q = \langle W \rangle . \tag{2.16}$$

Therefore, the difference $\Delta\mathcal{F}$ between the free energies of the initial and final states represents the maximum amount of work that can be extracted in practice from a spontaneous process with the same initial and final temperatures ($\Delta T = 0$) [36], or the minimum amount of work needed to reverse

it.

We will now apply the tools developed in chapter 1 and the definitions in this section to derive the Jarzynski equality (2.25) [49], a statistical identity that entails the work inequality (2.16) and relates the work done in a process to the free energy difference[2].

If $z_E = (z,\ z_R)$ denotes the microstate of a system (z) and of a heat reservoir at temperature T (z_R), our microscopic reversibility assumption implies that the direct and reverse transition probabilities between z_E and z'_E coincide (assuming, of course, that system and reservoir are isolated from everything else). When we coarse-grain our description and leave all the degrees of freedom in the reservoir out of the picture, then z, the coordinates and momenta of the system (disregarding the reservoir), corresponds to many different values of z_E. We write $\Omega(x_z)$ for the density of states in the reservoir associated with microstate z.

$$\Omega(x_z) = \int \delta(z' - z)\ dz'\ dz_R. \tag{2.17}$$

According to (1.30), the forward and reverse transition probabilities between z and z' satisfy[3]

$$\frac{p(z',\ t+\tau \mid z,\ t)}{p(\tilde{z},\ \tilde{t}+\tau \mid \tilde{z}',\ \tilde{t})} = e^{\frac{1}{k_B}(S_B(x_{z'}) - S_B(x_z))}. \tag{2.18}$$

We suppose that the initial states of direct and reverse processes linking z and z' have canonical probabilities, so that

$$\frac{\rho(z,\ t+\tau)\ p(z',\ t+\tau \mid z,\ t)}{\rho(\tilde{z}',\ \tilde{t}+\tau)\ p(\tilde{z},\ \tilde{t}+\tau \mid \tilde{z}',\ \tilde{t})} = e^{\frac{1}{k_B T}(H(z') - H(z) - \Delta\mathcal{F}) + \frac{1}{k_B}(S_B(x_{z'}) - S_B(x_z))}. \tag{2.19}$$

Note that the entropy difference signifies an increase in the Boltzmann entropy of the *reservoir*, as $S_B(x_z) = k_B \ln(\Omega(x_z))$. Supposing we have no

[2]The main idea behind the following proof was taken from [48].

[3]We use $\tilde{t}$ to denote time in the reverse process.

further information about the process, our definition of heat (2.10) states that the system must absorb

$$Q = -T\left(S_B(x_{z'}) - S_B(x_z)\right) \tag{2.20}$$

from the reservoir in the transition from x_z to $x_{z'}$. We combine the equation above with the one preceding it and obtain

$$\frac{\rho(z,\ t+\tau)\ p(z',\ t+\tau \mid z,\ t)}{\rho(\tilde{z}',\ \tilde{t}+\tau)\ p(\tilde{z},\ \tilde{t}+\tau \mid \tilde{z}',\ \tilde{t})} = e^{\frac{1}{k_B T}(H(z')-H(z)-\Delta\mathcal{F}-Q)}. \tag{2.21}$$

Jarzynski's discovery concerns the following work average over all the possible realisations of an isothermal process

$$\left\langle e^{-\frac{1}{k_B T}W} \right\rangle = \int\int e^{-\frac{1}{k_B T}W} \rho(z,\ t)\ p(z',\ t+\tau \mid z,\ t)\ dz\ dz'. \tag{2.22}$$

With (2.21) we can transform the integral from an average over direct trajectories into an average over reverse trajectories,

$$\begin{aligned}&\left\langle e^{\frac{1}{k_B T}W} \right\rangle \\ &= \int\int e^{-\frac{1}{k_B T}(H(z')-H(z)-\Delta\mathcal{F}-W-Q)} \rho(\tilde{z}',\ \tilde{t})\ p(\tilde{z}',\ \tilde{t}+\tau \mid \tilde{z},\ \tilde{t})\ dz\ dz'.\end{aligned} \tag{2.23}$$

Applying the first law of thermodynamics[4], $H(z') - H(z) = W + Q$,

$$\left\langle e^{\frac{1}{k_B T}W} \right\rangle = \int\int e^{-\frac{1}{k_B T}(-\Delta\mathcal{F})} \rho(\tilde{z}',\ \tilde{t})\ p(\tilde{z}',\ \tilde{t}+\tau \mid \tilde{z},\ \tilde{t})\ dz\ dz', \tag{2.24}$$

but $\Delta\mathcal{F}$ is just a constant determined by the initial and final canonical distributions, so we can factor it out of the integrals. Inside, we leave only a normalised joint probability density that integrates to unity. The resulting *Jarzynski equality* reads

$$\left\langle e^{-\frac{1}{k_B T}W} \right\rangle = e^{-\frac{1}{k_B T}\Delta F}. \tag{2.25}$$

[4]Note that the expression inside the integral refers to a single realisation between the microstates z and z', so we write $H(z)$ instead of E and W instead of $\langle W \rangle$.

Take the logarithm on both sides of (2.25), apply Jensen's inequality [50],

$$\left\langle -\frac{W}{k_B T} \right\rangle \leq \ln\left(\left\langle e^{-\frac{1}{k_B T}W}\right\rangle\right) = -\frac{\Delta\mathcal{F}}{k_B T}, \tag{2.26}$$

and multiply by $-k_B T$ to recover the work inequality (2.16).

Now we step back to think over our strategy. We posited a process with canonical initial and final states and paid attention to a specific transition. For the sake of concreteness, suppose that we keep the system isolated while we do work on it and then bring it back into contact with the reservoir and allow it to relax back to temperature T. As the work was carried out, part of the energy produced an increase $\Delta\mathcal{F}$ in the free energy, while the rest warmed the system up. When we resumed contact with the reservoir, the dissipated work flowed out as heat. Taken as a single constant-energy extended system, the entropy of the reservoir plus the system of interest equals the Boltzmann entropy. Therefore, *for this specific realisation*,

$$\begin{aligned} -W &= -W + T\,(S_B(x_{z'}) - S_B(x_z)) - T\,(S_B(x_{z'}) - S_B(x_z)) \\ &= -W + \Delta E - \Delta\mathcal{F} - Q. \end{aligned} \tag{2.27}$$

The first law simplifies this expression to $-W = -\Delta\mathcal{F}$ and we accomplish the simplification seen in the averages carried out to prove (2.25).

Using generalised canonical distributions (1.83), and keeping the initial values of all the Lagrange multipliers equal to their final values, we light upon an interesting result. Let $x_{E,f}$ represent a macrostate defined by the set of z that satisfy $H(z) = E$ and $F_1(z) = f_1,\ \ldots,\ F_k(z) = f_k$. Our equation (1.30) implies that

$$\begin{aligned} \frac{P_{H_0,F}(E,\ f)\ P_{H_0,F|H_t,F}(E',\ f' \mid E,\ f)}{P_{H_t,F}(E',\ f')\ P_{H_t,F|H_0,F}(E,\ f \mid E',\ f')} &= \frac{P_{H_0,F}(E,\ f)\ \Omega(x_{E',f'})}{P_{H_t,F}(E',\ f')\ \Omega(x_{E,f})} \\ &= e^{\frac{\Delta E - \Delta\mathcal{F}}{k_B T} + \sum_{i=1}^{k}\lambda_i \Delta f_i}. \end{aligned} \tag{2.28}$$

Remember that all the microstates in $x_{E,f}$ share the same values of the relevant phase functions.

$$H(z_1) = H(z_2) = E;\ F_i(z_1) = F_i(z_2) = f_i; \tag{2.29}$$

for all $z_1,\ z_2 \in x_{E,f}$, and $i = 1,\ 2,\ \ldots,\ k$. Hence, all processes from $x_{E,f}$ to $x_{E',f'}$ satisfy $H(z') - H(z) = E' - E$, $F_i(z') - F_i(z) = f_i' - f_i$. Let us define the function $\Delta\Phi$ as

$$\Delta\Phi(z,\ z') = H(z') - H(z) + k_B T \sum_{i=1}^{k} \lambda_i (F_i(z') - F_i(z)). \tag{2.30}$$

When we now average the exponential of $-\Delta\Phi/(k_B T)$ for all possible realisations connecting the initial distribution $P_{H_0,f}$ with the final $P_{H_t,f}$, we can again flip the probabilities like in (2.21),

$$\begin{aligned}
\left\langle e^{-\frac{1}{k_B T}\Delta\Phi} \right\rangle &= \int e^{-\frac{1}{k_B T}\Delta\Phi} P_{H_0,F}(E,\ f) \\
&\qquad \times P_{H_0,F|H_t,F}(E',\ f' \mid E,\ f)\ dE\ dE'\ df\ df' \\
&= \int e^{-\frac{1}{k_B T}\Delta\mathcal{F}} P_{H_t,F}(E',\ f') \\
&\qquad \times P_{H_t,F|H_0,F}(E,\ f \mid E',\ f')\ dE\ dE'\ df\ df' \\
&= e^{-\frac{1}{k_B T}\Delta\mathcal{F}},
\end{aligned} \tag{2.31}$$

and arrive at an analogue of Jarzynski's equality (2.25).

To determine the average dissipated work, we compare the evolution above to the adiabatic process connecting the initial and final states. In the latter process, the system starts off with a generalised canonical distribution $\rho(z;\ 0)$ (1.83) that then evolves in time obeying Hamilton's equations (1.42). The Hamiltonian $H(z,\ t)$ might depend on time, but the dynamics conserve the probability density along trajectories, $\rho(z;\ t) = \rho(\widetilde{\tilde{T}(\tilde{z};\ t;\ t)};\ 0)$, so

$$\rho(z;\ t) = \frac{1}{Z} e^{-\frac{1}{k_B T} H(\widetilde{\tilde{T}(\tilde{z};\ t;\ t)},\ 0) - \sum_{i=1}^{k} \lambda_i F_i(\widetilde{\tilde{T}(\tilde{z};\ t;\ t)})}. \tag{2.32}$$

Alternatively, we allow heat flows between the system and its environment. The Lagrange multipliers change their values from λ_i to λ_i' in response to

a variation in the expected values of the phase functions F_i. Assume we keep the λ_i' and the form of $H(z,\ t)$ constant at the end of the process for however long it takes the distribution to forget its previous history. The final distribution of the isothermal process corresponds to a new generalised canonical ensemble,

$$\rho'(z;\ t) = \frac{1}{Z'} e^{-\frac{1}{k_B T} H(z,\ t) - \sum_{i=1}^{k} \lambda_i' F_i(z)}. \tag{2.33}$$

When detailed balance holds (1.34), we can replace the ratio of forward and reverse transition probabilities between macrostates with probability densities. Although z represents microstates, when we consider the bath as part of our system, z corresponds to a set of microstates in the extended system. When we focus on transitions between $\widetilde{\tilde{T}(\tilde{z};\ t;\ t)}$ and z,

$$\frac{p(\widetilde{\tilde{T}(\tilde{z};\ t;\ t)},\ t \mid z,\ 0)}{p(\tilde{z},\ \tilde{t} \mid \tilde{T}(\tilde{z};\ t;\ t),\ 0)} = \frac{\rho(z;\ t)}{\rho'(z;\ t)}. \tag{2.34}$$

Inspired by our previous results, we examine the logarithm of this quotient.

$$\begin{aligned} k_B \ln\left(\frac{\rho(z;\ t)}{\rho'(z;\ t)}\right) &= \frac{1}{T}\left(H(z,\ t) - H\left(\widetilde{\tilde{T}(\tilde{z};\ t;\ t)},\ 0\right)\right) \\ &\quad + k_B \sum_{i=1}^{k}\left(\lambda_i' F_i(z) - \lambda_i F_i\left(\widetilde{\tilde{T}(\tilde{z};\ t;\ t)}\right)\right) \\ &\quad - \frac{1}{T}\Delta\mathcal{F}. \end{aligned} \tag{2.35}$$

We define a function Φ, with dimensions of energy, that changes along the trajectory both with z and with λ_i.

$$\Phi(z(t)) = H(z(t)) + k_B T \sum_{i=1}^{k} \lambda_i(t) F_i(z(t)). \tag{2.36}$$

Reversible change only accounts for the difference in the free energies $\Delta\mathcal{F}$. The rest of the energy we lose through dissipation. Defining the dissipated work as

$$W_{diss}(z(t)) = \Phi(z(t)) - \Phi(z(0)) - \Delta\mathcal{F}, \tag{2.37}$$

we realise that equations (2.35) and (2.37) entail

$$W_{diss}(z(t)) = k_B T \ \ln\left(\frac{\rho(z;\ t)}{\rho'(z;\ t)}\right). \tag{2.38}$$

The expected dissipation for the process follows from averaging with $\rho(z;\ t)$ because the initial state was described by $\rho(z;\ 0)$.

$$\langle W_{diss}\rangle = k_B T \ \mathrm{D}(\rho\|\rho'). \tag{2.39}$$

The derivation above generalises the discovery by Kawai, Parrondo and Van der Broeck [51] in the context of processes joining states in equilibrium at the same temperature[5]. The Kullback-Leibler divergence [52], defined by

$$\mathrm{D}(\rho\|\rho') = \int \rho(z;\ t) \ln\left(\frac{\rho(z;\ t)}{\rho'(z;\ t)}\right)\ dz, \tag{2.40}$$

quantifies how much ρ differs from ρ', and we will have much to say about it in the following pages. D does not satisfy the properties of a distance. In particular, $\mathrm{D}(\rho\|\rho')$ does not in general coincide with $\mathrm{D}(\rho'\|\rho)$. Nonetheless, $\mathrm{D}(\rho\|\rho') \geq 0$, with equality holding only when $\rho(z;\ t) = \rho'(z;\ t)$ almost everywhere.

2.4 Maximising relative entropy

The variational principle of maximum relative entropy coincides with Jaynes's maximum entropy formalism when the equilibrium distribution determines the probabilities over the values of the dynamical invariants. We can relax this condition slightly in the case of canonical and micro-canonical equilibria.

Intuitively, it would seem that probability distributions in phase space will spread out as far as possible, until they reach a state as close to equilibrium

[5]Kawai *et alii* provided an expression for any intermediate time t' along the process, but this follows from (2.39) through a change of variables $z \mapsto T(z;\ t;\ t - t')$.

as the constraints allow. When we measure "closeness" to equilibrium with the Kullback-Leibler divergence, we obtain a new variational principle[6].

We define the *relative entropy* as

$$\Delta S[\rho] = -k_B \mathrm{D}(\rho \| \rho^{eq}). \tag{2.41}$$

The aforementioned properties of D imply that $\Delta S \leq 0$, with equality holding when ρ coincides with the equilibrium ensemble. We can determine the distribution that differs least from equilibrium while still satisfying the constraints by imitating the method in Chapter 1, Section 1.6 and maximising the relative entropy (2.41) with the help of the method of Lagrange multipliers to include the relevant constraints.

How does this new variational principle relate to the maximum entropy formalism already discussed? When we turn our attention to the formal expression for the equilibrium ensemble (2.6), a simple connection between S and ΔS becomes apparent.

$$\begin{aligned}
\Delta S[\rho] &= -k_B \int \rho(z)\ \ln\left(\frac{\rho(z)}{m\ \frac{P_I(I(z))}{\Omega(I(z))}}\right)\ dz \\
&= -k_B \int \rho(z)\ \ln\left(\frac{\rho(z)}{m}\right)\ dz + k_B \int \rho(z)\ \ln\left(\frac{P_I(I(z))}{\Omega(I(z))}\right)\ dz \\
&= S[\rho] + k_B \int \mathrm{Tr}[\rho\ \delta(I-i)] \ln\left(\frac{P_I(i)}{\Omega(i)}\right)\ di.
\end{aligned} \tag{2.42}$$

We assume the last integral extends over all the values of i. Equation (2.42) shows that maximising the relative entropy will yield the same results as the maximum entropy method *whenever the final integral on the right disappears in the variation of* ΔS.

The probability distribution for the random variable I representing the dynamical invariants must remain fixed for isolated systems, as autonomous Hamiltonian mechanics cannot change the distribution of probabilities. In that case,

$$\mathrm{Tr}[\rho\ \delta(I-i)] = \mathrm{Tr}[\rho^{eq}\delta(I-i)] = P_I(i), \tag{2.43}$$

[6]This section presents results reported in [34].

and the integral over i in (2.42) becomes independent of ρ and vanishes when we calculate the variational derivative of ΔS with respect to ρ.

When we sent the first draft of our paper on relative entropy to the *Journal of Statistical Physics* [34], an anonymous reviewer pointed out that condition (2.43) could be relaxed when ρ^{eq} designates a generalised canonical distribution (1.83), for in that case we need assume only that the *expected* values of the dynamical invariants for the unknown distribution ρ coincide with those of the equilibrium ensemble. In other words, if we concede that

$$\mathrm{Tr}[\rho\ I] = \mathrm{Tr}[\rho^{eq} I], \tag{2.44}$$

then a generalised canonical ρ^{eq} allows us to rewrite the integral in (2.42) to make it independent of ρ,

$$k_B \int \rho(z)\ \ln\left(\frac{\rho^{eq}(z)}{m}\right)\ dz = k_B \int \rho^{eq}(z)\ \ln\left(\frac{\rho^{eq}(z)}{m}\right)\ dz = -S[\rho^{eq}]. \tag{2.45}$$

Thus, equation (2.42) becomes $\Delta S[\rho] = S[\rho] - S[\rho^{eq}]$. Furthermore, when ρ^{eq} represents a microcanonical ensemble (1.102), we just have to require that ρ vanishes whenever ρ^{eq} does to get the same expression.

The entropy formulae above contain a constant measure $m(z) = m$ because in classical Hamiltonian systems the measure turns out to be proportional to a power of Planck's constant h. In a three-dimensional N-particle system, for example, $m(z) = h^{-3N}$ as long as we can tell particles apart. If we wish to take indistinguishability into account, with N_1 particles of the first type, N_2 of the second type, and so on ($N_1 + N_2 + \ldots + N_m = N$), we must not overcount states by including the permutations among particles of the same type, so

$$m(z) = h^{-3N} \prod_{i=1}^{m} \frac{1}{N_i!}. \tag{2.46}$$

Instructors tend to mention this *Boltzmann counting* (2.46) in connection with the Gibbs paradox in order to demonstrate that classical statistical mechanics leads to incorrect predictions. They say that only quantum mechanics, where interchanging particles does not count as a real event,

can dissolve the paradox, although Jaynes showed in 1992 that Gibbs had a strikingly simple and more elegant solution already in the 19th century [53].

Given the equilibrium ensemble and the additional constraints on the unknown distribution ρ ($\mathrm{Tr}[\rho] = 1$ and $\mathrm{Tr}[\rho\, F_i] = f_i$), solving the variational equation $\delta\Delta S = 0$ gets us

$$\rho(z) = \frac{\rho^{eq}(z)\; e^{-\sum_{i=1}^{k} \lambda_i F_i(z)}}{\mathrm{Tr}\left[\rho^{eq}(z)\; e^{-\sum_{i=1}^{k} \lambda_i F_i(z)}\right]}, \tag{2.47}$$

which according to (2.42) should coincide with the maximum entropy result (1.83) as long as the problem satisfies one of the conditions that make both methods equivalent.

Our relative entropy road overlaps Jaynes's approach in many circumstances, but it still provides a distinct way to calculate statistical ensembles. Sometimes it simplifies derivations. For instance, a system in contact with a heat reservoir at temperature T approaches a state of equilibrium described by the canonical ensemble (1.114). When the canonical ρ^{eq} appears in (2.47), equation (1.115) from Section 1.6 follows immediately, without any assumptions about the interactions between the system and the reservoir or any work to link the Lagrange multiplier β to the environment temperature T. There is no magic involved here. We simply get an easier derivation because the equilibrium ensemble ρ^{eq} already contains much of the relevant information.

Leaving aside whatever aesthetic worth our contribution might possess, its true value lies in our capacity to calculate ensemble averages in settings where the maximum entropy route performs poorly. We have devoted the following sections to examples of the benefits of relative entropy, after the next few pages, which I have dedicated to explaining a subtle point that could potentially lead to some confusion.

We do not doubt that the relevant distribution (2.47) embodies the correct solution to our problem for canonical equilibria, but there seems to be a mismatch between the maximum entropy and the relative entropy

formulae in other cases. The discrepancy is only apparent, though. When I began to think about relative entropy, I believed that we always get the same solution when we maximise (2.41), namely, the equilibrium ensemble times an appropriate normalisation factor multiplied by the exponential of $-\sum_{i=1}^{k} \lambda_i F_i$, but this rule might deceive us.

Take, for example, a slightly more general problem than considered up to this point. So far, we have assumed that our systems had a definite number of particles, but we rarely know the exact number of, say, atoms in a macroscopic experiment. Even when we enclose the system in an adiabatic container, the number of atoms may change from one run of the experiment to the next, without us considering that we have turned to a different system.

Let us investigate how to determine the appropriate joint distribution $\rho(z_N,\ N)$ representing the probability density for a system of N particles in the state z_N of the accessible N-particle phase space Γ_N. Integrating this function over z_N outputs the marginal probability of finding N particles in our system, $P(N)$,

$$P(N) = \int_{\Gamma_N} \rho(z_N,\ N)\ dz_N. \tag{2.48}$$

Moreover, the probability density $P(E,\ N)$ of encountering an N-particle system with energy E equals

$$P(E,\ N) = \int_{\Gamma_N} \rho(z_N,\ N)\ \delta(H_N(z_N) - E)\ dz_N. \tag{2.49}$$

H_N stands for the N-particle Hamiltonian.

If we wish to calculate the entropy of joint distributions over the z_N and N, then we must generalise our entropy functional to

$$S[\rho] = -k_B \sum_N \int_{\Gamma_N} \rho(z_N,\ N)\ \ln(h^{3N}\rho(z_N,\ N))\ dz_N, \tag{2.50}$$

and, similarly, the relative entropy becomes

$$\Delta S[\rho] = -k_B \sum_N \int_{\Gamma_N} \rho(z_N,\ N)\ \ln\left(\frac{\rho(z_N,\ N)}{\rho^{eq}(z_N,\ N)}\right)\ dz_N. \tag{2.51}$$

Jaynes's maximum entropy principle applied to (2.50) plus (2.49) and normalisation returns a peculiar relevant ensemble,

$$\rho(z_N,\ N) = \frac{P(H_N(z_N),\ N)\ e^{-\sum_{i=1}^{k}\lambda_i F_{i,N}(z_N)}}{\int_{\Gamma_N} e^{-\sum_{i=1}^{k}\lambda_i F_{i,N}(z_N)}\delta(H_N(z'_N)-H_N(z_N))\ dz'_N}, \tag{2.52}$$

which looks formally different from the obvious generalisation of (2.47), that is,

$$\rho(z_N,\ N) = \rho^{eq}(z_N,\ N)\ \frac{e^{-\sum_{i=1}^{k}\lambda_i F_{i,N}(z_N)}}{\mathcal{Z}}, \tag{2.53}$$

where $\mathcal{Z}$ represents the normalising factor.

The disagreement between (2.52) and (2.53) disappears when we carry out the operations carefully. Before we solve the variational problem, we reflect about whether we should include (2.49) as constraints. We would not need (2.49) if the energy and the number of particles were allowed to change as the system approached equilibrium, for then the equilibrium probabilities $P^{eq}(E,\ N)$ would not reveal any information about the initial values of E and N. Instead, we imagine an isolated system with fixed E and N, so that the equilibrium ensemble contains all the information about the probabilities of these invariants. On that account, we must formulate the variational equation specifying that the unknown joint distribution must yield the same probability densities over E and N as the equilibrium ensemble, that is,

$$\begin{aligned} P(E,\ N) &= \int_{\Gamma_N} \rho(z_N,\ N)\ \delta(H_N(z_N)-E)\ dz_N \\ &= \int_{\Gamma_N} \rho^{eq}(z_N,\ N)\ \delta(H_N(z_N)-E)\ dz_N. \end{aligned} \tag{2.54}$$

This equation represents a constraint for every pair of E and N. Including these constraints together with the conditions on the average values f_i in the variation of the relative entropy (2.51) leaves us with the problem of

maximising ΔC,

$$\Delta C = \Delta S - k_B \sum_{i=1}^{k} \lambda_i \left(\sum_N \int_{\Gamma_N} \rho(z_N,\ N)\ F_{i,N}(z_N)\ dz_N - f_i \right) - k_B \sum_N \int_0^\infty \mu(E,\ N) \Bigg(\int_{\Gamma_N} \rho(z_N,\ N)\ \delta(H_N(z_N) - E)\ dz_N - P(E,N) \Bigg)\ dE. \quad (2.55)$$

Note that we have included a Lagrange multiplier μ for each pair of E and N, as required by the constraints on $P(E,\ N)$. Equating the variational derivatives of ΔC to zero and solving for the unknown distribution brings us to

$$\rho(z_N,\ N) = \rho^{eq}(z_N,\ N)\ e^{-\mu(H_N(z_N),\ N) - 1 - \sum_{i=1}^{k} \lambda_i F_{i,N}(z_N)}, \quad (2.56)$$

which we can identify with (2.53) as long as we interpret $\mathcal{Z}$ as a function of E and N. In other words, by defining

$$\mathcal{Z}(E,\ N) = e^{\mu(E,\ N)+1}, \quad (2.57)$$

we identify (2.56) with (2.53). We insert (2.56) into (2.54) and solve for $\mathcal{Z}$.

$$\mathcal{Z}(E,\ N) = \frac{\int_{\Gamma_N} \rho^{eq}(z_N,\ N)\ e^{-\sum_{i=1}^{k} \lambda_i F_{i,N}(z_N)} \delta(H_N(z_N) - E)\ dz_N}{P(E,\ N)}. \quad (2.58)$$

The equilibrium ensemble (2.6) in this case reads

$$\rho^{eq}(z_N,\ N) = \frac{P(H_N(z_N),\ N)}{h^{3N} \Omega_N(H_N(z_N))}. \quad (2.59)$$

Equations (2.57)-(2.59) transform (2.56) into (2.52), so maximising the relative entropy does indeed lead to the same result as Jaynes's approach, as we expected.

2.5 Nonergodic behaviour

Relevant ensembles obtained with the principle of maximum relative entropy are better suited to the problem of sampling trajectories with unknown or cumbersome dynamical invariants.

Hamiltonian equations sometimes contain hidden dynamical invariants, which bind the trajectories to a subset within the phase space manifold defined by the values of the known dynamical invariants. In other problems, we may know about additional invariants that are just too cumbersome to handle analytically. In either case, relevant distributions calculated with Jaynes's maximum entropy principle might not reproduce the average values found in experiments or simulations.

For example, in Figure 2.3 I have drawn an isolated gas separated into two compartments by a diathermal wall. The experiment begins with both sides at the same temperature, but with a greater density of particles on the left. We allow the wall to move in response to the difference in pressures until the system reaches mechanical and thermal equilibrium. If we wished to describe the movement of the wall, we could use the average velocity $v(x)$ of the heavy particle calculated with the relevant distribution as a first crude approximation to the time rate of change of x,

$$\frac{dx}{dt} \approx v(x) = \mathrm{Tr}\left[\bar{\rho}\,\frac{dX}{dt}\right]. \tag{2.60}$$

Below the diagram of the gas experiment lies a simpler one-dimensional analogue. Here we replace the moving wall with a heavier particle. All the masses interact with each other and with the container walls through short-range truncated Lennard-Jones potentials (1.77). We know, of course, that the equations of motion conserve the number of particles to the left and right of the "wall". The fact that the particle coordinates progress from left to right

$$q_1 < q_2 < \ldots < q_{100} \tag{2.61}$$

complicates the analysis. How do we express this fact as the expected value of a function of the microstates?

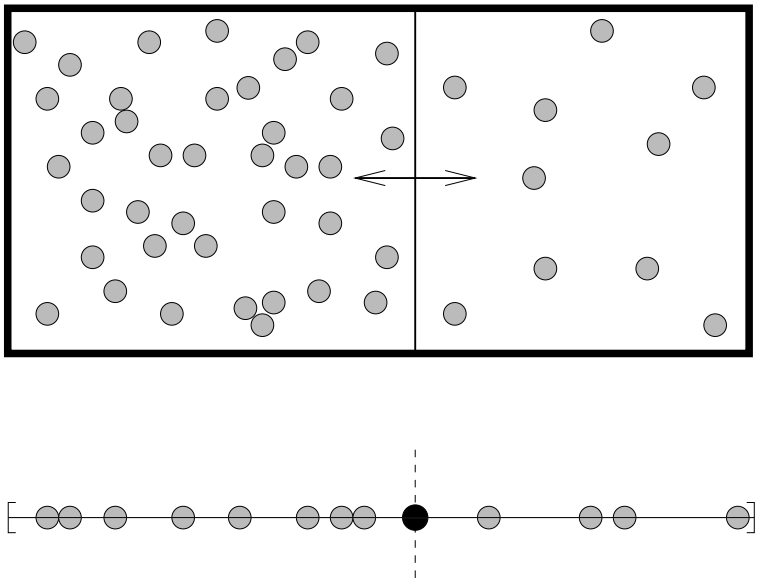

Figure 2.3: Two gases separated by a diathermal movable wall. The gases start at the same temperature, but with a greater density of particles in the compartment on the left. Below, a one-dimensional analogue of the problem replaces the movable wall with a heavier particle (in black).

We start our simulations with the heavy particle (mass $5m$) at rest in the middle ($X = 0$) of a simulation box of length 1000 σ, dividing the gas into two subsystems of 74 and 25 particles of mass m on the left and right, respectively. The position X of the heavy particle will serve as a relevant variable, and the function F will represent the force exerted on it in a given microstate. We will work out $f(x)$, the expected value of the force for the average position x.

If we wished to determine $f(x)$ numerically, Jaynes's method suggests that we fix X to different values x in the range of interest. For each value of x we run an equilibrium simulation and keep track of the force exerted on the fixed particle to calculate the expected value. Is it not easier to sample the equilibrium distribution and use that information to calculate nonequilibrium averages?

Regrettably, the subsystems might change their temperature as the heavy particle moves around so, before we carry out the equilibrium simulations, we should first run nonequilibrium simulations to find out the average temperatures of the subsystems as a function of x.

Here we propose a different approach. According to the discussion in Section 2.4, averages with the relevant distribution establish the connection between x and λ.

$$x(\lambda) = \mathrm{Tr}\left[\frac{\rho^{ref} e^{-\lambda X} X}{\mathrm{Tr}\left[\rho^{ref} e^{-\lambda X}\right]}\right]. \tag{2.62}$$

The reference distribution ρ^{ref} represents the actual probability density sampled in the long run, which may differ from the equilibrium distribution ρ^{eq} determined by the principle of maximum entropy and the information available. Whatever the initial condition, a single sufficiently long simulation of mixing dynamics should sample the accessible region of phase space to which ρ^{ref} assigns uniform probabilities. Interpret $z_1,\ z_2,\ \ldots,\ z_M$ as the microstates traversed by the numerical trajectory. We set $X = 0$ and let the bit-reversible molecular dynamics (1.75) figure out how X changes in time. Then we compute x for different values of λ with

$$x(\lambda) = \frac{\sum_{j=1}^{M} e^{-\lambda X(z_j)} X(z_j)}{\sum_{j=1}^{M} e^{-\lambda X(z_j)}}. \tag{2.63}$$

Using the data collected, we can invert the relation and plot λ as a function of x. A simple rule then takes us to the desired expected forces.

$$f(x) = \frac{\sum_{j=1}^{M} e^{-\lambda(x)X(z_j)} F(z_j)}{\sum_{j=1}^{M} e^{-\lambda(x)X(z_j)}}. \tag{2.64}$$

So the time average does in fact correspond to the expected value in the region of phase space accessible to the system. Figure 2.4 shows the numerical results calculated with three independent runs of the simulation. The (noisy) stable equilibrium position lies near $x = 240\ \sigma$, where the force vanishes. Beyond this point the particle is pushed back on average. The figure presents consistent results over most of the range sampled, but the edges of the plot show discrepancies.

I suspect that the curves bend away from each other at low and high values of x due to insufficient samples in the corresponding ranges, although the low values might be partially explained as an artefact of beginning with the heavy particle at rest (see page 88 for a discussion of a similar initialisation effect). When $\lambda = 0$, we obviously get equilibrium averages in (2.63) and (2.64). As we move away from equilibrium, we increase the distance from λ to zero and sets of rarely visited microstates receive greater weights. Reducing the error for these average values requires very long simulations that visit the rare states many times. A greater λ-range therefore demands much more intensive computation (but see the explanation of Figure (2.10) for a way to sample some of these extreme values).

The average values inferred with this method depend sensitively on the preparation of the initial state of the simulation. The initial macrostate may give rise to evolutions along separate paths for the relevant averages and the results will reflect these differences. For example, if we had assigned an initial thermal velocity to the heavy particle in our simulations, the runs would have led to disparate results (see Figure 2.5). In those cases, we must resort to multiple runs, although hopefully not as many as suggested by Jaynes's formalism. Interestingly, we can determine expected values by averaging over all the runs, but if we wish to *simulate* macroscopic phenomena, we should randomly pick the numerical results for just one of

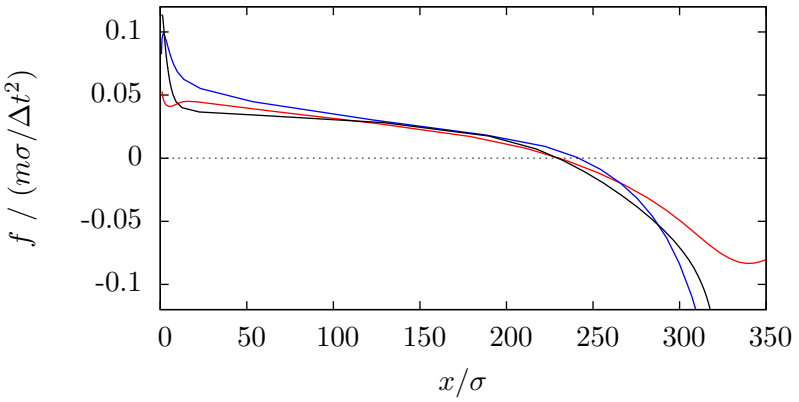

Figure 2.4: Expected value of the exerted force on the heavy particle in Figure 2.3 as a function of its expected position calculated with three different simulation runs. The diverging ends of the curves are probably due to insufficient samples for these extreme values.

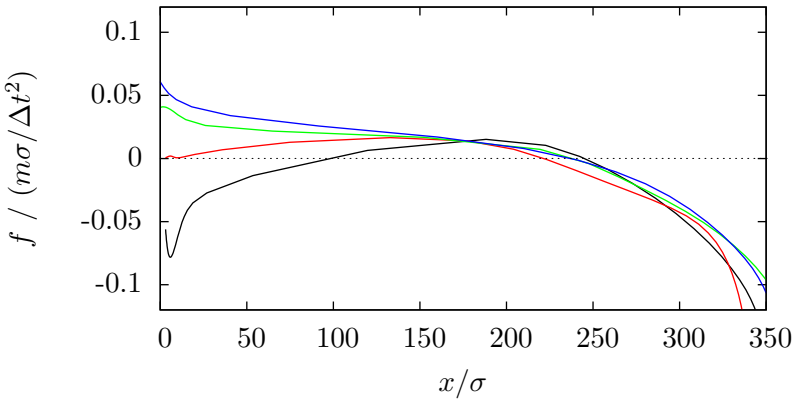

Figure 2.5: Expected value of the force exerted on the heavy particle in Figure 2.3 as a function of its expected position calculated for several initial velocities of the heavy particle. The dissonance on the left illustrates the dependence on the initial conditions.

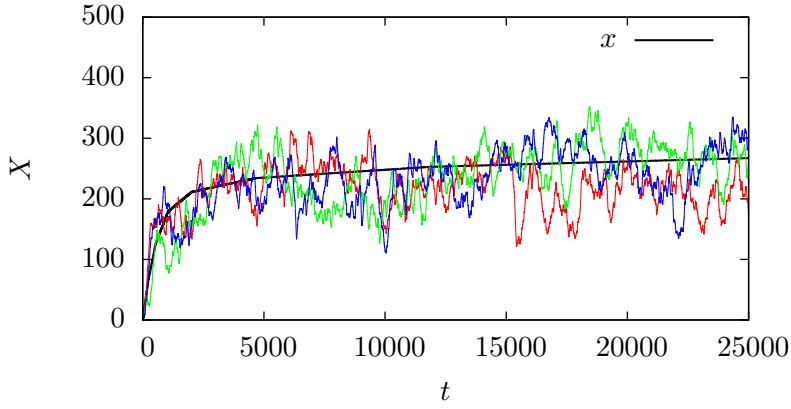

Figure 2.6: Position X of the heavy particle as a function of time during three simulation runs. The thick black line represents the expected value x as a function of time, estimated from numerical integration of the average velocity (2.60).

the runs.

When we carry out the calculation of the average velocity (2.60) with only the first simulation run, we see that the numerical integration of the results for $x(t)$ already tracks the realisations reasonably well, as shown in Figure 2.6.

The approach sketched above promises to simplify calculations when we cannot easily incorporate the structure of a system as relevant information in terms of average values. Protein simulation is a case in point. By recording virtual proteins relaxing to equilibrium, we should be able to predict nonequilibrium averages, at least in some range close to equilibrium.

The property of ergodicity becomes irrelevant when trajectories explore only part of the accessible volume in phase space over the time scales of interest. Despite the thermodynamical claim that carbon at standard temperature and pressure spends most of its time as graphite and that dia-

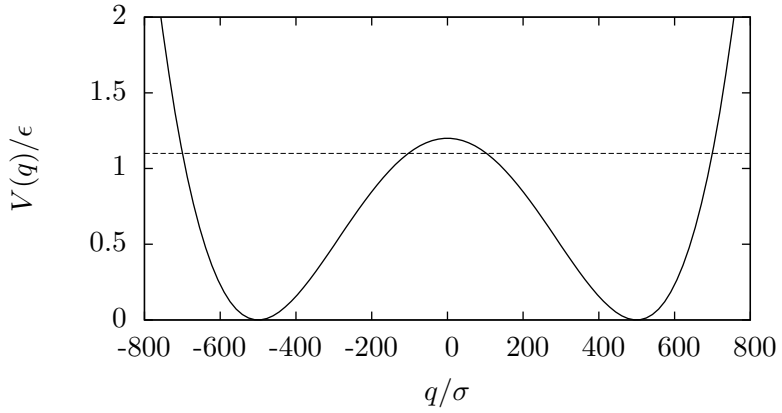

Figure 2.7: Double-well potential (2.65) confining $N = 100$ Lennard-Jones particles. The dotted line marks the average energy per particle E/N.

monds are unstable [54], we would not use the former allotrope's average values to describe the latter. Diamonds do not last forever, perhaps, but they do usually outlive us.

Similarly, many simulated systems classified as ergodic when we contemplate their ideal eternal trajectories behave very differently over the time scales of interest to us. Take the double-well potential in Figure 2.7,

$$V(x) = 1.2\epsilon \left(\left(\frac{x}{500\sigma} \right)^4 - 2 \left(\frac{x}{500\sigma} \right)^2 + 1 \right). \tag{2.65}$$

If we use it to confine one hundred Lennard-Jones particles and agree always to begin our simulations with all the particles on the left, then we will probably never find them all on the right, no matter how much computing power we invest on this problem, notwithstanding the fact that mirror images of the initial state exist (changing all the coordinate signs) with energies identical to the starting values.

I wrote a molecular dynamics program including the attractive part of

the Lennard-Jones potential,

$$V_{LJ}(r_{ij}) = 4\epsilon \left(\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right), \tag{2.66}$$

where again $r_{ij} = \|q_j - q_i\|$, the distance between particles i and j. The constant energy E for the dynamics was chosen to ensure that the average energy per particle E/N lay below the height of the central potential barrier. When we submitted our article [34], I mistakenly believed that this condition would guarantee the absence of ergodicity. Because we need more energy to move a particle to the right than we possess per particle, I argued, the system does not have enough energy to get all the particles over the barrier.

In a colloquium after the publication of [34], J. M. R. Parrondo expressed his concern about this point. He suspected that the conclusion could not be correct. In the ensuing discussion, C. Mejia Monasterio identified a mechanism that could transport all the particles to the right of the double-well whenever there was enough total energy to get two particles over the central barrier. The particles cannot all move to the right in one go, but if two particles cross over, one of them could go back with the energy of *both*, making it possible for *two* particles on the left to surmount the barrier. Repeating the process, we would eventually end up with the system in the right well. The bounded total energy, therefore, does not necessarily break the ergodicity.

Employing simulations with three and four particles, I managed to observe these whole-system transitions to the right, but we expect them to become less and less probable as the number of particles increases. When the molecular dynamics include one hundred particles (see Figure 2.8), we find that we have not yet reached equilibrium after times of $1.6 \times 10^5 \sqrt{m\sigma/\epsilon}$. We only need simulations of about half this length for the particles to divide themselves more or less evenly on both sides of the barrier. Averaging the kinetic temperature over the whole run, we obtain $T_{kin}/k_B = 1.2 \pm 0.1$ (mean $\pm$ standard deviation). We observed that the particles on the left and right display the same expected temperature, but the standard devia-

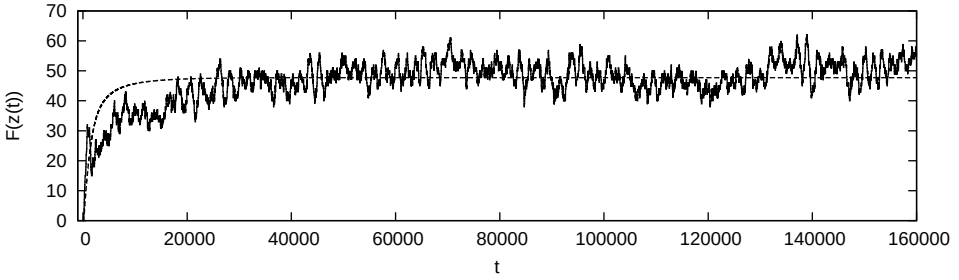

Figure 2.8: Number of particles with a positive coordinate value versus time for a system of $N = 100$ Lennard-Jones particles initially to the left of the potential barrier in figure 2.65. A classic Runge-Kutta fourth-order algorithm with time step equal to $10^{-4}\sqrt{m\sigma/\epsilon}$ was used to integrate the equations of motion. The total energy remained constant to four significant figures. Numerical integration with the flux density from figure 2.9 allowed us to estimate the average value of F as a function of time (dashed line).

tion doubles on the right!

In this problem, we would like our relevant variable $F(z)$ to represent the number of particles on the right

$$F(z) = \sum_{i=1}^{N} \theta(q_i). \tag{2.67}$$

We have used θ here to represent the Heaviside step function. Let $f = \mathrm{Tr}[\rho\ F]$ stand for the average number of particles on the right. The *flux density* $v(f)$,

$$\frac{df}{dt} \approx v(f) = \mathrm{Tr}\left[\rho \sum_{i=1}^{N} \delta(q_i)\ \frac{p_i}{m_i}\right], \tag{2.68}$$

gives us the expected speeds of the particles crossing the barrier.

Inserting the maximum entropy relevant distribution into (2.68) makes $v(f) = 0$, because we obtain odd integrands in the momenta p_i. Every accessible microstate $z = (q,\ p)$ value cancels a corresponding $\tilde{z} = (q,\ -p)$

value, making the trace vanish (the relevant distribution and Dirac delta function do not change when we invert p_i).

The relative entropy route suggests that we calculate $v(f)$ directly from the sampled microstates $\{z_j\}$.

$$v(f) = \frac{\sum_{j=1}^{M} e^{-\lambda(f)F(z_j)} \sum_{i=1}^{N} \theta(\frac{\sigma}{2} - |q_i|) \, \frac{p_i}{m_i}}{\sum_{j=1}^{M} e^{-\lambda(f)F(z_j)}}. \tag{2.69}$$

We replaced the Dirac delta function with an approximation, $\delta(q_i) \approx \theta(\sigma/2 - |q_i|)$, in order to make the computations feasible. Figure 2.9 presents the values of v as a function of f. By numerical integration, we estimated the evolution of f with time (represented with a dashed line in Figure 2.8).

We also find a suspicious bend in Figure 2.9 for f close to zero. To improve our calculations at this end we can repeatedly run simulations beginning in the left well and calculate averages over the different runs. This computation would have taken us a couple of months with our current facilities, but we can save time by noting that λ increases indefinitely as f approaches zero. For large λ (for example, $\lambda \approx 10$) microstates with $F(z_j) > 10$ hardly contribute to the average flux (2.69), so we can run the simulation until we get to eleven particles on the right and then start again.

Averaging over several runs confirmed the behaviour for small f represented in Figure 2.9, as shown in Figure 2.10. The data defies our intuition, though. When no particles have surmounted the barrier, the flux density should be greater than when one or two particles have crossed over. I do not believe that this effect has any real physical significance. Rather, it probably results as a consequence of how we initialise our system in the simulations. Our starting conditions spread the masses out, so that they do not overlap, and many end up at rest on the left well's slopes. Until they have time to mix, the system tends to compress on the left as they slide down towards the bottom of the well.

The point of view described in the last few pages contrasts with the *rare event sampling* approach. In the latter, we wish to obtain representative microstates from the whole set of accesible states. The problem then comes from the tiny transition probabilities to some regions in phase space. For

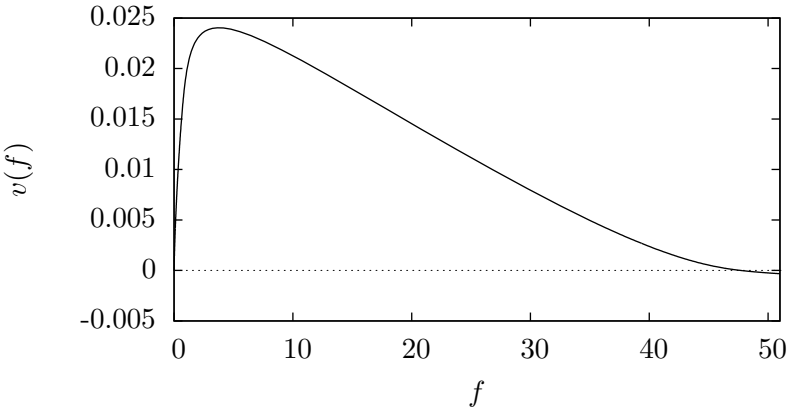

Figure 2.9: Flux density calculated with equation (2.69) versus average number of particles to the right of the potential barrier, f.

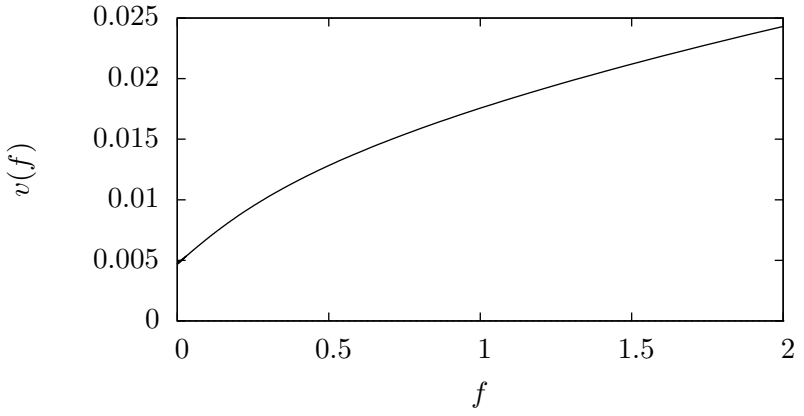

Figure 2.10: Flux density calculated from 100 simulation runs confirms the turn observed for small values of f in Figure 2.9.

example, we might wish to simulate the folding and unfolding of a protein, which occurs very rarely on the time scales of the numerical experiment. In that case, we would be interested precisely in the frequency with which it switches between configurations. But suppose we wanted to describe an irreversible transition instead, such as a spontaneous chemical reaction with a small activation barrier. Then we would begin with the reactants and concentrate on how they transform into products, forgetting about the reverse process back to the reactants. The method described in this section does not contribute anything new to the sampling of rare events. Rather, it is supposed to help us calculate the properties of systems which are trapped in a certain region of phase space over the time scales of interest, or undergoing irreversible transitions.

2.6 Multiple relevant variables

We investigate the possibility of using a single very long simulation to calculate averages for *different* sets of values of the simulation parameters.

I would like to mention some preliminary results concerning Bellman's curse of dimensionality [44]. Suppose we faced a problem with three relevant variables, instead of one. If we want to sample phase space for one hundred different values of each variable, then now we need a million simulations to cover all the combinations.

Could we use a single very long simulation to calculate the desired data? To answer the question, we simulate the movement of three disks (masses $m = 6.25$) interacting through truncated Lennard-Jones potentials with 197 unit-mass disks inside a square box of length 80 σ with periodic boundary conditions. The heavier disks were 2.5 times larger than the other disks. Fourth-order Runge-Kutta integration moves the system forward in time.

Guided by the reasoning in the previous section, we calculate the expected force on one of the heavy particles as a function of the position of the other two. We let d_1 denote the distance to the nearest disk, d_2 the

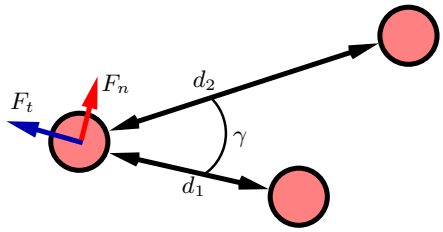

Figure 2.11: A numerical simulation of three disks interacting with smaller disks through truncated Lennard-Jones potentials tracked both components of the effective force exerted on one of the disks as a function of the positions of the other two.

distance to the other disk, and γ the angle drawn by the three particles (see Figure 2.11).

First we concentrate on the effective force as a function of the distance to the nearest disk, with the other one far away ($40 < d_2 < 50$). Figure 2.12 shows some noisy but encouraging data. We see an average attraction when the two disks come close to the repulsive part of the potential, suggesting the presence of a depletion interaction. No such effect appears when we look at the normal component of the force, F_n.

Next, we analyse what happens when we bring the other particle into the picture. We pick $7 < d_1 < d_2 < 8$, a range in which we find a net attraction between the nearby particles when the third particle is not present. We would like to monitor the effect of changing the angle, starting at $\gamma = \pi/2$. Figure 2.13 suggests that the position of the third particle has an effect on the effective force. These results are tentative, as they came in while I was working on this thesis, so they should be taken with a pinch of salt.

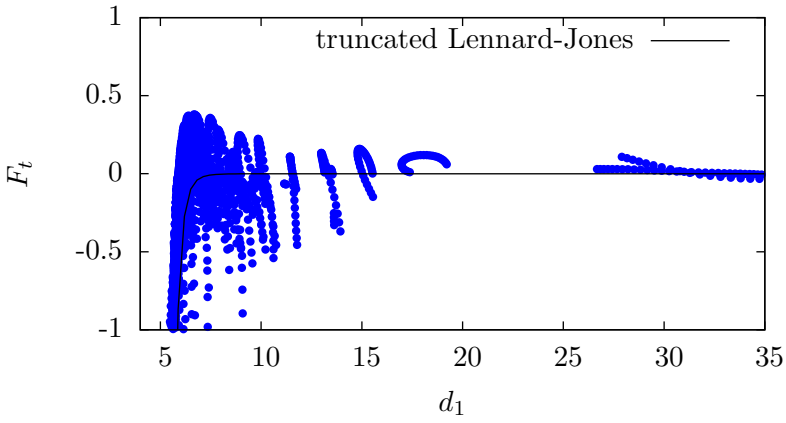

Figure 2.12: Tangential component of the force, F_t versus distance, d_1 (see Figure 2.11) when the third particle lies far away ($40 < d_2 < 50$). The hard-core Lennard-Jones repulsion (*black line*) develops a coarse-grained attractive depletion interaction.

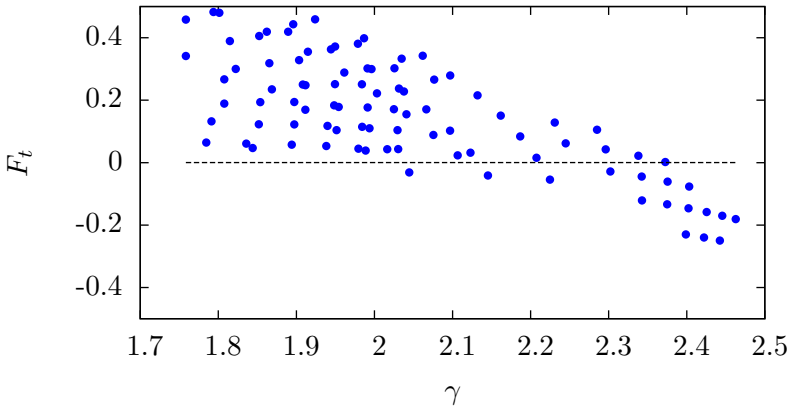

Figure 2.13: The effect of an additional particle on the attractive force shown in Figure 2.12 ($7 < d_1 < d_2 < 8$).

2.7 Nonequilibrium steady states

Using a reference probability distribution for given values of the parameters in a nonequilibrium steady state, we can use the principle of maximum relative entropy to calculate an approximation to the distribution for different values of the parameters.

Shortly after the *Journal of Statistical Physics* accepted our paper on relative entropy [34], I received an email from Puneet Patra, who works at the Indian Institute of Technology Kharagpur. He suggested extending the relative entropy approach to nonequilibrium steady state distributions and immediately began to work diligently on the project[7].

I could see no reason why maximising relative entropy should work in this context. After all, the steady state distributions do not satisfy the conditions that make maximising ΔS equivalent to Jaynes's principle. However, Patra's preliminary findings brought a paper on relative entropy that I had not read to my attention [57]. In it, Shore and Johnson proposed a principle similar to Bayes's rule for probabilities. Suppose we already possess some information that determines the *prior* probability density function ρ' for the state of a system (through maximum entropy, sampling, guessing, divine revelation or whatever). We then come across new information in the form of expected values of a set of phase functions, $\mathrm{Tr}[\rho\ F] = f$. The distribution that most resembles the prior but satisfies the new constraints on the average values follows from maximising the relative entropy with respect to ρ'.

Patra decided to explore one of the simplest systems known to display chaotic nonequilibrium steady state behaviour: a single unit mass trapped in a quartic potential $V(q) = q^4/4$ within a temperature field defined by

$$T(q) = 1 + \epsilon\ \tanh(q). \tag{2.70}$$

There are different ways to incorporate the action of heat reservoirs into

[7] This section briefly explains results that will appear in a forthcoming paper by Patra, Bhattacharya and myself [55]. See the arXiv preprint [56]

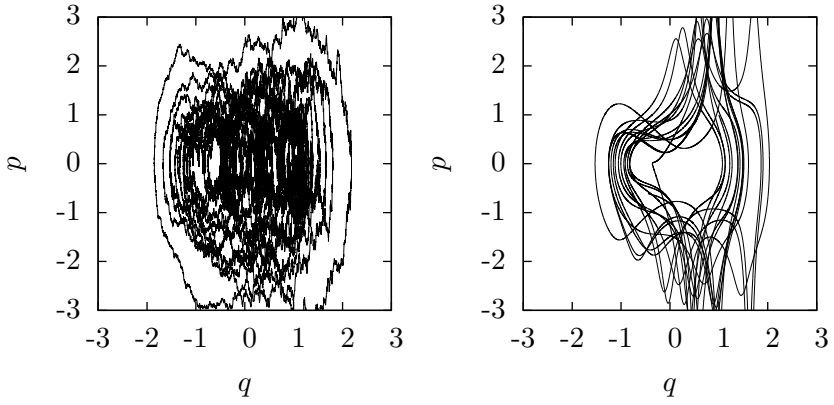

Figure 2.14: Phase space trajectories for a unit mass in a quartic potential under the influence of a position-dependent temperature field (2.70) with $\epsilon = 0.5$. The figure illustrates the large difference between the noisy Langevin dynamics (2.71) on the left and the smooth Hoover-Holian thermostat (2.73) on the right.

our dynamics[8]. One of the most widespread approaches looks to generalised Langevin equations [58] and modifies the canonical equations (1.42) to

$$\left.\begin{aligned} \dot{q} &= p, \\ \dot{p} &= -x^3 - \zeta p - \psi. \end{aligned}\right\} \tag{2.71}$$

The expected value of the stochastic force ψ equals zero. The fluctuation-dissipation theorem connects ψ to the damping constant ζ through

$$\left\langle \psi(t)\ \psi(t') \right\rangle = 2\zeta\ T(x)\ \delta(t - t'). \tag{2.72}$$

A popular alternative for sampling phase space makes use of deterministic time-reversible thermostats [19]. A particularly simple scheme, the

[8]Readers unfamiliar with constant temperature dynamics may wish to leave this section until after they have read sections 3.5 and 4.2.

Hoover-Holian thermostat [60], implements simultaneous integral control of the kinetic temperature and its fluctuation and generates an extended phase space distribution consistent with the canonical ensemble (1.114). Applied to the quartic oscillator, it leads to the following equations of motion.

$$\left.\begin{aligned} \dot{q} &= p, \\ \dot{p} &= -x^3 - \eta p - \xi p^3, \\ \dot{\eta} &= p^2 - T(q), \\ \dot{\xi} &= p^4 - 3\ T(q)\ p^2. \end{aligned}\right\} \tag{2.73}$$

Langevin and Hoover-Holian dynamics sample the same equilibrium distributions, but they follow contrasting trajectories (see Figure 2.14).

From the constraints on expected values of several phase functions, such as the heat flow, position and energy, and the maximum entropy principle, we calculated the least-biased distribution for a given value of ϵ. Then we began with a reference distribution ρ' for a *different* value of ϵ and worked out which distribution satisfied the constraints while maximising the relative entropy with respect to ρ'. Both methods generated reasonable approximations for the distribution, ρ, calculated with Langevin dynamics. In Patra's simulations, the relative entropy route won, while I found Jaynes's distribution closer to the right answer (See Figure 2.15). The disagreement was probably caused because we used different algorithms to follow the evolution of the system, but we are still working on the details.

For Hoover-Holian dynamics, the relative entropy solution was clearly superior, and it remained so even when the prior distribution ρ' differed considerably from ρ. In the figure, we show approximations for ρ when $\epsilon = 0.7$ and choosing $\epsilon = 0.4$ for the prior distribution. Through the reference distribution, the maximum relative entropy solution manages to incorporate information on the multifractal shape of the distribution in phase space.

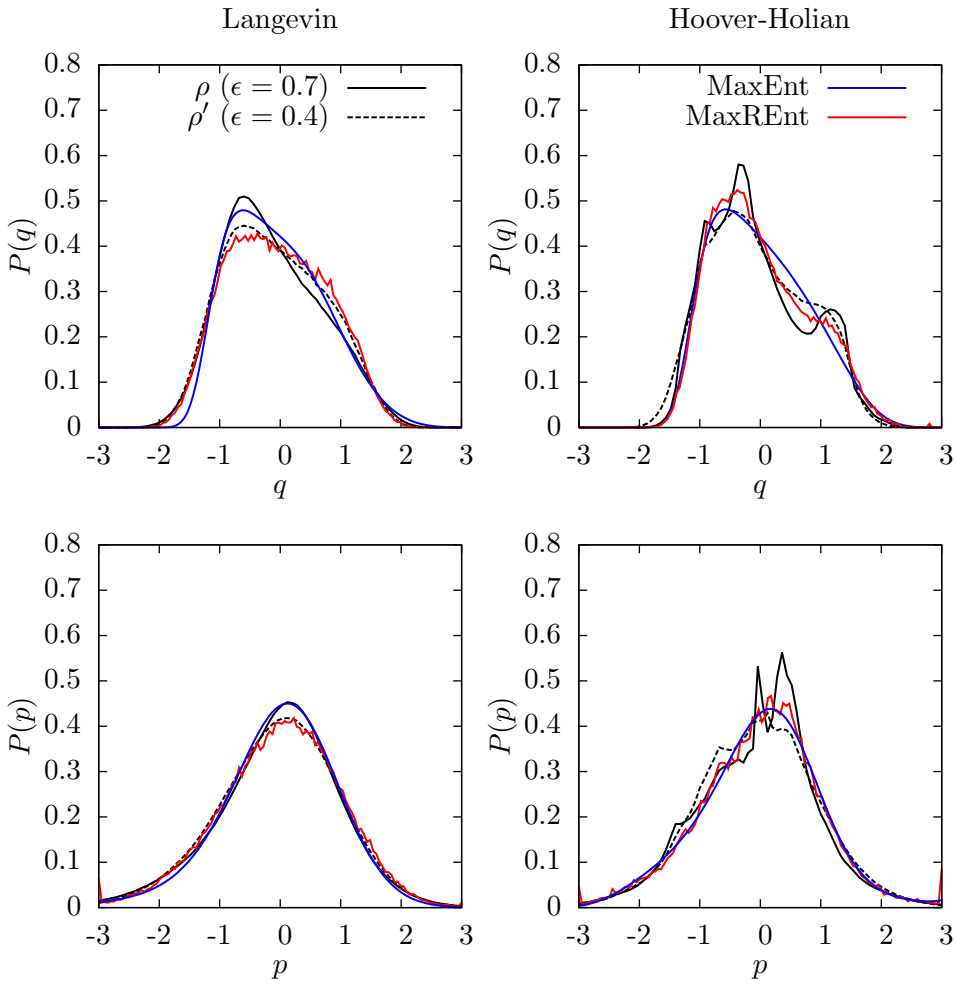

Figure 2.15: Nonequilibrium steady state marginal distributions for the quartic oscillator with $\epsilon = 0.7$. The blue line represents the maximum entropy approximation (MaxEnt) and the red line shows the maximum relative entropy solution (MaxREnt) calculated from the prior distribution ρ' with $\epsilon = 0.4$.

2.8 Dissipation and lag

We can set an upper bound on how much an evolved distribution lags behind the relevant distribution by determining the dissipated work. We prove the second law of thermodynamics for processes connecting equilibrium states.

From (2.42), we gather that the maximum entropy relevant distribution (1.83) need not always coincide with the result of maximising relative entropy. In particular, changing fields and thermodynamic forces will give rise to unequal distributions. A process in which we vary, say, the pressure and intensity of an external electric field acting on the system will change the probabilities for the dynamical "invariants", I. This means that, if ρ_i represents the maximum entropy distribution consistent with the information on the initial state and ρ_f the corresponding distribution for the final state, then in general $\mathrm{Tr}[\rho_i\ I] \neq \mathrm{Tr}[\rho_f\ I]$, so maximum entropy is no longer equivalent to maximising ΔS. What does relative entropy measure in these cases, then?

Vaikuntanathan and Jarzynski realised that the dissipated work theorem (2.39) by Kawaii *et alii* provided some insight into the constant-temperature evolution of probability distributions [59]. Instead of interpreting the distribution ρ' as a final state in the process, they proposed to think of it differently. Suppose λ stands for the values of the parameters determining the state of equilibrium (pressure, chemical potential and the intensity of an external field, for example)[9]. For every set of values of the parameters, we can define a maximum entropy equilibrium distribution $\rho^{eq}(\lambda)$, which $\rho(z;\ t)$ would approach asymptotically in time if λ were held fixed. If we devise an experimental protocol that makes λ depend on time, then the distribution $\rho(z;\ t)$ will lag behind $\rho^{eq}(\lambda(t))$ unless we carry out the variation of λ slowly enough.

[9]Some authors may wish to distinguish between conjugate variables (Lagrange multipliers) and external fields (or control parameters). Here we refer to both simply as "parameters".

Because the system must dissipate (2.39) in order to reach equilibrium, the Kullback-Leibler divergence establishes an "upper bound" on how much $\rho(z;\ t)$ can depart from $\rho^{eq}(\lambda(t))$. We will return to this point in Chapter 4 (Section 4.6).

Relative entropy cannot turn positive [52]. Mathematical consistency enforces

$$\Delta S(\rho(t)\|\rho^{eq}(\lambda(t))) \leq 0. \tag{2.74}$$

Similarly, as we mentioned in Section 1.4, mathematics teaches us that entropy remains invariant under Hamiltonian laws.

$$S[\rho(0)] = S[\rho(t)]. \tag{2.75}$$

Jaynes realised that we can use this fact to provide a very simple proof of the Second law of thermodynamics [10]. First, we note that the statistical mechanical interpretation of entropy S (1.23) gives us a solid foundation for the heat theorem (suitably generalised to (2.9)) and, therefore, that its value coincides with the thermodynamic entropy for equilibrium states, S_E, apart from a constant term, perhaps, which can be adjusted for some reference state. The principle of maximum entropy implies that, given an arbitrary distribution ρ,

$$S[\rho] \leq S[\rho^{eq}] = S_E. \tag{2.76}$$

Suppose we begin with an equilibrium distribution $\rho(z;\ 0) = \rho^{eq}(z)$ and that we perform some experimental manipulation that makes the system evolve adiabatically (under Hamiltonian mechanics) towards a different state of equilibrium with entropy S'_E. Then (2.75) and (2.76) bring us to the statement of the Second law,

$$S_E = S[\rho(0)] = S[\rho(t)] \leq S'_E, \tag{2.77}$$

making $S'_E - S_E \geq 0$, with equality holding when $\rho(t)$ coincides with the new equilibrium distribution.

We can easily rephrase Jaynes's proof in the language of relative entropy. Back in Section 2.4 we saw that canonical equilibria imply that

$\Delta S[\rho] = S[\rho] - S[\rho^{eq}]$. If we assume that $\rho(0) = \rho^{eq}(\lambda(0))$, then the adiabatic evolution to a new state of canonical equilibrium $\rho^{eq}(\lambda(t))$ must satisfy (2.74),

$$\begin{aligned}\Delta S(\rho(t)\|\rho^{eq}(\lambda(t))) &= S[\rho(t)] - S[\rho^{eq}(\lambda(t))] \\ &= S[\rho^{eq}(\lambda(0))] - S[\rho^{eq}(\lambda(t))] \leq 0, \end{aligned} \tag{2.78}$$

proving the Second law.

2.9 Summary

The simple fluctuation relation (1.30) we found in Chapter 1 guided our thinking about heat and dissipation. The theorem applies to any reversible Markov process, such as the wandering king example we encountered in Section 2.2. In the realm of statistical physics, it allowed us to infer nonequilibrium relations for work and dissipation like the work theorem and Jarzynski's equality. It also suggested that we focus on relative entropy.

Maximising relative entropy, we saw, constitutes a distinct variational principle from which to derive ensembles. The results coincide with Jaynes's maximum entropy formalism when the process considered does not alter the probabilities associated with the values of the dynamical invariants (although we can relax this condition slightly in the case of canonical and microcanonical equilibria). Although the relative entropy approach requires a reference distribution that we must determine in advance, it has some advantages over Jaynes's method: it simplifies the sampling of systems that are not ergodic, or effectively not ergodic over the time scales of interest; it simplifies the sampling of systems with many relevant variables; and we can use it to extend the results of a nonequilibrium steady state simulation to other states with similar values of the control parameters. Furthermore, the concept of relative entropy allowed a very simple proof of the Second law of thermodynamics and, through its relation to dissipation and lag, it gave us some clues about how systems drift away from the relevant distribution as time goes by.

Relative entropy methods and fluctuation theorems apply far beyond the realm of quasistatic processes. So far, we have set out some simple calculations of thermodynamic quantities without considering the intervening nonequilibrium process. In a couple of examples, we managed to calculate a rough approximation to the relaxation to equilibrium by sampling a nonequilibrium process. However, in realistic applications we will rarely have enough computational power to simulate the whole relaxation to equilibrium. The macroscopic variables will take way too long (in integration steps) to reach their final values. Therefore, if we wish to describe how macroscopic quantities evolve in time, we must turn to the theory of coarse-graining, which is the subject of the next chapter.

Chapter 3

Macroscopic Evolution

We explain how to derive dynamical equations for the expected values of phase functions that depend directly on the real dynamics, instead of the projected dynamics, as in Zwanzig's formalism.

"What is time?", puzzled Augustine of Hippo in his *Confessions*, "If no one asks me, I know. If I try to explain it to someone who asks, then I do not know"[1]. Time somehow intertwines this immediate familiarity and deep mystery. On reflection, in fact, we might even conclude that time is impossible, self-contradictory or non-existent, as Parmenides perhaps did[62]. The present, for instance, seems no more than the meeting point between what happened in the past and the not-yet-existent future. Even the great Newton, who wisely decided not to define time in his *Principia*, "as being well known to all", paradoxically wrote about the "flow" of time [63], but does flowing itself not require time?

Nevertheless, with his sharp analytical genius, Newton realised that we tend to confuse two different senses of time[2]. The first he called absolute, true and mathematical time. We could describe it in later Kantian terms as

[1] Confessions, Book XI, 14 [61].

[2] See *Scholium*, I, after the definitions at the beginning of the *Principia* [63].

the conditions that make change possible. The second, relative, apparent and common time referred to Aristotle's old explanation of time as "the measure of change with respect to 'before' and 'after'"[3]. The former meaning we will leave for the philosophers to discuss. Concerning the latter, we can shed some light on the subject.

We begin by avoiding the prima facie circularity of Aristotle's definition with an atemporal notion of 'before' and 'after'. We interpret 'before' as lower entropy, and 'after' as higher entropy. In *this* sense, time is left undefined for microscopic dynamics until we specify some way of grouping microstates into macrostates, for this determines the Boltzmann entropy associated with a microstate [5]. Furthermore, when an isolated macroscopic system reaches equilibrium, it no longer changes and time stops.

Measuring time involves two processes: a uniform cyclic motion (like the turning hands on a wrist watch), which divides different stages of change into comparable intervals, and a system that changes its Boltzmann entropy (like the person wearing the watch), which defines the arrow of time. Because the counting of cycles proceeds in the direction of growing entropy, time in an isolated system could conceivably reverse its direction if the entropy diminished, although this becomes increasingly unlikely the greater the number of particles in the system. In our analyses, we will not mention the cyclic clock, but we will assume its presence.

This chapter explains how to derive the laws of irreversible change from underlying reversible Hamiltonian dynamics[4] or, as Newton might have said, how the passing of common time affects the properties of matter.

3.1 The theory of coarse-graining

We recast Liouville's equation for autonomous Hamiltonian evolution as a closed integro-differential equation for the time evolution of the

[3] *Physics*, Book IV, 11 [64].

[4] See Section 3.8 at the end of this chapter for comments on how to infer macroscopic laws from non-Hamiltonian dynamics.

expected values of a set of relevant phase functions. Unlike in Zwanzig's theory, our method does not rely on projection operators.

Given the autonomous Hamiltonian function H and the initial probability distribution $\rho(z;\ 0)$, the statistical theory of macroscopic dynamics begins with Liouville's theorem (2.2) for the time evolution of ρ,

$$\frac{\partial \rho}{\partial t} = \{H,\ \rho\}.$$

We aim to rewrite this equation in terms of "macroscopic variables" [58]: the expected values f_i of a set of relevant phase functions $F_i(z)$. Combining the derivative of the definition of f_i (1.78) with Liouville's equation,

$$\frac{\partial f_i}{\partial t} = \frac{\partial}{\partial t}\ \mathrm{Tr}[\rho\ F_i] = \mathrm{Tr}[\{H,\ \rho\}\ F_i], \tag{3.1}$$

we find the equation that governs macroscopic change.

Even though the f_i stand for average values, this does not impose a serious restriction on our analysis. If we wish to study the whole probability distribution P_F, then we choose phase functions of the form $\Psi_f = \prod_{i=1}^{k} \delta(F_i - f_i) = \delta(F - f)$. The expected value of Ψ_f equals the probability of $F(z) = f$,

$$\mathrm{Tr}[\rho\ \Psi_f] = \mathrm{Tr}[\rho\ \delta(F - f)] = P_F(f). \tag{3.2}$$

The problem with (3.1) is that it does not express the variation of f in terms of f. Rather, it forces us to solve Liouville's equation first. If we could express ρ in terms of the relevant distribution $\bar{\rho}$ (the overbar distinguishes between the relevant distribution and the solution of Liouville's equation), then (3.1) would become a closed equation for f. In Chapter 1, we saw that the f_i determined $\bar{\rho}$ through the principle of maximum entropy. Therefore, we write ρ as the sum of a relevant part $\bar{\rho}$ plus an irrelevant part $\delta\rho$ (irrelevant, that is, to the calculation of the expected values).

$$\rho = \bar{\rho} + \delta\rho. \tag{3.3}$$

The mathematical properties of the Poisson bracket allow us to split the trace in (3.1) into two parts, because $\{H,\ \rho\} = \{H,\ \bar{\rho}\} + \{H,\ \delta\rho\}$.

$$\frac{\partial f_i}{\partial t} = \mathrm{Tr}[\{H,\ \bar{\rho}\}\ F_i] + \mathrm{Tr}[\{H,\ \delta\rho\}\ F_i]. \tag{3.4}$$

When the Hamiltonian equals a sum of kinetic and potential energies, $H = T_k(p) + V(q)$, then integration by parts shows that we can exchange the positions of $\bar{\rho}$ and F_i and change the sign of the integral,

$$\mathrm{Tr}[\{H,\ \bar{\rho}\}\ F_i] = -\mathrm{Tr}[\bar{\rho}\ \{H,\ F_i\}], = -\mathrm{Tr}\left[\bar{\rho}\ \frac{\partial F_i}{\partial t}\right], \tag{3.5}$$

and the same reasoning applies to the integral with $\delta\rho$.

$$\frac{\partial f_i}{\partial t} = -\mathrm{Tr}\left[\bar{\rho}\ \frac{\partial F_i}{\partial t}\right] - \mathrm{Tr}[\delta\rho\ \{H,\ F_i\}]. \tag{3.6}$$

The first term in the sum we call the *organised drift*, $v_i(f)$. Defining the Liouvillian operator as[5]

$$\hat{L}\mathrm{i}f = -\{H,\ f\}, \tag{3.7}$$

$v_i(f)$ becomes

$$v_i(f) = \mathrm{Tr}\left[\bar{\rho}\ \hat{L}\mathrm{i}F_i\right]. \tag{3.8}$$

Note the dependence of v on f through the relevant distribution, $\bar{\rho}$. The last term we have yet to express as a function of the expected values f. This can be achieved by solving the time evolution equation for $\delta\rho$ formally in terms of $\bar{\rho}$.

$$\frac{\partial}{\partial t}\delta\rho = \frac{\partial}{\partial t}(\rho - \bar{\rho}) = \{H,\ \rho\} - \frac{\partial\bar{\rho}}{\partial t} = \{H,\ \delta\rho\} + \{H,\ \bar{\rho}\} - \frac{\partial\bar{\rho}}{\partial t}. \tag{3.9}$$

Using the Liouvillian, we rewrite this equation.

$$\frac{\partial}{\partial t}\delta\rho = -\hat{L}\mathrm{i}\delta\rho - \hat{L}\mathrm{i}\bar{\rho} - \frac{\partial\bar{\rho}}{\partial t}. \tag{3.10}$$

[5]Textbooks usually write $\mathrm{i}\hat{L}$ or L instead of $\hat{L}\mathrm{i}$. The imaginary unit supposedly reminds the reader of the unitary character of the transformation generated by the Liouvillian operator.

We solve (3.10) as if $\bar{\rho}$ were a known function.

$$\delta\rho(z;\ t) = e^{-\hat{L}\mathrm{i}t}\delta\rho(z;\ 0) - \int_0^t e^{-\hat{L}\mathrm{i}(t-s)} \left(\hat{L}\mathrm{i}\bar{\rho}(z;\ s) - \frac{\partial\bar{\rho}(z;\ s)}{\partial s} \right)\ ds. \quad (3.11)$$

The easiest proof of (3.11) simply calculates the time derivative of $\delta\rho$ and verifies that it satisfies (3.10). We always assume that $\delta\rho(z;\ 0)$ vanishes everywhere, because our best guess at the initial ensemble equals a relevant distribution, $\rho(z;\ 0) = \bar{\rho}(z;\ 0)$. So substituting the remaining integral in (3.11) into (3.6), we get

$$\frac{\partial f_i}{\partial t} = v_i(f_i) - \mathrm{Tr}\left[\frac{\partial F_i}{\partial t} \int_0^t e^{-\hat{L}\mathrm{i}(t-s)} \left(\hat{L}\mathrm{i}\bar{\rho}(z;\ s) + \frac{\partial\bar{\rho}(z;\ s)}{\partial s} \right)\ ds \right]. \quad (3.12)$$

We have achieved a closed (and rather complicated) equation for f. While the expected values f_i determine the relevant distribution $\bar{\rho}$, we find that we have to deal with an *implicit* integro-differential equation, given the fact that $\bar{\rho}$ does not depend on the time explicitly, and its derivative should be interpreted as a shorthand for

$$\frac{\partial\bar{\rho}}{\partial t} = \sum_{i=1}^{k} \frac{\partial\bar{\rho}}{\partial f_i}\frac{\partial f_i}{\partial t}. \quad (3.13)$$

In the simplest case, the relevant distribution approximates the solution of Liouville's equation closely,

$$\frac{\partial\bar{\rho}}{\partial t} = -\hat{L}\mathrm{i}\bar{\rho}, \quad (3.14)$$

and the time integral in (3.12) cancels, leaving only

$$\frac{\partial f_i}{\partial t} = v_i(f). \quad (3.15)$$

This means that the relevant distribution coincides with the real distribution at all times, and $\delta\rho(z;\ t) = 0$ everywhere. The approximate evolution equations for the relevant variables in Chapter 2 had precisely this form.

Most frictionless problems in mechanics textbooks could be considered particular examples of this situation. We represent a pendulum bob or a falling stone as a point mass when they are actually made of a great number of atoms. The 'position' of the mass refers to the location of the centre of mass, $X(z)$, and 'velocity' to $dX(z)/dz \cdot \dot{z}$. If x and p stand for the expected values of X and P, and we can neglect the details of the interaction between the system and the suspending string or the surrounding air and reduce them to a single conservative force $F(x)$, then we have a Hamiltonian flow for the centre of mass position $X(z)$ and momentum $P(z)$. The drift averages work out as

$$v_x(x,\ p) = -\mathrm{Tr}[\bar{\rho}\ \{H,\ X\}] = \frac{p}{m}, \tag{3.16}$$

$$v_p(x,\ p) = -\mathrm{Tr}[\bar{\rho}\ \{H,\ P\}] = F(x). \tag{3.17}$$

Also, relevant distributions remain relevant as time goes by, because all the microstates in a given macrostate ($X(z) = x_0,\ P(z) = p_0$) evolve into the same macrostate at a later time ($X(z) = x_t,\ P(z) = p_t$). Therefore, $\partial\bar{\rho}/\partial t = -\hat{L}\mathrm{i}\bar{\rho}$ and we recover Hamilton's equations for the point mass from (3.12)

$$\left.\begin{aligned} \frac{\partial x}{\partial t} &= \frac{p}{m}, \\ \frac{\partial p}{\partial t} &= F(x). \end{aligned}\right\} \tag{3.18}$$

In this sense, we may view the classical equations of motion for macroscopic bodies as particular consequences of the theory of coarse-graining.

We have a fair amount of tedious work ahead of us as we rewrite the integral in (3.12) to make it friendly, so we had better break the process into smaller steps.

First, we hope our functions are well-behaved and that we can interchange the phase space and time integrals. Further, we separate the whole

integral into two parts,

$$
\begin{aligned}
&- \operatorname{Tr}\left[\frac{\partial F_i}{\partial t}\int_0^t e^{-\hat{L}\mathrm{i}(t-s)}\left(\hat{L}\mathrm{i}\bar{\rho}(z;\ s)+\frac{\partial\bar{\rho}(z;\ s)}{\partial s}\right)\ ds\right] \\
&\quad = -\int_0^t \operatorname{Tr}\left[(\hat{L}\mathrm{i}F_i)e^{-\hat{L}\mathrm{i}(t-s)}\hat{L}\mathrm{i}\bar{\rho}\right]\ ds - \int_0^t \operatorname{Tr}\left[(\hat{L}\mathrm{i}F_i)e^{-\hat{L}\mathrm{i}(t-s)}\frac{\partial\bar{\rho}}{\partial s}\right]\ ds.
\end{aligned}
\tag{3.19}
$$

In the first integral, we notice that the Liouvillian acts on the relevant distribution. We use the chain rule to transfer its action to the phase functions.

$$
\hat{L}\mathrm{i}\bar{\rho} = -\sum_{i=1}^{k}\frac{\partial\bar{\rho}}{\partial F_i}\{H,\ F_i\}. \tag{3.20}
$$

The variational principles for the entropy and relative entropy functionals (sections 1.6 and 2.4) taught us that relevant distributions corresponding to a set of expected values depend on the phase functions F through the exponential of $-\sum_{i=1}^{k}\lambda_i F_i$, which implies that

$$
\frac{\partial\bar{\rho}}{\partial F_i} = -\lambda_i\bar{\rho}. \tag{3.21}
$$

Combining the previous two equations, we conclude that

$$
\hat{L}\mathrm{i}\bar{\rho} = -\bar{\rho}\sum_{i=1}^{k}\lambda_i\ \hat{L}\mathrm{i}F_i. \tag{3.22}
$$

We will also make use of a handy property of the $e^{\hat{L}\mathrm{i}t}$ operator:

$$
\operatorname{Tr}\left[Ae^{-\hat{L}\mathrm{i}t}B\right] = \operatorname{Tr}\left[Be^{\hat{L}\mathrm{i}t}A\right], \tag{3.23}
$$

which follows from a change of variables $z \leftarrow \widetilde{\tilde{T}(\tilde{z};\ t;\ t)}$. Applying (3.22)

and (3.23) to the first of our integrals, we get

$$\int_0^t \mathrm{Tr}\left[(\hat{L}\mathrm{i}F_i)e^{-\hat{L}\mathrm{i}(t-s)}\hat{L}\mathrm{i}\bar{\rho}\right] = -\sum_{j=1}^{k}\int_0^t \lambda_j \mathrm{Tr}\left[\bar{\rho}\ (\hat{L}\mathrm{i}F_j)\ e^{\hat{L}\mathrm{i}(t-s)}(\hat{L}\mathrm{i}F_i)\right]\ ds. \quad (3.24)$$

Some additional notation will help keep our expressions compact. We define the *time correlation function* of A and B with respect to the distribution $\bar{\rho}$ as

$$\mathrm{C}_{\bar{\rho}}[A,\ B(t)] = \mathrm{Tr}\left[\bar{\rho}\ A\ e^{\hat{L}\mathrm{i}t}B\right]. \quad (3.25)$$

Introducing the definition in the integral above,

$$-\sum_{j=1}^{k}\int_0^t \lambda_j \mathrm{Tr}\left[\bar{\rho}\ (\hat{L}\mathrm{i}F_j)\ e^{\hat{L}\mathrm{i}(t-s)}(\hat{L}\mathrm{i}F_i)\right]\ ds. = -\sum_{j=1}^{k}\int_0^t \mathrm{C}_{\bar{\rho}}[\hat{L}\mathrm{i}F_j,\ \hat{L}\mathrm{i}F_i(s)]\ ds. \quad (3.26)$$

Now we concentrate on the last integral in (3.19), beginning with the time derivative of the relevant distribution, $\bar{\rho}$. We write $\bar{\rho}$ as

$$\bar{\rho}(z;\ t) = \frac{\rho^0(z)}{Z}e^{-\sum_{i=1}^{k}\lambda_i F_i(z)}. \quad (3.27)$$

Depending on whether the relevant distribution results from maximising the entropy or relative entropy functional, ρ^0 represents an appropriate constant function or the equilibrium distribution. In either case, it does not change in time, $\partial\rho^0/\partial t = 0$. As always, Z stands for the partition function,

$$Z = \mathrm{Tr}\left[\rho^0\ e^{-\sum_{i=1}^{k}\lambda_i F_i}\right]. \quad (3.28)$$

The Lagrange multipliers determine the value of Z and the relevant distribution, but they do not depend on time explicitly. Instead, they change as a result of the variation in the expected values, f_i. To calculate the time variation of $\bar{\rho}$ we must apply the chain rule all the way down to the time derivatives of f.

$$\frac{\partial \bar{\rho}}{\partial t} = -\bar{\rho} \sum_{j=1}^{k} \sum_{l=1}^{k} F_j \frac{\partial \lambda_j}{\partial f_l} \frac{\partial f_l}{\partial t} - \bar{\rho} \sum_{j=1}^{k} \sum_{l=1}^{k} \mathrm{Tr}[\bar{\rho} F_j] \frac{\partial \lambda_j}{\partial f_l} \frac{\partial f_l}{\partial t}. \tag{3.29}$$

Factoring out the common terms above,

$$\frac{\partial \bar{\rho}}{\partial t} = \sum_{j=1}^{k} \sum_{l=1}^{k} (F_j - f_j) \frac{\partial \lambda_j}{\partial f_l} \frac{\partial f_l}{\partial t}. \tag{3.30}$$

Let δF denote $(F_j - f_j)$. Inserting (3.30) into the last integral in (3.19) and applying (3.23) and (3.25), we rewrite the last term as

$$\int_0^t \mathrm{Tr} \left[(\hat{L}\mathrm{i} F_i)\ e^{-\hat{L}\mathrm{i}(t-s)} \frac{\partial \bar{\rho}}{\partial s} \right]\ ds = -\sum_{j=1}^{k} \sum_{l=1}^{k} \int_0^t \mathrm{C}_{\bar{\rho}}[\delta F_j,\ \hat{L}\mathrm{i} F_i(s-t)]. \tag{3.31}$$

Collecting our new expressions for the integrals (3.26) and (3.31), we transform equation (3.12) into

$$\begin{aligned} \frac{\partial f_i}{\partial t} \quad = \quad & v_i(f) - \sum_{j=1}^{k} \int_0^t \mathrm{C}_{\bar{\rho}}[\hat{L}\mathrm{i} F_j,\ \hat{L}\mathrm{i} F_i(s-t)]\ \lambda_j(f)\ ds \\ & - \sum_{j=1}^{k} \sum_{l=1}^{k} \int_0^t \mathrm{C}_{\bar{\rho}}[\delta F_j,\ \hat{L}\mathrm{i} F_i(s-t)] \frac{\partial \lambda_j}{\partial f_l} \frac{\partial f_l}{\partial s}\ ds. \end{aligned} \tag{3.32}$$

Equation (3.32) constitutes *one of the main results of this thesis dissertation.* Its value as a theoretical achievement lies in its exact nature, being an alternative expression of Liouville's theorem for an initially relevant distribution function. In contrast to Zwanzig's equations [65], (3.32)

does not include any projection operators (see Appendix A on projection operators). The noteworthy absence of Zwanzig's $1 - \mathrm{P}$ operator in the exponent makes our equation depend directly on the dynamical evolution of the relevant variables, instead of the projected dynamics. We will see later on how to derive a generalised Fokker-Planck equation that does not depend on the projected dynamics either (Section 3.3), and show how to extend (3.32) to non-Hamiltonian thermostated systems (Section 3.8).

Unfortunately, in addition to the integro-differential form of (3.32), the time derivatives of f appear also on the right hand side, hopelessly complicating even its numerical analysis, which might still be possible, but will no doubt become incredibly tedious.

3.2 Separation of time scales

If the relevant variables change very slowly over the typical time scales of molecular motion and the time correlation functions in (3.32) decay to zero quickly, then we can approximate the exact integro-differential equation with a memoryless partial differential equation.

MIT professor Sanjoy Mahajan constantly reminds his students (at MIT and elsewhere) that "when the going gets tough, the tough lower their standards" [66]. Following his advice, we will have to abandon exactness if we wish to make progress.

As we mentioned before, physical systems forget their past evolution when we consider long enough time scales. Remember that the state of our system depends on the variables we use to describe it [10]. As we saw in Chapter 1, the system forgets the past immediately if we specify its precise microstate, but memory effects appear as soon as we identify a set of microstates as the same macrostate.

We call a set of macroscopic variables *complete* [42] if their present values determine their future evolution. For example, the phenomenological equations of motion for a viscous fluid imply that knowledge of the momentum, mass and energy density fields at any instant determine the values of

these fields forever after. Magnetic hysteresis illustrates an *incomplete* set of variables, because we cannot determine how an external field will affect the magnetisation of a ferromagnet without knowledge of the past history of the variable.

Observable macroscopic magnitudes do not fluctuate wildly. If they did, we would not be able to measure them experimentally. Even if we could, their values at one time would tell us little about their immediate future. Stable macroscopic properties require slow evolutions when we view the phenomena on the molecular time scales. Therefore, we assume that over the time interval in which the integrands in (3.32) have not yet decayed to zero, the values of $\lambda(s)$ and $\partial\lambda(s)/\partial s$ remain virtually constant. These conditions allow us to replace $\lambda(s)$ with $\lambda(t)$, and $\partial\lambda(s)/\partial s$ with $\partial\lambda(t)/\partial t$. Our dynamic equations for f become

$$\frac{\partial f_i}{\partial t} = v_i(f) - \sum_{j=1}^{k} D_{ij}(f)\ \lambda_j(f) - \sum_{j=1}^{k}\sum_{l=1}^{k} N_{ij}(f)\frac{\partial \lambda_j}{\partial f_l}\frac{\partial f_l}{\partial t}. \tag{3.33}$$

We have introduced two new functions, defined by

$$D_{ij}(f) = \int_0^\tau \mathrm{C}_{\bar\rho}[\hat{L}\mathrm{i}F_j,\ \hat{L}\mathrm{i}F_i(s-t)]\ ds. \tag{3.34}$$

$$N_{ij}(f) = \int_0^\tau \mathrm{C}_{\bar\rho}[\delta F_j,\ \hat{L}\mathrm{i}F_i(s-t)]\ ds. \tag{3.35}$$

Note that D and N depend on f through the averaging with $\bar\rho$. We have written τ instead of t as the upper limit of integration. We saw in Section 1.1 that the conditions under which a macroscopic process was truly Markovian were very strict, and thence we expect the past history to have some influence on the process for a short time, though too short to measure with our macroscopic methods. Thus, over the time scales of interest, the time correlation functions in (3.32) should decay to zero almost immediately, so it should make no difference if we extend the upper limit of integration from t to some very large value τ on the molecular time scales.

With the slow variable hypothesis, we estimate the relative importance of the terms in (3.33) [67]. We take $\partial F/\partial t = \hat{L}\mathrm{i}F$ as a small quantity of

order $O(\tau)$. The organised drift (3.8), v_i, contains one factor of $\hat{L}\mathrm{i}F$, so it is formally of order $O(\tau)$. The function D contains the product $\hat{L}\mathrm{i}F_i$ times $\hat{L}\mathrm{i}F_j$ within a time integral, making it order $O(\tau^2)\ \Delta t$, with Δt standing for the time over which the integral decays. The last term includes $\partial f/\partial t$ (order $O(\tau)$ due to the organised drift) and E (order $O(\tau)\Delta t$). Keeping all the terms up to order $O(\tau^2)$,

$$\frac{\partial f_i}{\partial t} = v_i(f) - \sum_{j=1}^{k} D_{ij}(f)\ \lambda_j(f) - \sum_{j=1}^{k}\sum_{l=1}^{k} N_{ij}(f) v_l(f)\ \frac{\partial \lambda_j}{\partial f_l}. \qquad (3.36)$$

In Zwanzig's theory (see Appendix A), the presence of a projected dynamics operator makes it difficult to write down higher order terms for the time evolution of f. The second order approximation results from replacing the projected evolution with the real evolution in the time integral. The third order term would involve a complicated mix of exponentials, projection operators and exponentials of projection operators. Here, once we calculate the averages v, D and N, we can insert them recursively into (3.33). The whole process, though tedious, does not involve any mathematical difficulties (but see the discussion of Pawula's theorem below).

3.3 Heat transfer

We have already presented the equations for the time evolution of expected values. Here we derive the equations for the probability density corresponding to the macroscopic variables. We illustrate the method by deducing equations for the transfer of energy among systems.

The following pages explain how to derive equations for the evolution of the joint probability distribution, $P_F(f;\ t)$, for the values of the macroscopic variables. We will concentrate on heat transfer among several bodies at different temperatures. Should we wish to, we could easily generalise from this particular case.

Each system has its own Hamiltonian, $E_i(z_i)$, with z_i standing for the positions and momenta of all its degrees of freedom. We use $z = (z_1,\ z_2,\ \ldots,\ z_k)$ for the microstate of the combined system with Hamiltonian

$$H(z) = \sum_{i=1}^{k} E_i(z_i) + H_{int}(z). \tag{3.37}$$

As usual, we suppose that we can neglect the interaction energy compared to the energies of each subsystem, $H_{int}(z) \ll H_i(z_i)$. We indicate the macrostate with $E_1(z_1) = e_1,\ E_2(z_2) = e_2,\ \ldots,\ E_k(e_k) = e_k$, using the vector $\mathbf{e} = (e_1,\ e_2,\ \ldots,\ e_k)$. Given a probability distribution over the energies, $P_E(e)$, the corresponding relevant ensemble follows from maximising the relative entropy (2.41) subject to the constraints

$$P_E(\mathbf{e}) = \mathrm{Tr}[\rho\Psi_{\mathbf{e}}], \tag{3.38}$$

($\Psi_{\mathbf{e}}(z) = \prod_{i=1}^{k} \delta(E_i(z) - e_i) = \delta(E - \mathbf{e})$). The resulting distribution reads

$$\bar{\rho}(z) = \rho^{eq}(z)\frac{P_E(E_1(z_1),\ \ldots,\ E_k(z_k))}{\Omega(x_{E(z)})}. \tag{3.39}$$

The density of states $\Omega(x_{\mathbf{e}})$, representing

$$\Omega(x_{\mathbf{e}}) = \mathrm{Tr}[\rho^{eq}\delta(E - \mathbf{e})], \tag{3.40}$$

equals the probability of $\mathbf{e}$ at equilibrium $\Omega(x_{\mathbf{e}}) = P_E^{eq}(\mathbf{e})$. Under the conditions for microscopic reversibility, the distribution ρ^{eq} is flat, so Ω represents a genuine density of states for the energy distribution $\mathbf{e}$.

A few pages ago (page 229), we mentioned that a particular choice of relevant phase functions produces the evolution equation for P_E. We pick the delta function $\Psi_{\mathbf{e}}$ and substitute it for F in equation (3.12).

$$\frac{\partial}{\partial t}P_E(\mathbf{e},\ t) = \mathrm{Tr}[\bar{\rho}\ \hat{L}\mathrm{i}\Psi_{\mathbf{e}}] + \int_0^t (\hat{L}\mathrm{i}\Psi_{\mathbf{e}})e^{-\hat{L}\mathrm{i}(t-s)}(\hat{L}\mathrm{i}\bar{\rho} + \frac{\partial\bar{\rho}}{\partial s})\ ds. \tag{3.41}$$

Now, as we did in Section 3.1, we work through the terms searching for useful alternative expressions. We begin by transforming

$$\hat{L}\mathrm{i}\Psi_e = -\{H,\ \Psi_{\mathbf{e}}\} = -\sum_{i=1}^{k}\{H,\ E_i\}\ \frac{\partial\Psi_{\mathbf{e}}}{\partial E_i}, \tag{3.42}$$

with which we rewrite the first term as

$$\mathrm{Tr}[\bar{\rho}\ \hat{L}\mathrm{i}\Psi_{\mathbf{e}}] = \sum_{i=1}^{k}\mathrm{Tr}\left[\bar{\rho}\ (\hat{L}\mathrm{i}E_i)\ \frac{\partial\Psi_{\mathbf{e}}}{\partial E_i}\right] = -\sum_{i=1}^{k}\frac{\partial}{\partial e_i}\mathrm{Tr}[\bar{\rho}\ (\hat{L}\mathrm{i}E_i)\ \Psi_{\mathbf{e}}]. \tag{3.43}$$

The last step comes from the simple realisation that

$$\frac{\partial\Psi_{\mathbf{e}}}{\partial E_i} = \frac{\partial}{\partial E_i}\prod_{i=1}^{k}\delta(E_i - e_i) = -\frac{\partial}{\partial e_i}\prod_{i=1}^{k}\delta(E_i - e_i) = -\frac{\partial\Psi_{\mathbf{e}}}{\partial e_i}. \tag{3.44}$$

Thanks to the delta function, we can drag the probability distribution P_E out of the trace in

$$\mathrm{Tr}[\bar{\rho}\ (\hat{L}\mathrm{i}E_i)\ \Psi_{\mathbf{e}}] = \mathrm{Tr}[\rho^{eq}\ (\hat{L}\mathrm{i}E_i)\ \Psi_{\mathbf{e}}]\frac{P_E(\mathbf{e};\ t)}{\Omega(x_{\mathbf{e}})} \tag{3.45}$$

by writing $\bar{\rho}$ as in (3.39). The trace on the right divided by $\Omega(x_{\mathbf{e}})$ equals the expected value of $\hat{L}\mathrm{i}E_i$ when we distribute the energy among the subsystems according to $\mathbf{e}$. The shorter notation v_i, with

$$v_i(\mathbf{e}) = \frac{1}{\Omega(x_{\mathbf{e}})}\mathrm{Tr}[\rho^{eq}\ (\hat{L}\mathrm{i}E_i)\ \Psi_{\mathbf{e}}], \tag{3.46}$$

will denote the expected value, but remember that this definition differs from the organised drift (3.8) due to the averaging over the microcanonical

$$\hat{\rho}(z) = \delta(E(z) - \mathbf{e})\frac{\rho^{eq}(z)}{\Omega(x_{\mathbf{e}})} \tag{3.47}$$

instead of $\bar{\rho}$ (3.39).

When we deal with slow variables the evolution equation for $P_E(e;\ t)$ up to first order terms turns into

$$\frac{\partial}{\partial t}P_E(\mathbf{e};\ t) = -\alpha \sum_{i=1}^{k} \frac{\partial}{\partial e_i} v_i(\mathbf{e})\ P_E(\mathbf{e};\ t) + O(\tau^2). \tag{3.48}$$

Further on, we will argue that α unexpectedly turns out equal to approximately 2, as this factor takes into account part of the effects due to dissipation. For now, we can think of it as a parameter used to adjust the time scale for (3.48) to the results of numerical simulations or experiments.

The first order evolution equation (3.48) may well approximate a system's behaviour sufficiently closely in some cases. We will test it on a very simple Ising model for the transfer of energy. We will of course have to write discrete versions of our heretofore continuous equations, which might sound like a strange move. The reason why we decided to examine such a model is that typical examples of heat transfer in molecular dynamics have $v_i(f) = 0$, as we will show below (see Eq. ()).

We bring two one-dimensional chains of spins with periodic boundary conditions and Hamiltonians E and E' into contact by means of a small interaction term H_{int}. Then we obtain the Hamiltonian of the combined system by adding these three functions [68],

$$\begin{aligned} H(s,\ s') &= E(s) + E'(s') + H_{int}(s,\ s') \\ &= \left(-\frac{1}{2} \sum_{(i,\ j)} s_i s_j + N \right) \left(-\frac{1}{2} \sum_{(i,\ j)} s'_i s'_j + N' \right) + H_{int}(s,\ s') \\ &= E_T. \end{aligned} \tag{3.49}$$

The spins have two possible values, often referred to as "up" and "down", $s_i = +\frac{1}{2}$ and $s_i = -\frac{1}{2}$. Most of the total energy, E_T, resides either in system 1 or system 2, $H_{int}(z,\ z') \approx 0$. N and N' were just added to make the state of minimum energy lie at $H(z,\ z') = 0$. The brackets $(i,\ j)$ indicate that sites i and j are contiguous. Each spin s_i has three possible (degenerate) energy states:

1. a ground state, when $s_{i-1} = s_i = s_{i+1}$,
2. an intermediate state, when $s_{i-1} \neq s_{i+1}$, and
3. a state of maximum energy, when $s_{i-1} = s_{i+1} \neq s_i$.

We simulate the interaction between the chains by picking a random site. If we find the corresponding spin in the ground state, we leave it untouched; if in the intermediate state, we flip it. When we have chosen a site in the state of maximum energy, we flip it as well, but the conservation of energy requires that we also flip a spin in the ground state. Only the last event makes the transfer of energy between chains possible.

We have decided to focus on the simplest case, with only two Ising chains, so we have a single relevant phase function, E_1. The value of e_2 comes from subtracting e_1 from the total energy, $e_2 = E_T - e_1$. Our systems evolve in steps, so we write down a discretised version of (3.48),

$$\frac{P_E(e_1;\ t+\Delta t) - P_E(e_1;\ t)}{\Delta t} = -2\ \frac{v(e_1+\Delta e_1)P_E(e_1+\Delta e_1;\ t) - v(e_1)P_E(e_1;\ t)}{\Delta e_1}. \quad (3.50)$$

($\alpha = 2$). Unfortunately, $v(e_1+\Delta e_1)$ does not have the same value when we pick $\Delta e_1 > 0$ as when $\Delta e_1 < 0$. In fact, we cannot carry out the derivative in v because $v(e_1^-) \neq v(e_1^+)$ in general, so we replace (3.50) with a master equation that turns into (3.50) when $v(e_1^-) = v(e_1^+)$.

$$\begin{aligned}\frac{P_E(e_1;\ t+\Delta t) - P_E(e_1;\ t)}{\Delta t} =& B^-(e_1+|\Delta e_1|)P_E(e_1+|\Delta e_1|;\ t) \\ &+ B^+(e_1+|\Delta e_1|)P_E(e_1+|\Delta e_1|;\ t) \\ &- \left(B^+(e_1)+B^-(e_1)\right)P_E(e_1;\ t). \quad (3.51)\end{aligned}$$

The functions B^+ and B^- include $v(e^-)$ and $v(e^+)$,

$$B^-(e_1) = -\frac{v(e_1^-)}{\Delta e_1};\ B^+(e_1) = -\frac{v(e_1^+)}{\Delta e_1}. \quad (3.52)$$

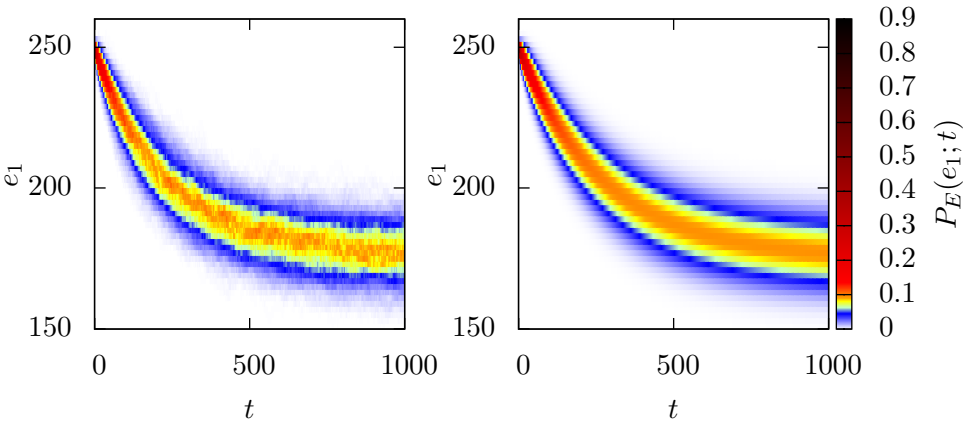

Figure 3.1: Probability distribution for e_1 versus time calculated with simulations (*left*) and with (3.53) (*right*). The Ising models had $N = N' = 500$ spins and initial energies set to $e_1 = 250$ and $E_T - e_1 = 100$.

At every time step, the systems can exchange two units of energy or none. B^+ and B^- quantify the probability of an upward or downward transition per unit time, respectively. Consequently, we choose our units so that $\Delta t = 1$ and $\Delta e_1 = 2$. Solving (3.51) for $P_E(e_1;\ t+1)$,

$$P_E(e_1;\ t+1) = B^-(e_1+2)P_E(e_1+2,\ t) + B^+(e_1-2)P_E(e_1-2,\ t) + \big(1 - (B^+(e_1) - B^-(e_1))\big)\, P_E(e_1,\ t), \quad (3.53)$$

which we can work out numerically for any initial distribution $P_E(e_1;\ 0)$ if we know the values of B^+ and B^-.

In Section 3.1 (page 105), we argued that the first order approximation considers the distribution over the states equal to the relevant distribution at all times. This means that we can calculate $B^-(e_1)$ assuming that all the configurations with the corresponding energy have the same probability

of occurrence.

$$B^-(e_1) = \frac{e_1}{N} \cdot \frac{(e_1 - 1)}{N - 1} \cdot \frac{N' - (E_T - e_1 + 1)}{N - (e_1 + 1) + N' - (E_T - e + 1)} \tag{3.54}$$

$$= \frac{e_1(e_1 - 1)(N' + e_1 - E_T - 1)}{N(N - 1)(N + N' - E - 2)}. \tag{3.55}$$

Similarly, the probability of system 1 absorbing two units of energy, B^+, equals the probability of system 2 giving them away,

$$B^+(e_1) = \frac{(E - e_1)(E - e_1 - 1)(N - e_1 - 1)}{N(N - 1)(N + N' - E_T - 2)}. \tag{3.56}$$

Figure 3.1 shows an excellent agreement between the dynamical simulations and the numerical integration of $P_E(e_1;\ t)$ with (3.53). Each Ising model consisted of 500 spins. The initial energies were $e_1 = 250$ and $e_2 = E_T - e_1 = 100$. The plot on the left, calculated by averaging over one thousand realisations, presents a slightly noisier version of the evolution on the right. Note that our "coarse-grained" equation gets both the average and the shape of the distribution right.

Making system 2 much larger than system 1, we expect the behaviour of the latter to approach the time-dependent statistics of an Ising model in contact with a heat reservoir [69], but only for high temperatures (see Section 3.5). To understand why, let us work out $P_E(e_1;\ t)$ when we set the initial values to $e_1 = 50$ and $e_2 = E_T - e_1 = 2$ in our 500-spin chains.

Figure 3.2 plots the probability distribution $P_E(e_1;\ t)$ for three different times. Even though the initial coarse-grained prediction coincides for a short time, the system starts to lag behind the relevant distribution before eventually catching up to the state of equilibrium. What causes the lag? Imagine a segment of our chains with the following configuration:

$$\ldots,\ -\frac{1}{2},\ -\frac{1}{2},\ +\frac{1}{2},\ -\frac{1}{2},\ -\frac{1}{2},\ \ldots \tag{3.57}$$

The probability of flipping the positive spin equals the probability of flipping one of the negative spins next to it. If it flips down, though, it becomes

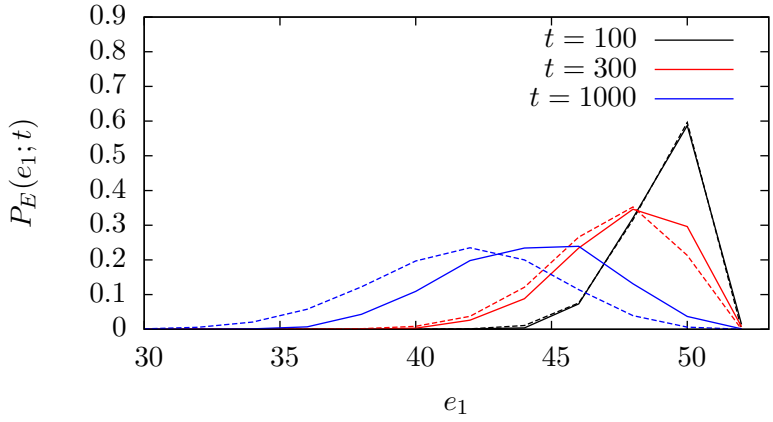

Figure 3.2: Probability distribution $P_E(e_1;\ t)$ at three different times. The solid lines correspond to averages over one thousand simulations. The dashed lines draw the coarse-grained evolution, calculated with (3.53). During the relaxation towards equilibrium, the simulated systems lag behind the theoretical prediction.

very unlikely that it flips up again. Moreover, if the negative spins on the sides flip,

$$\ldots,\ -\frac{1}{2},\ +\frac{1}{2},\ +\frac{1}{2},\ +\frac{1}{2},\ -\frac{1}{2},\ \ldots \tag{3.58}$$

then the central spin now lies in the ground state and is once again unlikely to change sign in one step. In other words, the spins tend to form "patches" of identical spins. At high temperatures, there is enough noise to make the Ising models forget these patterns, but low temperatures make some configurations for a given energy noticeably less probable, and they slow down the transfer of energy due to the scarcity of spins in a state of maximum energy. Just like we claimed in Section 2.8, the time-dependent distribution lags behind the relevant distribution because it cannot dissipate energy quickly enough.

Figure 3.2, following the heat transfer between low energy Ising chains, should suffice to convince us that we need at least a second order equation to describe some systems correctly. This brings us back to the task of rewriting the integrals in (3.41). We copy the strategy in Section 3.1 and start with an integral containing $\hat{L}\mathrm{i}\bar{\rho}$. The chain rule reveals that

$$\hat{L}\mathrm{i}\bar{\rho} = \rho^{eq} \sum_{i=1}^{k} (\hat{L}\mathrm{i}E_i) \frac{\partial}{\partial E_i} \left(\frac{P_E(E(z);\ t)}{\Omega(x_{E(z)})} \right). \tag{3.59}$$

We insert this expression into the first integral and obtain

$$\begin{aligned} &\mathrm{Tr}\left[(\hat{L}\mathrm{i}\bar{\rho})\ e^{\hat{L}\mathrm{i}(t-s)} \hat{L}\mathrm{i}\Psi_{\mathbf{e}} \right] \\ &\qquad = \sum_{i=1}^{k} \mathrm{Tr}\left[\rho^{eq} \frac{\partial}{\partial E_i} \left(\frac{P_E(E;\ s)}{\Omega(x_E)} \right) (\hat{L}\mathrm{i}E_i) e^{\hat{L}\mathrm{i}(t-s)} \hat{L}\mathrm{i}\Psi_{\mathbf{e}} \right]. \end{aligned} \tag{3.60}$$

Applying (3.42) and (3.44), the integral turns into

$$\begin{aligned} &\mathrm{Tr}\left[\rho^{eq} \frac{\partial}{\partial E_i} \left(\frac{P_E(E;\ s)}{\Omega(x_E)} \right) (\hat{L}\mathrm{i}E_i)\ e^{\hat{L}\mathrm{i}(t-s)} \hat{L}\mathrm{i}\Psi_{\mathbf{e}} \right] \\ &\quad = \sum_{j=1}^{k} \frac{\partial}{\partial e_j} \mathrm{Tr}\left[\rho^{eq} \frac{\partial}{\partial E_i} \left(\frac{P_E(E;\ s)}{\Omega(x_E)} \right) (\hat{L}\mathrm{i}E_i)\ e^{\hat{L}\mathrm{i}(t-s)} (\hat{L}\mathrm{i}E_j)\ \Psi_{\mathbf{e}} \right]. \end{aligned} \tag{3.61}$$

We multiply by the integral of $\Psi_{\mathbf{e}'}$ over $\mathbf{e}'$ to get P and Ω out of the trace,

$$\begin{aligned} &\mathrm{Tr}\left[\rho^{eq} \frac{\partial}{\partial E_i} \left(\frac{P_E(E;\ s)}{\Omega(x_E)} \right) (\hat{L}\mathrm{i}E_i)\ e^{\hat{L}\mathrm{i}(t-s)} (\hat{L}\mathrm{i}E_j)\ \Psi_{\mathbf{e}} \right] \\ &\quad = \int \mathrm{Tr}\left[\Psi_{\mathbf{e}'} \rho^{eq} (\hat{L}\mathrm{i}E_i)\ e^{\hat{L}\mathrm{i}(t-s)} (\hat{L}\mathrm{i}E_j)\ \Psi_{\mathbf{e}} \right] \frac{\partial}{\partial e_i'} \frac{P_E(\mathbf{e}';\ s)}{\Omega(x_{\mathbf{e}'})}\ d\mathbf{e}'. \end{aligned} \tag{3.62}$$

The trace contains two delta functions,

$$\mathrm{Tr}[\Psi_{\mathbf{e}'} \cdots \Psi_{\mathbf{e}}] = \mathrm{Tr}[\delta(E - \mathbf{e}') \cdots \delta(E - \mathbf{e})] = \delta(\mathbf{e}' - \mathbf{e})\ \mathrm{Tr}[\Psi_{\mathbf{e}'} \cdots]. \tag{3.63}$$

The outer delta function cancels the integral over $\mathbf{e}'$, replacing $\mathbf{e}'$ with $\mathbf{e}$ and leaving (3.60) as

$$\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial}{\partial e_j}\mathrm{Tr}\left[\Psi_{\mathbf{e}}\rho^{eq}(\hat{L}\mathrm{i}E_i)\ e^{\hat{L}\mathrm{i}(t-s)}(\hat{L}\mathrm{i}E_j)\right]\frac{\partial}{\partial e_i}\frac{P_E(\mathbf{e};\ s)}{\Omega(x_{\mathbf{e}})}$$
$$=\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial}{\partial e_j}\Omega(x_{\mathbf{e}})\ \mathrm{C}_{\hat{\rho}}[\hat{L}\mathrm{i}E_i,\ \hat{L}\mathrm{i}E_j(t-s)]\frac{\partial}{\partial e_i}\frac{P_E(\mathbf{e};\ s)}{\Omega(x_{\mathbf{e}})}. \quad (3.64)$$

The time correlation function depends on $\mathbf{e}$ through the microcanonical reference distribution $\hat{\rho}$ (3.47).

The last step involves transforming the trace containing $\partial\bar{\rho}/\partial s$,

$$\mathrm{Tr}\left[\frac{\partial\bar{\rho}}{\partial s}e^{\hat{L}\mathrm{i}(t-s)}\hat{L}\mathrm{i}\Psi_{\mathbf{e}}\right]=\mathrm{Tr}\left[\frac{\partial}{\partial s}\left(\frac{P_E(E;\ s)}{\Omega(x_{\mathbf{e}})}\right)\rho^{eq}e^{\hat{L}\mathrm{i}(t-s)}\hat{L}\mathrm{i}\Psi_e\right]. \quad (3.65)$$

Repeating the tricks in the previous paragraphs, we transfer the action of the Liouvillian to the relevant phase functions and extract P and Ω from the trace.

$$-\sum_{i=1}^{k}\frac{\partial}{\partial e_i}\int\mathrm{Tr}\left[\Psi_{\mathbf{e}'}\rho^{eq}e^{\hat{L}\mathrm{i}(t-s)}\hat{L}\mathrm{i}E_j\Psi_{\mathbf{e}}\right]\ \frac{\partial}{\partial s}\left(\frac{P_E(\mathbf{e}';\ s)}{\Omega(x_{\mathbf{e}'})}\right)\ d\mathbf{e}'$$
$$=-\sum_{i=1}^{k}\frac{\partial}{\partial e_i}\mathrm{Tr}\left[\Psi_{\mathbf{e}}\rho^{eq}e^{\hat{L}\mathrm{i}(t-s)}\hat{L}\mathrm{i}E_j\right]\ \frac{\partial}{\partial s}\left(\frac{P_E(\mathbf{e};\ s)}{\Omega(x_{\mathbf{e}})}\right). \quad (3.66)$$

The trace on the right, divided by $\Omega(x_{\mathbf{e}})$ equals $v_i(\mathbf{e}(t-s))$ or, in other words, the value of v_i (3.46) in the past given that the present macrostate is $\mathbf{e}$. We said in the previous section that we are mostly interested in slow variables, so we argue that the value of v in the past was not very different from its present value,

$$v(\mathbf{e}(t-s))\approx v(\mathbf{e}). \quad (3.67)$$

We have run into a very delicate hypothesis because we have no reason to believe that the integrand decays as time goes by. We must think through

this assumption carefully in every particular application. Although we will take it for granted in what follows, because it simplifies our expressions considerably, we write down our warning here: *beware equation* (3.67).

Because Ω does not depend on time,

$$-\sum_{i=1}^{k}\frac{\partial}{\partial e_i}v_i(\mathbf{e})\Omega(x_{\mathbf{e}})\frac{\partial}{\partial s}\left(\frac{P_E(\mathbf{e};\ s)}{\Omega(x_{\mathbf{e}})}\right)=-\sum_{i=1}^{k}\frac{\partial}{\partial e_i}v_i(\mathbf{e})\frac{\partial}{\partial s}P_E(\mathbf{e};\ s). \quad (3.68)$$

When we insert this expression into the time integral,

$$\begin{aligned}-\int_0^t\sum_{i=1}^{k}\frac{\partial}{\partial e_i}v_i(\mathbf{e})\frac{\partial}{\partial s}P_E(\mathbf{e};\ s)\ ds &= -\sum_{i=1}^{k}\frac{\partial}{\partial e_i}v_i(\mathbf{e})\int_0^t\frac{\partial}{\partial s}P_E(\mathbf{e};\ s)\ ds\\ &= -\sum_{i=1}^{k}\frac{\partial}{\partial e_i}v_i(\mathbf{e})P_E(\mathbf{e};\ s), \qquad (3.69)\end{aligned}$$

we find that the third term equals the first term! This explains our choice of $\alpha = 2$ above. Mixing in all the ingredients (3.48), (3.64), and (3.69), we obtain the evolution of $P_E(\mathbf{e};\ t)$ in the form of the generalised Fokker-Planck equation below.

$$\begin{aligned}\frac{\partial P_E(\mathbf{e},\ t)}{\partial t} =& -2\sum_{i=1}^{k}\frac{\partial}{\partial e_i}v_i(\mathbf{e})P_E(\mathbf{e};\ t)\\ &-\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial}{\partial e_j}\Omega(x_{\mathbf{e}})\int_0^t \mathrm{C}_{\hat{\rho}}[\hat{L}\mathrm{i}E_i,\hat{L}\mathrm{i}E_j(t-s)]\frac{\partial}{\partial e_i}\frac{P_E(\mathbf{e};\ s)}{\Omega(x_{\mathbf{e}})}\ ds.\end{aligned} \quad (3.70)$$

For the *exact* evolution equation, disregard equation (3.67) and replace the first term with

$$-\sum_{i=1}^{k}\frac{\partial}{\partial e_i}v_i(\mathbf{e})P_E(\mathbf{e};\ t)-\sum_{i=1}^{k}\frac{\partial}{\partial e_i}\int_0^t v_i(\mathbf{e}(t-s))\ \frac{\partial}{\partial s}P_E(\mathbf{e};\ s)\ ds.$$

Tradition holds that we should keep terms only up to order $O(\tau^2)$, perhaps because of Pawula's theorem. Pawula showed that if we use the

Kramers-Moyal expansion to derive a generalised Fokker-Planck equation from a master equation, then any finite truncation of the expansion with more than two terms implies that all the terms of order higher than two must vanish [70]. Furthermore, higher order truncations produce negative values for the distribution functions and require additional boundary conditions (some researchers have argued that these inconsistencies have only a small impact on the solutions [71, 72]).

Pawula's article warned the readers that revealing the inconsistency of keeping more than two terms was not the same as proving that the two terms were a good approximation. Though the detailed discussion of Pawula's theorem lies beyond the scope of this investigation, here I would simply like to say that a series expansion does not have to be unique. Both expressions below add to two.

$$1 + 1 + 0 + 0 + 0 + \ldots = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots \tag{3.71}$$

Including more terms improves the approximation on the right, but does not affect the left hand side beyond the second term. I see no reason why an alternative derivation of a Fokker-Planck equation should necessarily suffer from the same limitations as the Kramers-Moyal expansion, but this is a subject for further research. Nevertheless, our own derivation brought us to an exact generalised Fokker-Planck equation in which, surprisingly, there are no terms of order $O(\tau^3)$ or higher.

Zwanzig came up with a different expression for the exact generalised Fokker-Planck equation, also with only two terms [73]. As we have already mentioned, the main interest of (3.70), when compared to Zwanzig's result, lies in the absence of the projected dynamics evolution operator $\exp(\mathrm{Q}\hat{L}it)$ (see Appendix A).

Many Hamiltonians have the standard form

$$H(q,\ p) = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + V(q_1,\ \ldots,\ q_N), \tag{3.72}$$

with the potential energy independent from the values of the momenta. If our combined system had such a Hamiltonian, then the interaction term in (3.37) would not include the momenta.

$$H(q,\ p) = \sum_{j=1}^{k} \sum_{i=1}^{N_j} \left(\frac{p_{j,i}^2}{m_{j,i}} + V(q_{j,1},\ \ldots,\ q_{j,N_j}) \right) + H_{int}(q_{1,1},\ \ldots,\ q_{j,i},\ \ldots,\ q_{k,N_k}). \quad (3.73)$$

The partial derivatives of H_{int} with respect to any momentum would vanish,

$$\frac{\partial H_{int}}{\partial p_{j,i}} = 0, \quad (3.74)$$

making the Poisson bracket $\{H,\ E_j\}$ equal

$$-\hat{L}\mathrm{i}E_j = \{H_{int},\ E_j\} = \sum_{i=1}^{N_j} \frac{\partial H_{int}}{\partial q_{j,i}} \frac{p_{j,i}}{m_{j,i}} \quad (3.75)$$

(using the fact that $\{E_i,\ E_j\} = 0$). In turn, this result implies that v_j (3.46) must vanish as well, because the expression in the trace becomes an odd function of the momenta

$$v_i(\mathbf{e}) = \frac{1}{\Omega(x_\mathbf{e})} \mathrm{Tr} \left[\rho^{eq} \left(\sum_{i=1}^{N_j} \frac{\partial H_{int}}{\partial q_{j,i}} \frac{p_{j,i}}{m_{j,i}} \right) \Psi_\mathbf{e} \right] = 0. \quad (3.76)$$

This result holds only when the bodies have no net momentum in any direction. Similarly, the non-diagonal diffusion coefficients contain integrals over odd functions of the momenta, so $D_{ij}(\mathbf{e}) = 0$ for $i \neq j$, leaving only the diagonal terms in the Fokker-Planck equation,

$$D_{jj}(\mathbf{e}) = \int_0^\infty \mathrm{C}_{\hat{\rho}} \left[\sum_{i=1}^{N_j} \mathbf{F}_i \cdot \mathbf{v}_i,\ \sum_{i=1}^{N_j} \mathbf{F}_i(t-s) \cdot \mathbf{v}_i(t-s) \right] ds. \quad (3.77)$$

Here, $\mathbf{F}_i$ stands for the total external force on particle i, and $\mathbf{v}_i$ for its velocity,

$$\mathbf{F}_i(t) = -\frac{\partial H_{int}}{\partial q_i};\ \mathbf{v}_i = \frac{p_i}{m_i}. \tag{3.78}$$

Time arguments indicate the presence of an exponential operator,

$$F(t-s) = e^{\hat{L}\mathrm{i}(t-s)}F. \tag{3.79}$$

3.4 Ideal gases

The generalised Fokker-Planck equation for heat flow (3.70) correctly predicts the evolution of the probability distribution for the energies of a homogenous mixture of two ideal gases initially at different temperatures.

A theorist will proudly contemplate the exact equation for a phenomenon. Practical considerations soon teach us that exact equations often work only as a framework within which to write approximate formulae for real applications. In the case of heat transfer, we will commonly take for granted that the bodies in contact experience relative changes in their internal energies much more slowly than the changes in kinetic energies of the particles they are made from[6].

The slow variable version of (3.70) brings the probability P_E outside the time integral.

$$\begin{aligned}\frac{\partial}{\partial t}P_E(\mathbf{e},\ t) = &-2\sum_{i=1}^{k}\frac{\partial}{\partial e_i}v_i(\mathbf{e})P_E(\mathbf{e};\ t)\\ &-\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial}{\partial e_j}\Omega(x_{\mathbf{e}})D_{ij}(\mathbf{e})\ \frac{\partial}{\partial e_i}\frac{P_E(\mathbf{e};\ t)}{\Omega(x_{\mathbf{e}})}.\end{aligned} \tag{3.80}$$

[6]Part of the material in this section was presented as a poster [74] during the FisEs 2012 conference in Palma.

The *diffusion coefficient*,

$$D_{ij}(\mathbf{e}) = \int_0^\infty \mathrm{C}_{\hat{\rho}}[\hat{L}\mathrm{i}E_i,\ \hat{L}\mathrm{i}E_j(t-s)]\ ds, \tag{3.81}$$

depends on $\mathbf{e}$ through $\hat{\rho}$ (3.47). In most practical settings, the term for v will vanish (3.76), as will the off-diagonal diffusion coefficients (3.77). We can adapt the resulting equation to many different situations, bearing in mind that two conditions must be met, as we will see in the following chapter. First, the bodies must not carry out macroscopic work on each other. Second, energy must travel relatively quickly within the bodies, so that temperature gradients do not have enough time to develop. We will come back to these conditions in the next chapter.

We approximate the time correlation function in D_{jj} (3.81) by

$$\mathrm{C}_{\hat{\rho}}[\hat{L}\mathrm{i}E_j,\ \hat{L}\mathrm{i}E_j(t-s)]\ ds \approx e^{-(t-s)/\tau}\ \mathrm{Tr}\left[\hat{\rho}\ (\hat{L}\mathrm{i}E_j)^2\right], \tag{3.82}$$

which coincides with the time correlation function at $t = s$, and may or may not resemble it for smaller values of s. In any case, our approximation implies that

$$D_{jj}(\mathbf{e}) \approx \tau\ \mathrm{Tr}\left[\hat{\rho}\ (\hat{L}\mathrm{i}E_j)^2\right], \tag{3.83}$$

and we will leave it for numerical or experimental data to decide if this is reasonable.

Note that the approximation (3.83) makes it possible in principle to calculate the diffusion coefficients analytically for simple potential energies. Neglecting the interaction energy, the microcanonical equilibrium distribution becomes

$$\rho^{eq}(z) = \frac{\delta(\sum_{i=1}^k E_i - E_T)}{\Omega(x_{E(z)})}, \tag{3.84}$$

and it can be brought out of the trace in (3.83),

$$D_{jj}(\mathbf{e}) = \tau\ \frac{\delta(\sum_{i=1}^k e_i - E_T)}{(\Omega(x_\mathbf{e}))^2} \sum_{i=1}^{N_j} \mathrm{Tr}\left[\left(\frac{p_{j,i}}{m_{j,i}}\frac{\partial H_{int}}{\partial q_{j,i}}\right)^2 \Psi_\mathbf{e}\right]. \tag{3.85}$$

When all the particles in the system have the same mass, $m_{j,i} = m_j$, we can carry out the integral

$$\begin{aligned}\int \frac{p_{j,i}}{m_j}\delta(E_j - e_j)\ dp_j &= \frac{1}{N_j}\int \left(\sum_{i=1}^{k}\frac{p_{j,i}^2}{2m_j}\right)\delta(E_j - e_j)\ dp_j \\ &= \frac{e_j - V_j(q_j)}{N_j}\int \delta(E_j - e_j)\ dp_j. \end{aligned} \tag{3.86}$$

The last step follows from writing the Hamiltonian for system j as $E_j(q_j,\ p_j) = \sum_{i=1}^{N_j} p_{j,i}^2/2m_j + V_j(q_j)$, so the delta function allows us to replace the kinetic energy in the integral with $e_j - V_j$. The remaining integral on the right equals the surface of an N_j-dimensional hypersphere with radius $\sqrt{2m_j(e_j - V_j(q))}$.

$$\int \delta(E_j - e_j)\ dp_j = \frac{2\pi^{N_j/2}}{\Gamma\left(\frac{N_j}{2}\right)}\left(2m_j(e_j - V_j(q_j))\right)^{(N_j-1)/2}. \tag{3.87}$$

Here, Γ refers to Euler's gamma function,

$$\Gamma(s) = \int_0^\infty x^{s-1}e^{-s}\ dx. \tag{3.88}$$

Integrating over p in the expression for the diffusion coefficient yields

$$\begin{aligned} D_{jj}(\mathbf{e}) =& \tau\ \frac{\delta(\sum_{i=1}^{k} e_i - E_T)}{(\Omega(x_\mathbf{e}))^2}\frac{\pi^{N_j/2}(2m_j)^{(N_j-1)/2}}{N_j\Gamma\left(\frac{N_j}{2}\right)} \\ &\times \sum_{i=1}^{k}\int \left(\frac{\partial H_{int}}{\partial q_{j,i}}\right)(e_j - V_j(q_j))^{(N_j+1)/2}\ dq_j \left(\prod_{i\neq j}\Omega(y_{e_i})\right). \end{aligned} \tag{3.89}$$

The sets y_{e_i} at the end contain the microstates z_i such that $E_i(z_i) = e_i$.

Even though we can evaluate (3.89) for some simple potentials, this does not mean that it is practical to do so. For ideal gases we can use

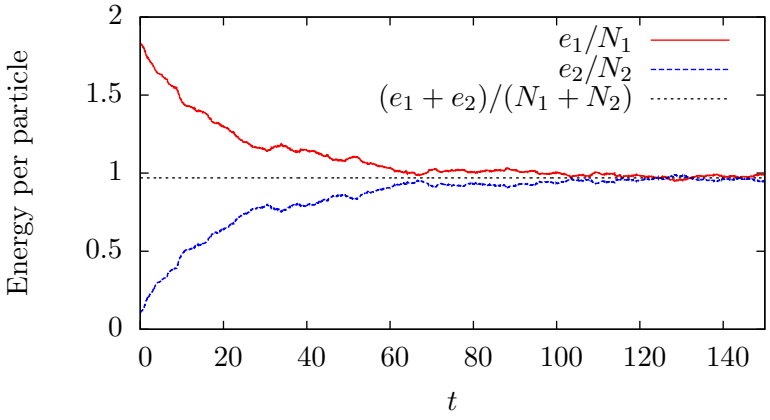

Figure 3.3: Time evolution of the average energy per particle in a homogeneous mixture of two identical hard-sphere gases ($N_1 = N_2 = 1000$) at different initial temperatures. The hotter gas (*in red*) begins with an energy $e_1 \approx 1.7$, while the other gas (*blue*) starts off with $e_2 \approx 0.3$. The dotted line indicates the average energy per particle for the combined system.

the approximation $V_j(q_j) \approx 0$, but the expressions will still be long and complicated when we wish to track more than three or four particles. In most cases, running a molecular dynamics simulation solves the problem quickly with sufficient accuracy.

It may seem that we have gained nothing if we now need to resort to molecular dynamics and solve the trajectories numerically. On the contrary, thanks to the microcanonical $\hat{\rho}$, we can calculate the diffusion coefficients numerically with *equilibrium* molecular dynamics or Monte Carlo sampling [75].

A mixture of two ideal gas systems at different temperatures will illustrate the relaxation towards equilibrium. We imagine gases dilute enough to make the mean free time comparable to the time it takes them to mix. The molecular dynamics simulations were implemented with an event-driven algorithm [21] for hard spheres. Although hard-sphere interactions do not

count as independent of the momentum in the strict sense, we will assume that they approximate Lennard-Jones potentials or something similar, which do.

The conservation of energy allows us to forget about one of the internal energies, just like in the Ising chains example. We only need to register the values of e_1, which should obey the equation

$$\frac{\partial P_E(e_1;\ t)}{\partial t} = \frac{\partial}{\partial e_1}\Omega(x_{e_1})D(e_1)\ \frac{\partial}{\partial e_1}\frac{P_E(e_1;\ t)}{\Omega(x_{e_1})}. \tag{3.90}$$

Then $e_2 = E_T - e_1$. By simple inspection, we can tell that the probability distribution at equilibrium equals $\Omega(x_{e_1})$ (3.40). We write $\Omega(x_{e_1})$ for brevity, instead of $\Omega(x_{e_1,E_T-e_1})$, meaning the density of states when system 1 has internal energy e_1 and system 2 has $e_2 = E_T - e_1$. This becomes important when we calculate the derivative of its logarithm with respect to e_1. Because we neglect the interaction energy,

$$\Omega(x_{e_1,E_T-e_1}) = \Omega(y_{e_1})\ \Omega(w_{E_T-e_1}). \tag{3.91}$$

The macrostates y_e and $w_{e'}$ apply to systems 1 and 2, respectively.

Ω contains the information on the equilibrium probability of each value of e_1 and hence we can use it to quantify equilibrium fluctuations. We see in Figure 3.4 that the probability of finding e_1/N_1 departing significantly from its expected value diminishes as the number of particles increases.

Instead of working out the value of Ω numerically, it is much simpler to use the definition of temperature (1.104). If we know the value of e_1, then the derivative of entropy with respect to e_1 works out as

$$\begin{aligned}\frac{\partial S_B(x_{e_1})}{\partial e_1} &= k_B\ \frac{\partial}{\partial e_1}\ln(\Omega(x_{e_1})) \\ &= k_B\ \frac{\partial}{\partial e_1}\ln(\Omega(y_{e_1})) - k_B\ \frac{\partial}{\partial(E_T - e_1)}\ln(\Omega(w_{E_T-e_1})) \\ &= \frac{1}{T_1(e_1)} - \frac{1}{T_2(e_2)}\end{aligned} \tag{3.92}$$

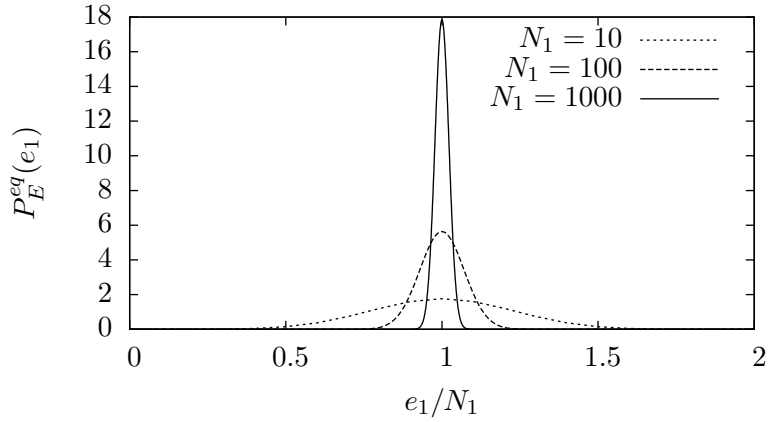

Figure 3.4: Equilibrium probability distribution for the energy per particle e_1/N_1 for an N_1-particle ideal gas mixed with another gas ($N_2 = N_1$). The total energy was set to $E_T = N_1 + N_2$. As the number of particles increases, the probability distribution narrows.

(remembering the meaning of x_{e_1} (3.91)). Carrying out the inner derivative in (3.90)

$$\frac{\partial P_E(e_1;\ t)}{\partial t} = \frac{\partial}{\partial e_1} D(e_1)\ \left(\frac{\partial P_E(e_1;\ t)}{\partial e_1} - P_E(e_1;\ t) \frac{\partial}{\partial e_1} \ln(\Omega(x_{e_1})) \right) \quad (3.93)$$

and introducing (3.92),

$$\begin{aligned} \frac{\partial P_E(e_1;\ t)}{\partial t} = &- \frac{\partial}{\partial e_1} \left(\frac{1}{k_B T_1(e_1)} - \frac{1}{k_B T_2(E_T - e_1)} \right)\ P_E(e_1;\ t) \\ &+ \frac{\partial}{\partial e_1} D(e_1)\ \frac{\partial P_E(e_1;\ t)}{\partial e_1}. \end{aligned} \quad (3.94)$$

We interpret the function $T_j(e_j)$ as the temperature that system j would have if it were isolated and in equilibrium with internal energy equal to e_j. The law of equipartition tells us that our ideal gases satisfy

$$T_1(e_1) = \frac{2}{3} \frac{e_1}{N_1 k_B},\ T_2(E_T - e_1) = \frac{2}{3} \frac{E_T - e_1}{N_2 k_B}. \quad (3.95)$$

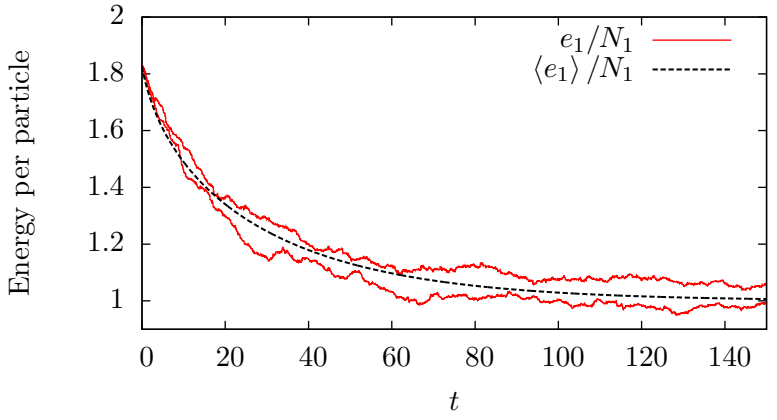

Figure 3.5: Two instances of the evolution of e_1/N_1 calculated with molecular dynamics (*red*) and the evolution of the mean value $\langle e_1 \rangle /N_1$ as predicted by equation (3.96) once the free parameter D_0 has been adjusted ($N_1 = N_2 = 1000$).

Equation (3.94) exhibits two terms with a very clear physical interpretation. The drift term, with one partial derivative, responds to temperature differences and drives the systems towards thermal equilibrium. The diffusion term at the end adjusts the shape of the distribution even when the system has reached the equilibrium temperature.

We will compare (3.94) to the results of our numerical simulations. Before we carry out the numerical integration of P_E, we need to know the value of the diffusion coefficient as a function of the energy, $D(e_1)$. For the moment, we will boldly treat it like a constant, $D(e) = D_0$.

In many cases, we can neglect fluctuations because macroscopic distributions often peak sharply around their average value [42], $P(e_1;\ t) \approx \delta(e_1 - \langle e_1 \rangle)$. With a constant diffusion coefficient, an easy equation for $\langle e_1 \rangle$ results. Multiplying equation time e_1 and integrating by parts with respect

to e_1,

$$\begin{aligned}\frac{\partial \langle e_1 \rangle}{\partial t} &= D_0 \int_0^\infty \left(\frac{1}{k_B T_1(e_1)} - \frac{1}{k_B T_2(E_T - e_2)} \right) \; P(e_1;\; t) \; de_1. \\ &= D_0 \; \left(\frac{1}{k_B T_1(e_1)} - \frac{1}{k_B T_2(E_T - e_2)} \right). \end{aligned} \tag{3.96}$$

Figure 3.5 compares a couple of realisations to the numerical solution of (3.96) once we have adjusted D_0. The predicted behaviour tracks the realisations reasonably closely, but the fluctuations are obviously still relevant for a system of this size (see Figure 3.4).

Turning to the calculation of the diffusion coefficient, we have shown above that $D(e_1)$ equals some time τ multiplied by the expected value of $(\hat{L}iE_1)^2$ over all the microstates such that $E_1(z_1) = e_1$. We can carry out this average by simulating the movement of the gases while ignoring the effects of collisions between system 1 and system 2.

To estimate τ, we search for the value that makes the time correlation function vanish. The collisions that transfer energy between system 1 to system 2 have no statistical correlations, so τ should equal the mean free time for system 1, $\tau \approx 0.025$, that is, the average interval separating a collision involving a particle from system 1 and another from system 2.

Figure 3.6 contains the averages calculated with molecular dynamics and a parabolic fit.

The integration of partial differential equations like (3.94) poses many technical difficulties. Scientists often calculate many runs of the stochastic process associated with a Fokker-Planck equation to construct a histogram that displays the evolution of the probabilities or to determine how the expected values change. Here, the numerical integration was carried out with the Crank-Nicolson method [76], which avoided the instabilities when it was provided with a smooth initial distribution. Clearly, we cannot apply the same technique to distributions in many dimensions.

Figure 3.7 compares several realisations of the molecular dynamics simulations to the evolution of the probabilities predicted by (3.94). The excellent agreement between the theory and the simulations constitutes an argument in favour of Boltzmann's molecular chaos hypothesis (Stosszahlansatz).

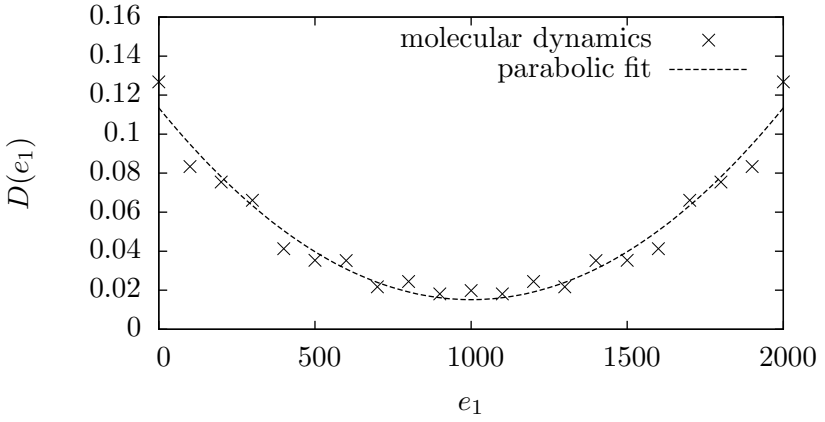

Figure 3.6: Diffusion coefficient versus the energy of a gas in a homogeneous mixture of two identical hard-sphere gases. The crosses display the simulation results, and the curve draws the best parabolic fit, which we used in the numerical integration of (3.94).

Through collisions, gases quickly forget their previous states, making the velocities involved in any given interaction completely independent of the previous collision. In the following chapter, we will analyse two anharmonic chains equilibrating. There we will show that the presence of memory effects makes the evolution towards equilibrium quite different from what we see in Figure 3.7.

3.5 Thermostating

This section envisages systems in contact with a heat bath and explains how to derive the coarse-grained equations of motion when we leave the exact state of the bath out of our description.

The easiest class of molecular dynamics algorithms simulates isolated systems, yet real experiments are often set up in thermal equilibrium with

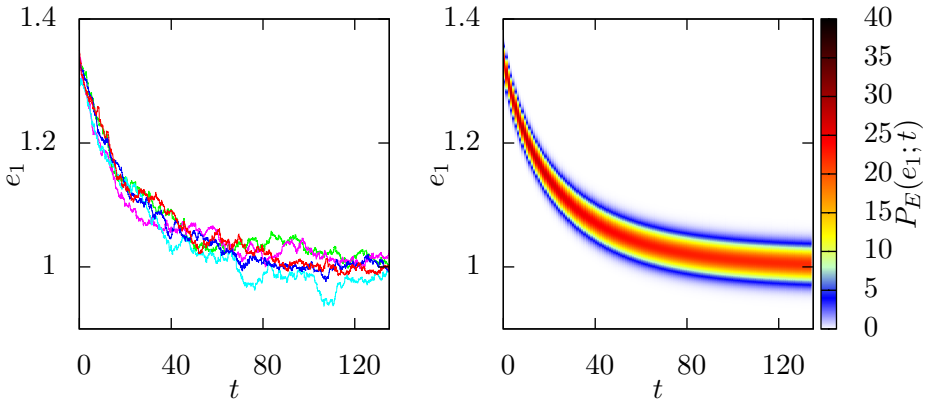

Figure 3.7: Several realisations of the evolution of the energy e_1 of a hard-sphere gas (*left*) interacting with another identical gas ($N_1 = N_2 = 1000$) compared to the evolution of the probability of e_1 in time (*right*).

their environments. If we wish to compare empirical data to simulations, we need a computational device that removes excess heat or lets it flow into the system in response to changes in the temperature. The popular deterministic option will be discussed in Section 4.2. Here we will concentrate on stochastic dynamics.

Let us bring back our Ising models, with an extremely large system 2. In that case, energy variations in system 1 will hardly affect the temperature of system 2, $T_2 = T$, so system 1 will gradually approach T. When we focus on a single site, we find a simple two-state system in the midst of a sea of spin sites at temperature T. In Chapter 2, Section sec:MaxREnt, we derived the least-biased probability distribution for a body in contact with a heat reservoir. The probabilities were proportional to the exponential of $\epsilon/(k_B T)$, where ϵ represents the energy of the body.

Instead of using a one-dimensional chain, I decided in favour of a 10×10 model with periodic boundary conditions, as it was still easy to simulate,

but complex enough to illustrate phase change. The dynamic evolution was followed with a Markov-chain Monte Carlo programme [75] known as the Metropolis algorithm: we choose a random spin and see if it moves to a state with higher probability if we flip it. When it does, we change its state. Otherwise, we flip it with a probability equal to the new probability divided by that of the present state.

When describing relaxation towards equilibrium, molecular dynamics differ from Markov-chain Monte Carlo algorithms by a constant factor in the time scale [77], which depends on the interaction driving the spin flips. Here we imagine that the Monte Carlo algorithm constitutes the real dynamics that we wish to coarse-grain.

We began with a random initial state. All configurations were equally likely. The binomial distribution gave us the initial probability P_M for each value of the magnetisation. If u represents the number of up spins,

$$P_M(2u - N) = \binom{N}{u} \left(\frac{1}{2}\right)^N . \tag{3.97}$$

At high temperatures, the probability remains concentrated around $M = 0$, with (3.97) representing the equilibrium distribution at infinite temperature. As the temperature lowers, the distribution widens until, at sufficiently low temperatures, the system achieves a net magnetisation. Surprisingly, our simple coarse-grained model (3.51) predicts these features in a qualitative fashion.

As in Section 3.3, we calculated the transition probabilities in the master equation (3.51) by treating all the microstates with a given magnetisation as equally likely. Figure 3.8 compares the simulation data to the approximate evolution of the probabilities. The blue histogram was constructed by storing the magnetisation of a thousand simulation runs at time $t = 9000$, well into the equilibrium regime for $k_B T = 5$. The theoretical curve was not too bad, considering that we had neglected the diffusion term completely.

The red curve clearly overestimated the speed at which the model reached its net magnetisation due to the lag we already mentioned, nor did it get the shape of the distribution right. Nonetheless, it failed only

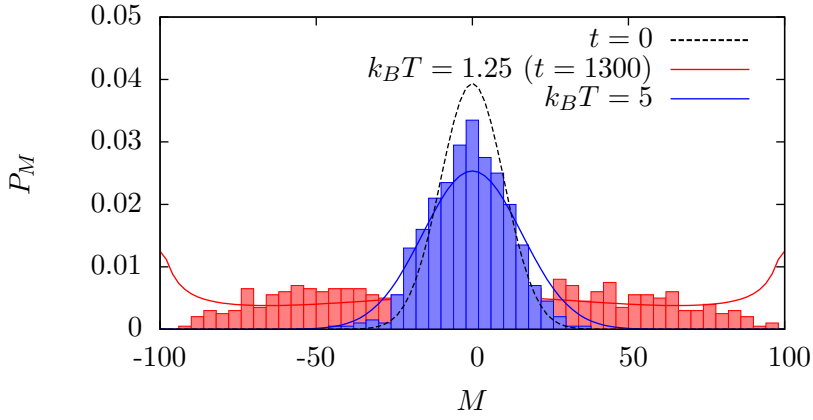

Figure 3.8: Initial distribution (*dashed line*) and evolved distributions for the magnetisation of a 10×10 Ising model with periodic boundary conditions in weak contact with a heat reservoir. Solid lines represent the approximation calculated numerically with (3.51). The histograms include data from 1000 simulation runs. The blue lines plot the equilibrium regime for $k_BT = 5$ and the red lines display a snapshot of the evolution towards the magnetised states at $k_BT = 2.5$.

when the model cooled down. The approximate equation actually mirrored the first stages of the transient evolution towards equilibrium quite accurately, as shown in Figure 3.9.

We have already explained that typical models have potentials that do not depend on the momenta, and that $v_i = 0$ in that case. Therefore, trying to apply the first order approximation (3.48) is out of the question.

Heat reservoirs keep their temperature constant when their energy fluctuates[7], $T_2(E_T - e_1) = T$, so the slow variable equation describing relaxation

[7]See Feynman's discussion of the constancy of temperatures in Section 1.1 of his *Statistical Mechanics. A set of Lectures* [30].

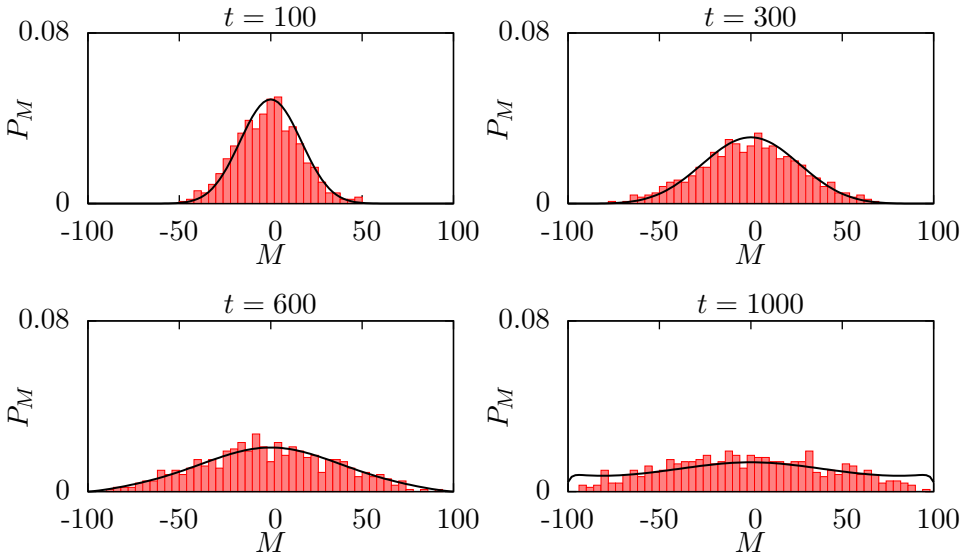

Figure 3.9: Initial stages in the evolution towards a magnetised state in the 10×10 Ising model with periodic boundary conditions in weak contact with a heat reservoir at $k_B T = 1.25$. The histograms display data from 1000 simulation runs at the corresponding times. The solid line draws the master equation prediction (3.51).

to equilibrium with the reservoir becomes

$$\frac{\partial P_E(e_1;\ t)}{\partial t} = -\frac{\partial}{\partial e_1}\left(\frac{1}{k_B T_1(e_1)} - \frac{1}{k_B T}\right) P_E(e_1;\ t) + \frac{\partial}{\partial e_1} D(e_1)\ \frac{\partial P_E(e_1;\ t)}{\partial e_1}. \tag{3.98}$$

Markov-chain Monte Carlo methods effectively sample stationary distributions, but when the distributions change we need an independent method to establish the changing relative probabilities of the energies corresponding to each tentative step. Equation (3.98) provides precisely this dynamic distribution, encouraging the investigation of nonequilibrium Markov-chain Monte Carlo sampling.

How about molecular dynamics? What laws does a system obey when in contact with a heat reservoir? To answer this question, we need the Fokker-Planck equation for the evolution of the microstate of our system of interest, z_1. We hypothesise that the state of our system does not significantly affect the behaviour of the reservoir, which remains at the same temperature T throughout. If we know the total energy of the combined system, E_T, the microstate z_1 determines the energy stored in the bath.

We follow the same steps explained in Section 3.3 to derive a Fokker-Planck equation for the positions and momenta. Once again, the Hamiltonian $H(z) = E_1(q,\ p) + E_2(z') + H_{int}(q,\ p,\ z')$ consists of a system Hamiltonian, E_1, a reservoir Hamiltonian, E_2, and a negligible interaction term, H_{int}. To keep the formulae compact, let us agree to denote positions with odd indices and save the even numbers for momenta.

$$f_{2i-1} = q_i;\ f_{2i} = p_i. \tag{3.99}$$

These will be our relevant variables. The equation for the evolution of P_F emerges as

$$\begin{aligned}\frac{\partial}{\partial t} P_F(f;\ t) = &-\sum_{i=1}^{k} \frac{\partial}{\partial f_i}\left(2\ v_i(f) + \sum_{j=1}^{k} \frac{D_{ij}(f)}{k_B}\frac{\partial S_B(x_f)}{\partial f_j}\right) P_F(f;\ t) \\ &+ \sum_{i=1}^{k}\sum_{j=1}^{k} \frac{\partial}{\partial f_j} D_{ij}(f) \frac{\partial}{\partial f_i} P(f;\ t).\end{aligned} \tag{3.100}$$

Neglecting the system-reservoir interaction energy, $S_B(x_f)$ equals the Boltzmann entropy of the heat reservoir with energy $E_T - E_1(z_1)$ because the "macrostate" x_f includes the information on the precise microstate z_1.

$$S_B(x_f) = \int m\ \delta\left(E_2(z_2) - (E_T - E_1(z_1))\right)\ dz_2, \tag{3.101}$$

where m inserts the right measure constant. The constant temperature of the reservoir implies that

$$\frac{1}{T} = \frac{\partial S_B(x_f)}{\partial(E_T - e_1)} = -\frac{\partial S_B(x_f)}{\partial e_1}. \tag{3.102}$$

We can then write the derivatives of S_B in (3.100).

$$\frac{\partial S_B(x_f)}{\partial f_j} = -\frac{1}{T}\frac{\partial E_1}{\partial f_j}. \tag{3.103}$$

For standard Hamiltonians, these derivatives equal the negative velocities and forces divided by the reservoir temperature, because

$$\frac{\partial E_1}{\partial f_{2i-1}} = \frac{\partial E_1}{\partial q_i} = \frac{p_i}{m_i},\ \frac{\partial E_1}{\partial f_{2i}} = \frac{\partial E_1}{\partial p_i}. \tag{3.104}$$

The expression (3.100) above describes the evolution of the probability distribution associated with the following stochastic differential equation (in the Ito sense) [58]

$$df_i = \left(2\ v_i(f) - \sum_{j=1}^{k} \frac{D_{ij}(f)}{k_B T}\frac{\partial E_1}{\partial f_j}\right)\ dt + \sum_{j=1}^{k} G_{ij}(f)\ dW_j. \tag{3.105}$$

The W_j refer to independent Wiener processes,

$$dW_i\ dW_j = \delta_{ij}\ dt. \tag{3.106}$$

The δ_{ij} represents the Kronecker delta. The G_{ij} coefficients must satisfy

$$\sum_{j=1}^{k} G_{ij} G_{lj} = 2k_B T\ D_{il}, \tag{3.107}$$

or $GG^T = D$. The functions v, D and S_B will obviously depend on the temperature (or the internal energy) of the heat reservoir apart from the macrostate x_f, but we have not indicated it explicitly, as it ultimately depends on the constant E_T, the total energy for the combined system.

Hamilton's equations already tell us the time variation of q given p.

$$\dot{q}_i = \frac{p_i}{m_i}. \tag{3.108}$$

We do not have a corresponding deterministic law for the momenta because they change as a consequence of both the forces between the particles in system 1 and the forces exerted by the reservoir. This fact shows up directly in the v coefficients.

$$\begin{aligned} v_{2i}(f) &= \frac{1}{\Omega(x_f)} \int \delta(q - q') \; \delta(p - p') \; \hat{L} i p_i' \; dz' \\ &= \frac{\partial E_1}{\partial q_i} + \mathrm{Tr} \left[\hat{\rho} \; \frac{\partial H_{int}}{\partial q_i} \right]. \end{aligned} \tag{3.109}$$

I do not think we can progress any further without specifying the details of the interactions among the degrees of freedom and the reservoir. The easiest particular case is surely the Brownian particle, for then E_1 reduces to a free particle Hamiltonian

$$E_1(q, \; p) = E_1(p) = \frac{p^2}{2m}. \tag{3.110}$$

We take for granted that the reservoir does not exert a net force on the particle. Many cancellations take place automatically, and the stochastic differential equation turns into the Langevin equation

$$\begin{aligned} dq &= \frac{p}{m} \; dt, \\ dp &= -\frac{D}{m k_B T} \; p \; dt + 2 k_B T \; \sqrt{D} \; dW, \end{aligned} \tag{3.111}$$

where D could conceivably depend on the state $(q, \; p)$.

Saying that we get a generalised Langevin equation (3.105) for the motion of our system in phase space might not sound very exciting. We already know that Langevin dynamics correctly reproduce the statistical behaviour of a system in equilibrium at a given temperature. True, but pay attention to the word '*equilibrium*'. In principle, (3.105) should also describe the transient relaxation towards equilibrium. We must not overstate our case, though. We based our reasoning on the slow dynamics presupposition. Hence, we cannot use our results to simulate systems when q and p vary over the same time scales as the coordinates and momenta in the reservoir, which is unfortunately the most common case.

3.6 The heat equation

Newton's law of cooling and the one-dimensional heat equation embody approximations to our coarse-grained theory of energy transfer.

Equation (3.96) looks like Newton's law of cooling [78],

$$\frac{dQ}{dt} = \tilde{D}\ (T_2 - T_1), \tag{3.112}$$

where Q represents the energy flowing from system 2 to system 1. Instead of temperature differences, our equation used inverse temperatures. For high absolute temperatures and small temperature differences, the expressions give similar predictions. For example, take two identical systems at different temperatures, as before. The final equilibrium temperature equals $T = (T_1 + T_2)/2$. Define ΔT as $(T_1 - T_2)/2$, so that $T_1 = T + \Delta T$ and $T_2 = T - \Delta T$.

$$\frac{1}{k_B T_1} - \frac{1}{k_B T_2} = \frac{T_2 - T_1}{k_B\ (T + \Delta T)(T - \Delta T)}. \tag{3.113}$$

When $\Delta T \ll T$, then

$$\frac{1}{k_B T_1} - \frac{1}{k_B T_2} \approx \frac{T_2 - T_1}{k_B T^2}, \tag{3.114}$$

and equation (3.96) turns into (3.112) by choosing $\tilde{D} = D/(k_B T^2)$.

We will investigate how to derive an equation for heat conduction from the idea behind (3.96). We begin with a one-dimensional temperature profile for an insulated metal bar. We picture many cross-sectional slices of width dx. According to our derivation above, energy may flow to contiguous slices. We represent the expected energy at point x and time t with $e(x;\ t)$.

$$\begin{aligned}\frac{\partial e(x;\ t)}{\partial t} &= D\left(\frac{2}{T(e(x;\ t))} - \frac{1}{T(e(x+dx;\ t))} - \frac{1}{T(e(x-dx;\ t))}\right) \\ &\propto -\frac{\partial^2}{\partial x^2}\left(\frac{1}{T(e(x;\ t))}\right). \end{aligned} \tag{3.115}$$

This suggests the following equation for heat conduction in one dimension:

$$\frac{\partial e(x;\ t)}{\partial t} + \alpha\ \frac{\partial^2}{\partial x^2}\left(\frac{1}{T(e(x;\ t))}\right) = 0, \tag{3.116}$$

with α depending on the material properties of the bar. The analogue equation in three dimensions reads

$$\frac{\partial}{\partial t} e(r;\ t) + \alpha\ \nabla^2\left(\frac{1}{T(e(r;\ t))}\right) = 0, \tag{3.117}$$

At high temperatures and small temperature differences, we follow the approximation used above for Newton's law of cooling, plus the Dulong-Petit law [79],

$$e(x;\ t) \approx 3N(x)\ k_B T(x;\ T). \tag{3.118}$$

$N(x)$ stands for the number of particles at position x. Inserting both approximations into (3.116), we recover the well-known heat equation,

$$\frac{\partial}{\partial t} T(x;\ t) \propto \frac{\partial^2}{\partial x^2} T(x;\ t). \tag{3.119}$$

It is interesting to compare the solutions of equations (3.116) and (3.119). We use the Dulong-petit law to simplify the former and solve the following expressions numerically for an initial temperature profile shaped like a

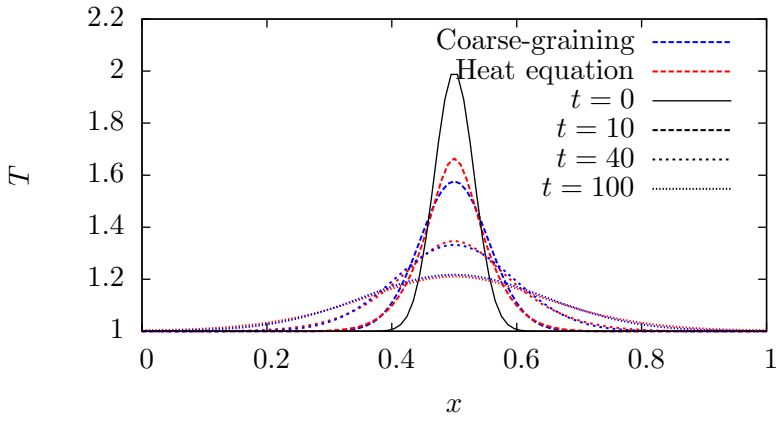

Figure 3.10: Numerical solutions of the heat equation (3.121) (*red*) and the coarse-grained heat conduction (3.120) (*blue*) for three different times ($\alpha_1 = 1$, $\alpha_2 = 1.5$). The solid black line shows the initial temperature profile.

normal distribution.

$$\frac{\partial T}{\partial t} + \alpha_1 \frac{\partial^2 T}{\partial x^2} = 0. \tag{3.120}$$

$$\frac{\partial T}{\partial t} - \alpha_2 \frac{\partial^2}{\partial x^2}\left(\frac{1}{T}\right) = 0. \tag{3.121}$$

Figure 3.10 demonstrates that the solutions look very much alike.

3.7 Fluctuation-dissipation

Fluctuation-dissipation relations connect equilibrium properties with diffusion away from equilibrium, and they follow automatically from the formalism explained in this chapter

The surprising connection between equilibrium fluctuations and nonequilibrium diffusion coefficients proved by Callen and Welton [41] has now become a routine calculation in nonequilibrium statistical mechanics [67]. Given a Fokker-Planck equation for a process (3.100), the steady state solution (if it exists) makes the partial derivative with respect to time vanish. The diffusion term then equals the negative drift term and we can often find a relation between the v and D coefficients.

The aforementioned result holds for equilibrium. The derivation of generalised Fokker-Planck equations from Liouville's equation (2.2) provides this connection automatically. Equations like (3.81) for the diffusion coefficients,

$$D_{ij}(f) = \int_0^t \mathrm{C}_{\hat{\rho}}[\hat{L}\mathrm{i}F_j,\ \hat{L}\mathrm{i}F_i(t-s)]\ ds, \tag{3.122}$$

could be interpreted as fluctuation-dissipation theorems that hold for any state, however far from equilibrium. The Langevin equation (3.111), for example, emerges with a connection between the friction coefficient and the intensity of the noise through D.

3.8 Compressible flow in phase space

We have developed the theory of coarse-graining for isolated Hamiltonian systems, which satisfy Liouville's theorem for incompressible motion of the phase fluid. In some cases, we can extend the theory to equations of motion that generate nonequilibrium steady states, which generate *compressible* flows in phase space.

This chapter began with Liouville's equation for a Hamiltonian system and derived an equation for the macroscopic variables. Interestingly, the main idea extends beyond Hamiltonian systems. Appendix A briefly explains why the projection operator version of our theory applies to compressible flows in phase space.

When we analyse nonequilibrium steady states, such as those generated by contact with reservoirs at different temperatures, or constant tempera-

ture systems in a shear flow, for example, we usually leave the reservoirs out of our description. Liouville's theorem for the incompressible flow of probability in phase space no longer applies then, so we use a more general idea inspired by the motion of compressible fluids. The time variation of probability at a point in phase space must equal the negative divergence of the flow,

$$\frac{\partial \rho}{\partial t} = -\sum_{i=1}^{N} \left(\frac{\partial}{\partial q_i}(\rho\ \dot{q}_i) + \frac{\partial}{\partial p_i}(\rho\ \dot{p}_i) \right). \tag{3.123}$$

This brings us to a generalised version of Liouville's equation [19].

$$\frac{\partial \rho}{\partial t} = -\hat{L}\mathrm{i}\rho - \rho\ \sum_{i=1}^{N} \left(\frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right). \tag{3.124}$$

We can write this equation briefly as

$$\frac{\partial \rho}{\partial t} = -\mathrm{L}\rho, \tag{3.125}$$

if we define the new L operator as

$$\mathrm{L} = \hat{L}\mathrm{i} + \sum_{i=1}^{N} \left(\frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right) = \hat{L}\mathrm{i} + \Lambda. \tag{3.126}$$

The phase space compression factor Λ disappears when the motion obeys Hamilton's equations (1.42),

$$\Lambda(q,\ p) = \sum_{i=1}^{N} \left(\frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right) = \sum_{i=1}^{N} \left(\frac{\partial^2 H}{\partial p_i \partial q_i} - \frac{\partial^2 H}{\partial q_i \partial p_i} \right) = 0. \tag{3.127}$$

L is linear, so it allows us to repeat the steps at the beginning of Section 3.1 for the evolution of ρ. We start with

$$\mathrm{L}\rho = \mathrm{L}\bar{\rho} + \mathrm{L}\delta\rho. \tag{3.128}$$

Then we write the equation for the time evolution of $\delta\rho$ and solve it formally in terms of $\bar{\rho}$.

$$\delta\rho(t) = e^{-\mathrm{L}t}\delta\rho(0) - \int_0^t e^{-\mathrm{L}(t-s)}\left(\mathrm{L}\rho + \frac{\partial\bar{\rho}}{\partial s}\right)\, ds. \tag{3.129}$$

We can follow our derivation further when we have a Hermitian L, that is, when

$$\mathrm{Tr}[A\mathrm{L}B] = -\mathrm{Tr}[(\mathrm{L}A)B]. \tag{3.130}$$

McPhie *et al.* have shown that L becomes Hermitian in the long time limit for a mixing system under the influence of a deterministic thermostat[8] [80]. The resulting equation for the evolution of the relevant phase function then looks just like (3.12), only replacing $\hat{L}\mathrm{i}$ with L.

$$\frac{\partial f_i}{\partial t} = \mathrm{Tr}[\rho\ \mathrm{L}F_i] + \int_0^t \mathrm{Tr}\left[(\mathrm{L}F_i)e^{-\mathrm{L}(t-s)}\mathrm{L}\bar{\rho}\right]\, ds + \int_0^t \mathrm{Tr}\left[(\mathrm{L}F_i)e^{-\mathrm{L}(t-s)}\frac{\partial\bar{\rho}}{\partial s}\right]\, ds. \tag{3.131}$$

L still satisfies a version of the Schrödinger-Heisenberg duality (see Appendix A),

$$\mathrm{Tr}\left[Ae^{-\mathrm{L}t}B\right] = \mathrm{Tr}\left[Be^{\mathrm{L}t}A\right], \tag{3.132}$$

which we can use to rewrite the traces in (3.131). Keeping track of the new Λ factor and remembering (3.22),

$$\mathrm{L}\bar{\rho} = \bar{\rho}\ \left(\sum_{i=1}^{k}\lambda_i\hat{L}\mathrm{i}F_i + \Lambda\right). \tag{3.133}$$

Therefore, the more general version of equation (3.32) contains an extra

[8] See Section 4.2 for an explanation of deterministic thermostats.

term which takes the phase space compression into account.

$$\begin{aligned}\frac{\partial f_i}{\partial t} =&\mathrm{Tr}[\bar{\rho}\ \mathrm{L}F_i] - \sum_{j=1}^{k}\int_0^t \mathrm{C}_{\bar{\rho}}[\hat{L}\mathrm{i}F_j,\ \mathrm{L}F_i(s-t)]\ \lambda_j(f)\ ds \\ &- \sum_{j=1}^{k}\int_0^t \mathrm{C}_{\bar{\rho}}[\Lambda,\ \mathrm{L}F_i(s-t)]\ \lambda_j(f)\ ds \\ &- \sum_{j=1}^{k}\sum_{l=1}^{k}\int_0^t \mathrm{C}_{\bar{\rho}}[\delta F_j,\ \mathrm{L}F_i(s-t)]\frac{\partial\lambda_j}{\partial f_l}\frac{\partial f_l}{\partial s}\ ds. \end{aligned} \tag{3.134}$$

Here, the time arguments indicate evolution by means of the L operator,

$$F(t) = e^{\mathrm{L}t}F(0). \tag{3.135}$$

In Section 3.3, we learnt how to transform an exact equation for the expected values f_i into a generalised Fokker-Planck equation for P_F, by choosing the delta functions Ψ_f as the relevant variables. The first integral in (3.134) transforms into

$$\begin{aligned}\mathrm{Tr}[\bar{\rho}\ \mathrm{L}\Psi_f] &= -\sum_{i=1}^{k}\frac{\partial}{\partial f_i}\mathrm{Tr}[\bar{\rho}\ (\hat{L}\mathrm{i}F_i)\ \Psi_f] + \mathrm{Tr}[\bar{\rho}\Lambda\Psi_f] \\ &= -\sum_{i=1}^{k}\frac{\partial}{\partial f_i}v_i(f)\ P_F(f;\ t) + \mathrm{Tr}[\hat{\rho}\ \Lambda]\ P_F(f;\ t), \end{aligned} \tag{3.136}$$

where v represents the same function given in (3.46), and $\hat{\rho}$ the micro-canonical distribution for the macrostate x_f (Cf. (3.47)). When ρ equals the relevant distribution $\bar{\rho}$ all the other terms cancel. We see then that the probability P_F changes due to a drift (first term) and to the phase space compression (second term). All the subsequent terms add corrections in response to the presence of a non-zero $\delta\rho$. We will encounter no further technical difficulties in their mind-numbing calculation.

First we take the trace involving $\mathrm{L}\bar{\rho}$, and apply (3.133) and the chain rule to transfer the action of the Liouvillian from Ψ to F,

$$\begin{aligned}\mathrm{Tr}\left[(\mathrm{L}\bar{\rho})\ e^{\mathrm{L}(t-s)}\mathrm{L}\Psi_f\right] = &-\sum_{i=1}^{k}\frac{\partial}{\partial f_i}\mathrm{Tr}\left[(\hat{L}\mathrm{i}\bar{\rho})\ e^{\mathrm{L}(t-s)}(\hat{L}\mathrm{i}F_i)\Psi_f\right]\\ &-\sum_{i=1}^{k}\frac{\partial}{\partial f_i}\mathrm{Tr}\left[\bar{\rho}\ \Lambda\ e^{\mathrm{L}(t-s)}(\hat{L}\mathrm{i}F_i)\Psi_f\right]\\ &+\mathrm{Tr}\left[\hat{L}\mathrm{i}\bar{\rho}\ e^{\mathrm{L}(t-s)}\Lambda\ \Psi_f\right]\\ &+\mathrm{Tr}\left[\bar{\rho}\ \Lambda\ e^{\mathrm{L}(t-s)}\Lambda\ \Psi_f\right]. \qquad (3.137)\end{aligned}$$

Now we use (3.59) and rewrite the traces above in terms of time correlation functions, beginning with the traces with $\hat{L}\mathrm{i}\bar{\rho}$,

$$\begin{aligned}&\mathrm{Tr}\left[\hat{L}\mathrm{i}\bar{\rho}\ e^{\mathrm{L}(t-s)}\hat{L}\mathrm{i}F_i\ \Psi_f\right]\\ &\qquad=\sum_{j=1}^{k}\mathrm{C}_{\hat{\rho}}[\hat{L}\mathrm{i}F_j,\ \hat{L}\mathrm{i}F_i(t-s)]\ \Omega(x_f)\ \frac{\partial}{\partial f_j}\left(\frac{P_F(f;\ s)}{\Omega(x_f)}\right), \qquad (3.138)\end{aligned}$$

and

$$\begin{aligned}&\mathrm{Tr}\left[\hat{L}\mathrm{i}\bar{\rho}\ e^{\mathrm{L}(t-s)}\Lambda\ \Psi_f\right]\\ &\qquad=\sum_{j=1}^{k}\mathrm{C}_{\hat{\rho}}[\hat{L}\mathrm{i}F_j,\ \Lambda(t-s)]\ \Omega(x_f)\ \frac{\partial}{\partial f_j}\left(\frac{P_F(f;\ s)}{\Omega(x_f)}\right). \qquad (3.139)\end{aligned}$$

The other two traces become

$$\begin{aligned}\mathrm{Tr}\left[\bar{\rho}\ \Lambda\ e^{\mathrm{L}(t-s)}\Lambda\ \Psi_f\right] &= \mathrm{C}_{\hat{\rho}}[\Lambda,\ \Lambda(t-s)]\ P_F(f;\ s), \qquad (3.140)\\ \mathrm{Tr}\left[\bar{\rho}\ \Lambda\ e^{\mathrm{L}(t-s)}\hat{L}\mathrm{i}F_i\Psi_f\right] &= \mathrm{C}_{\hat{\rho}}[\Lambda,\ \hat{L}\mathrm{i}F_i(t-s)]\ P_F(f;\ s). \qquad (3.141)\end{aligned}$$

The same tricks bring the remaining traces in the exact equation to

$$\begin{aligned}
\mathrm{Tr}\left[\bar{\rho}\; e^{\mathrm{L}(t-s)}\mathrm{L}\Psi_f\right] &= -\sum_{i=1}^{k}\frac{\partial}{\partial f_i}\mathrm{C}_{\hat{\rho}}\left[\Lambda,\ \hat{L}\mathrm{i}F_i(t-s)\right]\ P_F(f;\ s) \\
&\quad +\mathrm{C}_{\hat{\rho}}\left[\Lambda,\ \Lambda(t-s)\right]\ P_F(f;\ s), \\
\mathrm{Tr}\left[\frac{\partial\bar{\rho}}{\partial s}\; e^{\mathrm{L}(t-s)}\mathrm{L}\Psi_f\right] &= -\sum_{i=1}^{k}\frac{\partial}{\partial f_i}\mathrm{Tr}\left[\hat{\rho}e^{\mathrm{L}(t-s)}\hat{L}\mathrm{i}F_i\right]\frac{\partial}{\partial s}P_F(f;\ s). \\
&\quad +\mathrm{Tr}\left[\hat{\rho}e^{\mathrm{L}(t-s)}\Lambda\right]\frac{\partial}{\partial s}P_F(f;\ s). \qquad (3.142)
\end{aligned}$$

Putting all the parts together and introducing the slow variable hypothesis and our dangerous equation (3.67) to slightly simplify the results, we get the following gigantic formula in all its gory details.

$$\begin{aligned}
\frac{\partial}{\partial t}P_F(f;\ t) = &-2\sum_{i=1}^{k}\frac{\partial}{\partial f_i}v_i(f)\ P_F(f;\ t) + 2\ \mathrm{Tr}[\hat{\rho}\ \Lambda]\ P_F(f;\ t) \\
&-\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial}{\partial f_i}D_{ij}(f)\Omega(x_f)\ \frac{\partial}{\partial f_j}\left(\frac{P_F(f;\ s)}{\Omega(x_f)}\right) \\
&-\sum_{i=1}^{k}\int_0^t\frac{\partial}{\partial f_i}\mathrm{C}_{\hat{\rho}}\left[\Lambda,\ \hat{L}\mathrm{i}F_i(t-s)\right]\ ds\ P_F(f;\ t) \\
&+\sum_{j=1}^{k}\int_0^t\mathrm{C}_{\hat{\rho}}[\hat{L}\mathrm{i}F_j,\ \Lambda(t-s)]\ ds\ \Omega(x_f)\ \frac{\partial}{\partial f_j}\left(\frac{P_F(f;\ t)}{\Omega(x_f)}\right) \\
&-\sum_{i=1}^{k}\int_0^t\frac{\partial}{\partial f_i}\mathrm{C}_{\hat{\rho}}\left[\Lambda,\ \hat{L}\mathrm{i}F_i(t-s)\right]\ ds\ P_F(f;\ t) \\
&+2\int_0^t\mathrm{C}_{\hat{\rho}}\left[\Lambda,\ \Lambda(t-s)\right]\ ds\ P_F(f;\ t). \qquad (3.143)
\end{aligned}$$

The functions v and D have their usual meanings (Cf. (3.46) and (3.77)).

We view equation (3.143) as a general framework within which to derive equations for particular problems. I do not see how it could be applied

directly in any circumstances, but it might serve as a starting point for further research, after fearlessly crossing out some terms.

3.9 Summary

This chapter addressed the main theme in the thesis: the theory of coarse-graining without projection operators. Statistical mechanics infers the properties of substances in thermal equilibrium by assigning probabilities to the microstates and using them to calculate the expected values of phase functions. If we wish to extend this approach to nonequilibrium processes, we must determine how probability distributions evolve in time. Zwanzig developed a theory of irreversible change for ensembles. It relied on the mathematical theory of projection operators and rewrote Liouville's equation as a closed expression for the evolution of macroscopic variables.

Here we have derived a similar theory without any reference to projection operators. The main advantage of our point of view over Zwanzig's lies in the absence of any reference to projected dynamics in the equations. Instead, they contain only the real dynamics. This simplifies the analytical calculations and the connection with molecular dynamics averages. The price to pay for these simplifications is a few extra terms in the exact equations.

Through the slow variable hypothesis and algebraic manipulations, we transformed the general expressions into the equations for Brownian motion, thermalisation and heat conduction. From the latter, we inferred the heat equation in one dimension as an approximation to the underlying Hamiltonian dynamics. The nonequilibrium equations yield fluctuation-dissipation relations and Green-Kubo coefficients, just like in Zwanzig's theory.

Towards the end, we explained how to extend our reasoning to systems under the influence of deterministic thermostats. We view both the Hamiltonian and thermostated dynamical equations as a framework within which to find approximate expressions for particular problems.

All the concrete predictions in our examples were built on the assump-

tion of a separation between the molecular time scales and the macroscopic evolution. Even with slow macroscopic variables, the future properties of a system may depend on its history when the irrelevant part of the distribution does not decay quickly. In the following chapter, we will examine two problems in which the internal energy of a system changes slowly when viewed on the molecular time scales, but with a distribution that departs from the relevant ensemble. We will observe a behaviour unlike the monotonic approach to equilibrium which we have described in this chapter.

Chapter 4

Memory Effects

We argue that the logarithmic thermostat recently proposed by Campisi *et al.* is ineffective because its "constant temperature" property applies only to very large time scales. We suggest how to estimate the importance of memory effects.

One day, at tea time at my parent's place some years ago, I remember I was sitting on the living room carpet with my laptop. My mother was on the sofa reading Proust's *À la recherche du temps perdu*. Putting the book down and leaving her reading glasses on the tea table, she turned to me and told me about some episode in the novel. I have forgotten the exact details of the conversation, but we ended up talking about remembering that particular moment years later. As far as I know, I recall this scene not because my mind can reach through spacetime and perceive events gone by, but because I have become a different person from my younger self. The structure of my brain has changed, and now, when I dip a madeleine into my tea, I am reminded of Proust and transported back to the carpet in my parent's living room.

In much the same way as a computer does, we store memory in the present state of our bodies. Our past experiences make up who we are

today in a very real sense. We interact with our environments and change, and the cumulative change over our lives affects how we will react to the next situation.

In physics, we apply the word 'memory' to a wider range of phenomena than just the remembrance of things past. Whenever we observe that the evolution of a system depends on its previous history and not just on its present state, we refer to this fact as a memory effect. As already mentioned in Chapter 1, we assume without proof that, in principle, a detailed enough description of the state of a system will eliminate memory effects completely. Incomplete information propagates unpredictability, which we can sometimes remedy by considering the history of a process. A very simple thought experiment will make the point clear.

Imagine a statistical mechanics exam which asks us to calculate the kinetic energy stored in a copper disk hanging on the laboratory wall. The problem statement provides the density and atomic mass of copper, the dimensions of the disk and the room temperature T. The law of equipartition of energy states that the total kinetic energy equals $3Nk_BT/2$, with N being the number of atoms in the copper. We use the dimensions of the disk to calculate its volume, multiply by the density to get its mass, divide by the atomic mass to find the number of moles and, finally, multiply by Avogadro's number to determine the number of atoms, N.

An annoying know-it-all student might insist that we did not have all the necessary information to solve the problem. What if the disk were part of the pendulum on a grandfather clock? Then we would also need to know the amplitude of the oscillation and its present (or past) position and velocity. What if the disk had been brought out of the freezer? We would have to be told how long it had been hanging on the wall and the temperature in the freezer. And so on.

Interestingly, the history of the copper disk loses its relevance as time goes by, assuming we leave it on the wall. The oscillations will dampen out and the disk will warm to room temperature. Soon we will be unable to tell by examining the disk if this morning it was oscillating or in the refrigerator.

The tendency to "forget" past states is obviously not confined to cop-

per disks. The complex dynamics of many-body systems cause matter to approach states of thermodynamic equilibrium characterised by the values of a handful of parameters, independently of their previous states. When I pour milk into my tea, for instance, it billows and spreads through the water. The next time I have tea, the milk will form into a cloud again, but not the same cloud. The exact details of its shape depend on minute differences that climb up the length scales, amplified by the unstable dynamics, until they become evident at plain sight. Nevertheless, a few minutes later the cup always contains the same calm surface and homogeneous colour, which depends only on the amount of tea and milk.

Oblivion spreads through matter leaving homogeneous and dull landscapes in its stead. We feel its action, however, over very different time scales, depending on the case. Pressure waves through the air in a room dissipate in the blink of an eye, vortices in a stirred liquid take much longer, and erosion sluggishly rounds mountain peaks over geological ages. Memory effects are primarily a question of time scales. If we wait long enough, the properties of any system will depend only on the values of its dynamical invariants and on its temperature.

4.1 Definitions of temperature

The definition of temperature in terms of entropy (1.104) usually becomes difficult to work with in the context of nonequilibrium simulations. Instead of calculating the temperature from the probability densities (through the entropy), several alternative definitions of temperature as a function of the microstate have been proposed.

Maxwell defined the temperature of a body as "its thermal state considered with reference to its power of communicating heat to other bodies" [81], with higher temperature bodies warming lower temperature bodies. On the same page, he introduced what would eventually come to be known as the Zeroth law of thermodynamics:

> Law of Equal Temperatures.—*Bodies whose temperatures are equal to that of the same body have themselves equal temperatures.*

The first chapter defined temperature (1.104) as the inverse variation of entropy with energy, $T = (\partial S/\partial E)^{-1}$, without any mention of heat exchange. Now, macroscopic systems usually transfer heat between them through weak interactions, meaning that their internal energy overwhelms the relatively small interaction energy. We then say that the Boltzmann entropy is an extensive property, or that the entropy of a combined system equals the sum of the entropies of its parts [34]. For a two-part system,

$$S_B(E) = S_{B,1}(E_1) + S_{B,2}(E - E_1). \tag{4.1}$$

The total energy in the combined system, E, equals the sum of the energy in system one, E_1, plus the energy in system two. Our fluctuation theorem (1.30) implies that, when we bring the systems into thermal contact and isolate them, we will likely observe that E_1 changes until the Boltzmann entropy of the combined system reaches a maximum. Differentiating (4.1) with respect to E_1, we see that the state of maximum entropy satisfies

$$\frac{\partial}{\partial E_1} S_{B,1}(E_1) - \frac{\partial}{\partial (E - E_1)} S_{B,2}(E - E_1) = 0, \tag{4.2}$$

which, according to our definition of temperature (1.104), amounts to saying that the temperature of system one equals that of system two. Furthermore, the combined system must also end up with the same temperature, as differentiation of (4.1) with respect to E leads to

$$\frac{1}{T} = \frac{\partial}{\partial E} S_B(E) = \frac{\partial}{\partial (E - E_1)} S_{B,2}(E - E_1). \tag{4.3}$$

Equations (4.2) and (4.3) apply to any number of systems in thermal contact, because we can always single out a particular subsystem and name it system one, and group all the others together as system two. Therefore, as the Zeroth law of thermodynamics states, systems at the same temperature form an equivalence class. Thermometers are simply bodies with some

measurable property that depends on their internal energies, which we use to define a temperature scale. To find the temperature of a system, we let the thermometer equilibrate with it and read off the value on the scale.

Our statistical statement about the evolution of weakly interacting systems, that (most of the time) they evolve towards states with the same temperature, overlooks thermal fluctuations. Instead of denying the existence of states of thermal equilibrium, we interpret the absence of net heat transfer as an average property. This conjures up a problem for nonequilibrium statistical mechanics. If we understand that our system has the same temperature as another (in particular, a thermometer) when they have reached equilibrium and no longer exchange heat on average, then how should we define the temperature of a nonequilibrium state?

One possibility would be to fall back on our definition of temperature (1.104) and work out the entropy of the state and then calculate its derivative. As we saw in Chapter 3, it may already be very difficult to determine the distribution representing the nonequilibrium state, not to mention working out the value of the entropy functional for this distribution. Apart from this practical difficulty, we might prefer a definition of temperature that depends on the system's microstate, instead of the states of an ensemble of systems.

Among the many different ways to define temperature as a phase function, the most popular receives the name of *kinetic* or ideal gas temperature [19]. It emerges from the law of equipartition, which states that at equilibrium each component of the linear momentum, p_α, has an expected value of $\langle p_\alpha \rangle = k_B T/2$. Hence, the total kinetic energy at any instant should be very nearly proportional to $k_B T$. For a d-dimensional system,

$$\sum_{i=1}^{N} \frac{p_i^2}{2m_i} \approx \frac{d}{2}\, N k_B T. \tag{4.4}$$

The expression above turns into an exact equality for ideal gases. To define the kinetic temperature, we simply solve the equation above for T and take

for granted that it applies also to nonequilibrium states.

$$T = \frac{1}{dNk_B} \sum_{i=1}^{N} \frac{p_i^2}{m_i}. \tag{4.5}$$

In particular applications, such as the study of shockwaves, when a strong anisotropy develops in the kinetic energy, we might prefer to define the temperature as a tensor. Along the x direction,

$$T_{xx} = \frac{1}{Nk_B} \sum_{i=1}^{N} \frac{p_{x,i}^2}{m_i}, \tag{4.6}$$

and we define T_{yy} and T_{zz} in a similar fashion. For an equilibrated system, $\langle T_{xx} \rangle = \langle T_{yy} \rangle = \langle T_{zz} \rangle$.

Rugh proposed a different dynamical notion, based on the geometry of the energy surfaces in phase space [82],

$$\frac{1}{k_B T} = \left\langle \nabla \cdot \left(\frac{\nabla H}{\|H\|^2} \right) \right\rangle. \tag{4.7}$$

As Rugh noted, (4.5) and (4.7) depend directly on the dynamics and not on the entropy, which makes them independent of any particular choice of relevant variables.

Rugh's work inspired several alternative definitions of temperature [83], the most general of which reads [84]

$$k_B T = \frac{\langle \nabla H \cdot B \rangle}{\langle \nabla \cdot B \rangle}. \tag{4.8}$$

B stands for any continuous and differentiable vector in phase space. Choosing $B = (0,\ 0,\ \ldots,\ 0,\ p_1,\ p_2,\ \ldots,\ p_N)$ outputs the equipartition theorem. A different choice of B yields an alternative definition. If we select B along the gradient of the potential, $B = \nabla V$, the instantaneous temperature is sometimes referred to as the *configurational* temperature [85].

$$k_B T = \frac{\|\nabla V\|^2}{\nabla^2 V}. \tag{4.9}$$

At equilibrium, the time averages of (4.5) and (4.9) both coincide with the thermodynamic temperature (1.104), but they differ as we move away into the nonequilibrium realm. Which definition should we consider correct?

If pressed to give my opinion, I would defend Maxwell's criterion and say that when a body tends to give away part of its energy through thermal contact with another, it has a higher temperature. This lends more weight to the entropic definition. In practice, though, not much work gets done with this view because of the difficulty in determining the nonequilibrium entropy.

An interesting particular case arises when we consider a one-dimensional *monocyclic motion*. A particle moves within a confining potential $V(q,\ \lambda)$ (where λ stands for an external control parameter) for which there exists only one closed trajectory for each value of the total energy E and the parameter λ. If the Hamiltonian function has the form

$$H(q,\ p) = \frac{p^2}{2m} + V(q,\ \lambda), \tag{4.10}$$

then we can always write the solution of Hamilton's equations of motion as the integral [86]

$$t = \sqrt{\frac{m}{2}} \int \frac{dq}{\sqrt{E - V(q,\ \lambda)}}. \tag{4.11}$$

We determine the period of oscillation by finding the turning points, that is, the values of q that satisfy

$$V(q,\ \lambda) = E. \tag{4.12}$$

Call these points q_A and q_B. The period equals twice the time taken in going from q_A to q_B, or

$$2t_{AB} = \sqrt{2m} \int_{q_A}^{q_B} \frac{dq}{\sqrt{E - V(q,\ \lambda)}}. \tag{4.13}$$

Now, when we compare the period (4.13) to the density of states for E and λ,

$$\Omega(x_{E,\lambda}) = \int \delta(H(q,\ p) - E)\ dq\ dp, \tag{4.14}$$

we realise, integrating out the p dependence, that $\Omega(x_{E,\lambda}) = 2t_{AB}$, so the Boltzmann entropy for $x_{E,\lambda}$ equals

$$S_B(x_{E,\lambda}) = k_B \ \ln(2t_{AB}). \tag{4.15}$$

Campisi and Hänggi recently proposed [87] to approach monocyclic systems with the Helmholtz theorem [88], defining the temperature as

$$T = \frac{1}{k_B}\frac{\Phi(x_{E,\lambda})}{\Omega(x_E,\lambda)}, \tag{4.16}$$

where Φ represents the phase space volume for $H(z) \leq E$,

$$\Phi(x_{E,\lambda}) = \int_0^E \Omega(x_{e,\lambda})\ de. \tag{4.17}$$

Helmholtz found a way to express how the density of states changed as a consequence of variations in the system energy and parameter λ with the same structure as the classic heat theorem,

$$dS = \frac{dE}{T} - \frac{d\mathcal{F}}{T}. \tag{4.18}$$

S represented the thermodynamic entropy, E the internal energy, and $\mathcal{F}$ the free energy. In Chapter 2, we saw how to define the free energy in terms of the partition function and derived a more general relation (2.9), which gave a statistical mechanical interpretation of the heat theorem (4.18). The monocyclic analogue of the free energy,

$$\mathcal{F} = \left\langle \frac{\partial V}{\partial \lambda} \right\rangle, \tag{4.19}$$

tells us the average variation of the potential energy with λ.

For the heat theorem to work, we need a different notion of entropy, known as the *Hertz entropy* [89, 87], S_H,

$$S_H(x_{E,\lambda}) = k_B \ \ln(\Phi(x_{E,\lambda})). \tag{4.20}$$

Now, when we calculate the derivatives of the Hertz entropy,

$$\frac{\partial S_H(x_{E,\lambda})}{\partial E} = k_B \frac{\Omega(x_{E,\lambda})}{\Phi(x_{E,\lambda})} = \frac{1}{T}, \tag{4.21}$$

$$\frac{\partial S_H(x_{E,\lambda})}{\partial \lambda} = \frac{k_B}{\Phi(x_{E,\lambda})} \left\langle \frac{\partial V}{\partial \lambda} \right\rangle, \tag{4.22}$$

we find that the inverse of T (4.16) becomes the integrating factor,

$$dS_H = \frac{dE}{T} - \frac{\mathcal{F}}{T}. \tag{4.23}$$

Helmholtz's theorem (4.23) provided inspiration for Boltzmann's statistical mechanical investigation of the foundations of thermodynamics [89, 90]. It derived a paradigmatic thermodynamic expression from mechanical postulates. A key point to remember here, as it will come up again when we analyse the logarithmic oscillator in Section 4.3, is the fact that the temperature definition (4.16) involves averaging over an entire period, $2t_{AB}$. For time scales smaller than this interval, the heat theorem (4.23) no longer holds.

4.2 Deterministic thermostats

We can incorporate constraints into Hamilton's equations of motion which make the system sample states at constant temperature. The algorithms depend not only on the chosen definition of temperature, but also on how we implement the constraint.

Differing notions of temperature will obviously rule alternative thermostated dynamics. Even when we agree on the definition and pick, for example, the kinetic temperature (4.5), we may still choose from alternative implementations of the constant-temperature laws. Within the field of deterministic dynamics, Hoover and coworkers derived the isokinetic equations of motion

from Gauss's principle of least constraint[1] [91, 92, 93]. By constraining the kinetic temperature to equal

$$\sum_{i=1}^{N} \frac{p_i^2}{2m_i} = \frac{d}{2}\, N k_B T, \tag{4.24}$$

an extra frictional force appears in Newton's second law, which becomes

$$m\ddot{q}_i = -\frac{\partial H}{\partial q_i} - \zeta m\dot{q}_i. \tag{4.25}$$

The friction coefficient ζ depends on the microstate.

$$\zeta = \frac{-\frac{\partial H}{\partial q} \cdot q}{m\dot{q} \cdot \dot{q}}. \tag{4.26}$$

Although isokinetic dynamics (4.25) do not satisfy Liouville's theorem, they retain the properties of memorylessness and reversibility (with the friction coefficient changing sign in the reverse trajectory)[2]. Most importantly, while they keep the kinetic temperature fixed at a constant value, they nevertheless recreate the canonical distribution in configuration space [96].

In 1983, Shūichi Nosé came up with an extended system method that sampled the complete canonical distribution in phase space [97]. How he discovered this approach remains a mystery. William G. Hoover, who had the chance to speak to Nosé and some of his collaborators on the matter, has often remarked that he was unable to get a clear impression of what led Nosé to his breakthrough. Here I present my own reconstruction. Perhaps Nosé followed a different path, but the derivation below flows quite naturally

[1] See [18], Chapter IV, Section 8, for Gauss's physical analogy to the method of least squares.

[2] Dettmann and Morriss found a Hamiltonian from which isokinetic dynamics followed for a particular choice of the total energy [94]. Hoover and Leete discovered an alternative Hamiltonian which could be constrained to give isokinetic dynamics different from (4.25) [95].

from Hamiltonian dynamics, and it is consistent with Nosé's interpretation in terms of virtual variables and time rescaling.

In Chapter 1, Section 1.4, we saw how to incorporate fluctuations in the system energy into Hamiltonian mechanics by means of a reparametrisation that included the time among the coordinates (1.50). The Hamiltonian vanished in the resulting variational problem, which was solved by imposing the constraint $K = 0$, where $K = H(q,\ p) + p_t$ (and p_t stood for the momentum conjugate to the time coordinate t). We then interpreted the momentum p_t as the negative of the system Hamiltonian.

When in contact with a heat bath, the energy leaving the system flows out into the bath so, assuming the system and bath isolated from the rest of the universe, the sum of their Hamiltonian functions equals a constant energy E.

$$H(q,\ p) + H_{bath}(z_{bath}) = E. \tag{4.27}$$

We can read this last equation as the constraint on K if we identify p_t with $H_{bath}(z_{bath}) - E$. Therefore, the isothermal equations of motion follow from (1.51) by choosing $K = H(q,\ p) + H_{bath}(z_{bath}) - E$, though we cannot go any further until we specify how the bath Hamiltonian changes with the free parameter.

Nosé must have noticed that t' scaled the velocities in the reparametrised equations, $\dot{q} = q'/t'$, (see the discussion leading up to equation (1.49) in Chapter 1), as velocity rescaling enjoyed great popularity at the time in the field of molecular dynamics. Following his convention, we write t' as $1/s$ and rename the independent parameter to τ. The kinetic energy in the new variables,

$$T_k(q) = \frac{p^2}{2ms^2}, \tag{4.28}$$

included the original momenta as p/s, with the real time satisfying

$$t = \int_0^\tau \frac{d\tau}{s}. \tag{4.29}$$

The next question to answer was how the change in kinetic energy affected the heat bath.

For most systems with many degrees of freedom, the density of states grows exponentially with increasing energy (spin gases are one of the rare exceptions [98]). Nosé decided to represent the heat bath with a single degree of freedom by situating a particle with mass Q in a logarithmic potential. The Hamiltonian function for this *logarithmic oscillator* (as Campisi and coworkers have proposed to call it [99]) reads

$$H_{log}(s,\ p_s) = \frac{p_s^2}{2Q} + \lambda\ \ln(|s|), \tag{4.30}$$

with λ representing an arbitrary external parameter. Being a monocyclic system, the density of states equals the period $2t_{AB}$ (4.13), as we saw above.

$$\Omega(x_{E,\lambda}) = \sqrt{2m}\int_{s_A}^{s_B} \frac{ds}{\sqrt{E - \lambda\ \ln(|s|)}}. \tag{4.31}$$

The turning points coincide with the states with vanishing potential energy, $\lambda\ \ln(|s|) = E$, so

$$s_A = e^{\frac{E}{\lambda}};\ s_B = -e^{\frac{E}{\lambda}}. \tag{4.32}$$

We ignore, for the time being, the steep singularity at $s = 0$ and integrate (4.31) to confirm that the logarithmic oscillator density of states depends exponentially on its energy.

$$\Omega(x_{E,\lambda}) = \sqrt{\frac{8\pi m}{\lambda}} e^{\frac{E}{\lambda}}. \tag{4.33}$$

The behaviour of the one-particle bath should be completely characterised by the parameter λ. This makes us suspect that λ must be some function of the bath temperature.

Inserting H_{log} as the bath Hamiltonian in K, we get

$$K = \sum_{i=1}^{N} \frac{p_i^2}{2m_i s^2} + V(q) + \frac{p_s^2}{2Q} + \lambda\ \ln(|s|) - E. \tag{4.34}$$

The time coordinate t no longer appears explicitly in the equation for K. However, we are allowed to express the constraint $K = 0$ in terms of any

set of conjugate variables of our choice [18]. Therefore, we express K as a function of q, p, s and p_s, and the variational problem (1.50) generates the following equations of motion (see (1.51) and assume that $s > 0$):

$$\left.\begin{aligned} \frac{dq_i}{d\tau} &= \frac{p_i}{m_i s^2}, \\ \frac{dp_i}{d\tau} &= -\frac{\partial H}{\partial q_i}, \\ \frac{ds}{d\tau} &= \frac{p_s}{Q}, \\ \frac{dp_s}{d\tau} &= \frac{1}{s}\left(\sum_{i=1}^{N} \frac{p_i^2}{m_i s^2} - \lambda\right). \end{aligned}\right\} \tag{4.35}$$

The last equation for the variation of p_s is particularly interesting. There we see that, when $p_s = 0$ and

$$\lambda = \sum_{i=1}^{N} \frac{p_i^2}{m_i s^2}, \tag{4.36}$$

the energy flux between the system and the bath stops, as when the system reaches equilibrium with the heat reservoir. Remember that the original momenta were expressed as p/s, so the right hand side equals the real kinetic temperature (4.5) divided by dNk_B. Consequently, we set

$$\lambda = dNk_BT. \tag{4.37}$$

With an ergodic extended system, Nosé dynamics (4.35) sample the microcanonical ensemble (1.102) for the total energy E, but we expect the real coordinates and momenta q and p/s to sample the canonical ensemble (1.114). By calculating the partition function for the original variables, Nosé demonstrated this fact rigorously [97].

With the reparametrisation of time, Nosé transformed the dynamics of a system interacting with its environment into an autonomous Hamiltonian evolution. A conversion back to real time,

$$t \leftarrow \int \frac{d\tau}{s};\ q_i \leftarrow q_i;\ p_i \leftarrow \frac{p_i}{s};\ s \leftarrow s;\ p_s \leftarrow \frac{p_s}{s}; \tag{4.38}$$

transforms the equations of motion (4.35) into

$$\left.\begin{aligned} \dot{q}_i &= \frac{p_i}{m_i}, \\ \dot{p}_i &= -\frac{\partial H}{\partial q_i} - \frac{p_s}{Q} p_i, \\ \dot{s} &= \frac{s^2 p_s}{Q}, \\ \dot{p_s} &= \frac{1}{s}\left(\sum_{i=1}^{N} \frac{p_i^2}{m_i s^2} - dNk_BT\right) - \frac{sp_s}{Q}. \end{aligned}\right\} \tag{4.39}$$

We have lost the canonical structure (1.42), although the function K, which no longer satisfies the properties of a Hamiltonian function, is still conserved by the motion when expressed in these variables. The equation for the system accelerations follows from calculating the time derivative of $\dot{q}$.

$$\ddot{q}_i = \frac{\dot{p}_i}{m_i} = -\frac{1}{m_i}\frac{\partial H}{\partial q_i} - \frac{p_s}{Q}\dot{q}_i. \tag{4.40}$$

The first term on the right we interpret as the monogenic forces [18], while the second represents a friction. If we define the friction coefficient as $\zeta = p_s/Q$, the expression above looks just like the isokinetic equations (4.25), but now the friction changes according to the pair of equations

$$\left.\begin{aligned} \dot{\zeta} &= \frac{1}{Qs}\left(\sum_{i=1}^{N} \frac{m_i \dot{q}_i^2}{s^2} - dNk_BT\right) - \frac{s\zeta}{Q}, \\ \dot{s} &= s^2\zeta. \end{aligned}\right\} \tag{4.41}$$

The motion (4.40) samples the canonical ensemble with reversible trajectories. In contrast to (4.25), the present value of the friction depends on the history of the process, which means that a memory effect has crept into our equations.

The year following the publication of Nosé's dynamics brought an unexpected turn of events. In a single stroke, Hoover simplified the real-time

dynamics (4.40) and introduced the popular equations of motion known today as Nosé-Hoover dynamics [100]. Starting with the equations for the virtual variables (4.35), he scaled back to real time in a noncanonical way, replacing $d\tau$ with $s\ dt$. As we did above, he eliminated the s variable from the equations for p and q, resulting in the now famous pair of equations:

$$\left.\begin{aligned} \ddot{q}_i &= -\frac{1}{m_i}\frac{\partial V(q)}{\partial q_i} - \zeta\dot{q}. \\ \dot{\zeta} &= \frac{1}{Q}\left(\sum_{i=1}^{N} m_i\dot{q}_i^2 - dNk_BT\right). \end{aligned}\right\} \tag{4.42}$$

Hoover proved that the equations conserve the probability distribution

$$\rho(q,\ p,\ \zeta) = \frac{e^{-\frac{1}{k_BT}(H(q,\ p)+\frac{1}{2}Q\zeta^2)}}{\int e^{-\frac{1}{k_BT}(H(q,\ p)+\frac{1}{2}Q\zeta^2)}dq\ dp\ d\zeta}, \tag{4.43}$$

which becomes the canonical ensemble after integration over ζ. Writing out the continuity equation for the probability (3.124),

$$\frac{\partial\rho}{\partial t} + \sum_{i=1}^{N}\frac{\partial}{\partial q_i}(\rho\dot{q}_i) + \sum_{i=1}^{N}\frac{\partial}{\partial p_i}(\rho\dot{p}_i) + \frac{\partial}{\partial\zeta}(\rho\zeta) = 0, \tag{4.44}$$

relatively simple algebraic calculations show that the partial derivatives on the left all add up to zero [100]. Conversely, if we assume that the equation for $\ddot{q}$ in (4.42) conserves $\rho(q,\ p,\ \zeta)$ (4.43) and that $\dot{\zeta}$ does not depend on ζ, then the friction coefficient must satisfy the second equation in (4.42) [100].

Nosé-Hoover dynamics possess many interesting properties. In particular, they promote heat flow when the system interacts with more than one heat reservoir [101, 102] and they satisfy the Jarzynski equality (2.25) [49].

Whoever felt uneasy with the noncanonical time scaling before (4.42) need not worry. Dettmann and Morriss showed how to derive these equations without it in 1996 [103]. They proposed to replace our $K = 0$ constraint above with the equivalent $K' = sK = 0$. Solving the variational

problem with (1.51) generates the Nosé-Hoover equations directly. If we interpret $K' + sE$ as a Hamiltonian function (as Dettmann and Morriss, and later Hoover, did), then the Nosé-Hoover equations (4.42) govern the motion of our system. However, if we insist on regarding $K' = 0$ as a constraint, then we can convert (4.42) back to real time. The resulting equations look like (4.40) with some differences in the scaling factor locations, but the same partition function proof used by Nosé works in this case too, and the real-time variables sample the canonical distribution (assuming ergodicity). Unfortunately, the equations involve cumbersome handling in comparison with Hoover's simple scheme.

The research by Nosé and Hoover led to the proposal of many other equations of motion for canonical dynamics, such as the Hoover-Holian dynamics introduced in Section 2.7 or Nosé-Hoover chains [104]. The alternative methods generally implement an integral control of some quantity. For example, if we wish to monitor a higher velocity moment, like $\dot{q}^3$, in our equations, then the dynamical equations

$$\left.\begin{aligned} m_i\ddot{q}_i &= -\frac{\partial V}{\partial q_i} - \frac{m_i^2}{k_BT}\zeta\dot{q}^3, \\ \dot{\zeta} &= \frac{k_BT}{Q}\sum_{i=1}^{N}\left(\frac{m_i^2\dot{q}_i^4}{(k_BT)^2} - 3\frac{m_i\dot{q}_i^2}{k_BT}\right), \end{aligned}\right\} \tag{4.45}$$

also conserve the canonical distribution in phase space. As we mentioned in Chapter 2, the Hoover-Holian equations control both the kinetic temperature and its fluctuation.

Just this year, Patra and Bhattacharya have proposed a new thermostat that controls the kinetic and configurational temperatures at the same time [105], so I expect there is room for many more discoveries in this area for years to come.

4.3 The logarithmic thermostat affair

The logarithmic oscillator induces the canonical distribution in other systems that come into weak contact with it (assuming ergodicity of the combined system). However, because the exchange of energy takes place over such vast time scales, the logarithmic oscillator does not perform efficiently as a thermostat, whether in computer simulations or experiments.

When I began to browse the Internet looking up current research on computational thermostats, a recent article in the arXiv by a team working in Augsburg caught my eye. Campisi, Zhan, Talkner and Hänggi had entitled their contribution *Logarithmic Oscillators: Ideal Hamiltonian Thermostats* [99]. Logarithmic potentials, they explained, exist in real systems. For instance, a uniformly charged wire generates a logarithmic electric potential in its vicinity [106]. What if we could use logarithmic oscillators to interact with a another system? Campisi *et al.* believed that such a setup might serve as a thermostat, especially for small atomic clusters at low temperatures.

In this approach, we read $K + E$ as if it were a Hamiltonian function, $H(q,\ p,\ s,\ p_s)$, which we rewrite as

$$H(q,\ p,\ s,\ p_s) = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + V(q_i) + \frac{p_s^2}{2Q} + \lambda \ln(|s|) + \left(\frac{1}{s^2} - 1\right) \sum_{i=1}^{N} \frac{p_i^2}{2m_i}. \quad (4.46)$$

This brings the kinetic energy back to its standard $p^2/(2m)$ form and adds an interaction term H_{int} at the end,

$$H_{int}(p,\ s) = \left(\frac{1}{s^2} - 1\right) \sum_{i=1}^{N} \frac{p_i^2}{2m_i}. \quad (4.47)$$

Campisi left the exact form of the interaction term unspecified, but he assumed it was small compared to the other terms. In the expression above, this corresponds to $s \approx 1$.

The density of states for the isolated logarithmic oscillator has an exponential dependence on the energy (4.33), and the curious property that the derivative of the Hertz entropy with respect to energy predicts a temperature equal to $k_B\lambda$ (for the corresponding definition of temperature (4.16)),

$$\frac{1}{T} = \frac{\partial S_H}{\partial E} = \frac{k_B}{\lambda}, \tag{4.48}$$

which depends only on the external parameter and not on the energy. Now, λ comes from the shape of the logarithmic potential, so Campisi *et al.* concluded that they could use the potential to set the temperature of the oscillator to some desired value, $\lambda = k_B T$, and the oscillator, in turn, could interact with some other system and bring it to the same temperature when both systems reached thermal equilibrium. In other words, a single particle played the role of an ideal Hamiltonian heat bath: it had an infinite heat capacity!

Campisi, Zhan and Hänggi probably got the idea for their discovery from their investigation of negative heat capacities [107]. Before 1968 [108], physicists working on statistical mechanics assumed that all systems have positive specific heats, as the canonical ensemble (1.114) characterises equilibrium at a certain temperature T, and we can calculate the heat capacity $\partial E/\partial T$ with [31]

$$\frac{\partial E}{\partial T} = -\frac{\partial}{\partial T}\frac{\partial}{\partial (k_B T)^{-1}}\ \ln(Z) = \frac{1}{(k_B T)^2}\left(\left\langle H^2\right\rangle - \left\langle H\right\rangle^2\right) > 0, \tag{4.49}$$

which is clearly a positive quantity. Nonetheless, astronomers had long known that adding energy to a star would have the effect of cooling it down [109], contradicting our conclusion (4.49). Where could we have gone wrong in such a simple proof? As Thirring pointed out [110], in using the canonical ensemble we have already assumed that the system equilibrates to the temperature of the reservoir, but a system with a negative heat capacity would never equilibrate with an external bath. If the bath were hotter, then

the system would absorb heat from the bath and cool down, leading it to absorb more heat. If the bath were cooler, then the system would give some of its energy away and heat up. Whatever the relative temperatures, contact with a constant temperature environment would trigger a runaway process.

Campisi *et al.* investigated the behaviour of single-particle systems with negative specific heats in their interaction with a neutral ballistic particle confined in the same volume by reflecting walls [107]. Imitating the arguments given by astrophysicists [110], let us examine the virial theorem for a particle in a central confining potential $V(r) = \lambda r^n$, with $r^2 = \sum_i q_i^2$, and n a non-zero integer. The Hamiltonian equals

$$H(q,p) = \sum_{i=1}^{3} \frac{p_i^2}{2m} + \lambda r^n = E. \tag{4.50}$$

If the potential bounds the motion of our particle, then λ and n must share the same sign, but they may be positive or negative. We define the virial as $G = p \cdot r$, and calculate its time derivative using Hamilton's equations (1.42),

$$\frac{\partial G}{\partial t} = p \cdot \dot{r} + \dot{p} \cdot r = \frac{p^2}{m} - n\lambda r^n. \tag{4.51}$$

The distance r covers the range from 0 to some maximum value r_{max} determined by the energy,

$$r_{max} = \sqrt[n]{\frac{E}{\lambda}}. \tag{4.52}$$

When the particle reaches the points where $r = r_{max}$, the velocity vanishes and $G = 0$. When it approaches the origin, then $\lim_{r\to 0} G$ does not go to infinity, as long as $n \geq -2$ (for other cases, we would have to assume that the particle cannot come arbitrarily close to the origin). In addition, the momentum must lie in the range $[0, \sqrt{2mE}]$. Being continuous (except at the origin) the virial is bounded. This implies the vanishing of the time average of (4.51),

$$\left\langle \frac{\partial G}{\partial t} \right\rangle_t = \lim_{t\to\infty} \frac{1}{t} \int_0^t \frac{\partial G}{\partial \tau}\, d\tau = \lim_{t\to\infty} \frac{G(t) - G(0)}{t} = 0. \tag{4.53}$$

Combining (4.53) with (4.51), we realise that

$$\left\langle \frac{p^2}{m} \right\rangle_t = n\lambda \left\langle r^n \right\rangle_t . \tag{4.54}$$

On the left, we have the average kinetic temperature, $\langle T \rangle$ (4.5), and, on the right, an average that depends on the energy. Positive values of n make the expected value of r^n grow with increasing energy, but a negative n makes it decrease. In other words, the particle cools down when we add energy to the system if the potential has a negative n value[3].

Campisi *et al.* must have wondered about the limiting case that separates the systems with positive heat capacities from the negative ones. If we repeat the virial argument for a $V(r) = \lambda \ \ln(r)$ potential, we end up with

$$\left\langle \frac{p^2}{m} \right\rangle_t = \lambda, \tag{4.55}$$

which does not depend on the energy E. Here, setting $\lambda = dk_B T$ for some desired temperature T fixes the average kinetic temperature of the d-dimensional oscillator. This result coincides with our previous conclusion when $d = 1$.

What happens when the logarithmic oscillator interacts weakly with a system of interest with Hamiltonian $H_I(q,\ p)$ and a positive heat capacity? Suppose the combination of both systems samples the microcanonical ensemble, with

$$H(q,\ p,\ r_s,\ p_s) = H_I(q,\ p) + H_{log}(r_s,\ p_s) + H_{int} = E. \tag{4.56}$$

We use s_i to represent the coordinates of the logarithmic particle, and p_{s_i} for the conjugate momenta. We define the distance r_s to the origin with $r_s^2 = \sum_i s_i^2$. We further assume that we can neglect the interaction energy when we compare it with the energies of the logarithmic oscillator and

[3]Thirring noted that, as an explanation for the negative heat capacity of stars, the argument is, at best, incomplete [110].

system of interest, $H_{int} \approx 0$. We wish to derive the probability distribution for $(q,\ p)$.

$$\rho(q,\ p) = \frac{\Omega(x_{E-H_I(q,p),\lambda})}{\Omega(y_E)}. \tag{4.57}$$

The numerator represents the density of states for the logarithmic oscillator when $H_{log} = E - H_I(q,\ p)$.

$$\Omega(x_{E-H_I(q,p),\lambda}) = \sqrt{\frac{\pi}{2Qd\lambda}}\ e^{\frac{d}{\lambda}(E-H_I(q,p))}. \tag{4.58}$$

The denominator contained the density of states for the combined system when the total energy equals E,

$$\begin{aligned} \Omega(y_E) &= \int \delta(H(q,\ p,\ r_s,\ p_s) - E)\ r_s^d\ dq\ dp\ dr_s\ dp_s. \\ &= \int \Omega(x_{E-H_I(q,p),\lambda})\ dq\ dp. \end{aligned} \tag{4.59}$$

The exponential of $(d/\lambda)\ E$ factors out of the numerator and denominator, converting (4.57) into the canonical ensemble.

We have shown that *if* the combined system samples the microcanonical distribution, *then* the system of interest has a canonical distribution (in the available energy range $[0, E]$). As in our erroneous proof of (4.49), we have implicitly assumed that our system will equilibrate with the logarithmic oscillator, but the latter has no intrinsic tendency towards equilibrium. As we argued before, when systems with positive heat capacities exchange heat with another system, their temperature approaches the temperature of the latter. In contrast, negative heat capacities foster the opposite behaviour. The logarithmic oscillator, with its infinite heat capacity, displays no tendency either way.

Imagine two logarithmic oscillators with different values of the λ parameter interacting weakly. For concreteness, we pick an attractive gravitational force between the particles. Furthermore, we situate the singularity of both logarithmic potentials at the same point (for convenience, as we can

rewrite the argument without this assumption). Our experience with positive heat capacities suggests that energy will flow from the particle with the higher average kinetic temperature (higher λ) to the other particle. Because they do not change their temperature in the process, the heat flow will go on for ever. The high temperature particle will loose energy and spiral in towards the singularity, while the other will move further out indefinitely. But then the hotter particle will attract the other inwards, and the cold particle will pull the inner particle outwards, making the energy flow the other way.

Real heat reservoirs have positive heat capacities. When we put them in contact with a hotter body, they absorb heat from it and their temperature increases, albeit negligibly. Without this statistical tendency, we should expect logarithmic thermostats to react incredibly slowly to temperature differences.

When I joined the discussion, Hoover had already written an arXiv contribution [111] in response to the claim by Campisi *et al.* that the logarithmic thermostat was Hamiltonian, in contrast to the Nosé-Hoover equations, which were not. Hoover wanted to emphasise that we can use Hamiltonian mechanics to derive equations that trace out *exactly the same trajectories* in phase space as those generated by Nosé-Hoover dynamics, and he provided several examples illustrating the "Hamiltonian nature" of Nosé-Hoover dynamics.

In response, Campisi *et al.* explained their point in more detail [112]: the Nosé-Hoover equations do not satisfy Liouville's theorem for incompressible flow in phase space. True, but this fact had already been pointed out by Hoover in his original 1985 paper [100]. In fact, Hoover has repeatedly insisted on the virtues of the non-Hamiltonian properties, which make it possible to produce the multifractal distributions associated with deterministic dissipative systems[4]. In the previous section, we found that the reparametrisation of our action integral (1.50) led quite naturally to the Nosé-Hoover equations. They do not satisfy Liouville's theorem, though,

[4]See, for example, [19, 101, 113, 114].

due to the constraint on the energies[5].

I was struck by Campisi's enthusiasm for the experimental realisation of the logarithmic thermostat. As I stressed at the end of Section (4.1), the definition of temperature applies to time scales larger than the period of oscillation, which the energy magnifies exponentially. For the one-dimensional trajectory, the turning points and the period depend exponentially on the energy, $s_{max} \propto \pm e^{E/\lambda}$, $2t_{AB} \propto e^{Ed/\lambda}$, as we have already shown, (4.13) and (4.33). If the particle moves in a circular orbit, we can solve the acceleration equation,

$$\frac{\dot{r}_s^2}{r_s} = \frac{\lambda}{m r_s}, \tag{4.60}$$

for the velocity $\dot{r}_s$, and obtain a speed that does not depend on the radius of the orbit,

$$\dot{r}_s = \sqrt{\frac{\lambda}{m}}. \tag{4.61}$$

The energy determines the radius. Inserting the expression for the speed into the Hamiltonian (4.30) and solving for r_s,

$$r_{s,orb} = e^{\frac{E}{\lambda} - \frac{1}{2}}. \tag{4.62}$$

Then the orbital time also depends exponentially on E,

$$t_{orb} = \frac{2\pi r}{\dot{r}} = 2\pi \sqrt{\frac{m}{\lambda e}}\; e^{\frac{E}{\lambda}}. \tag{4.63}$$

Given arbitrary initial conditions, the oscillator will follow a two-dimensional trajectory that will not usually be closed [86], and it will have turning points between $r_{s,orb}$ and s_{max}. The time between two maximum distances from

[5] In a more recent article, Campisi and Hänggi seem to have forgotten all about this discussion, because they say that their method differs from the Nosé-Hoover equations "in the sense that the thermostated dynamics are obtained directly from Hamilton's equations of motion, with no need to perform time scaling nor the use of non-canonical transformations". As we saw above, Nosé-Hoover equations can also be obtained directly from Dettmann's Hamiltonian by choosing the right initial conditions.

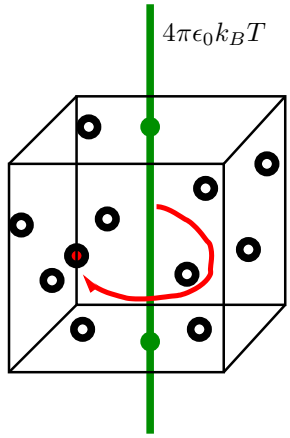

Figure 4.1: The experiment suggested by Campisi *et al.* [99]. Neutral atoms in a box interact with an ion (in red) that feels the electric field created by a charged wire with a linear charge density equal to $4\pi\epsilon_0 k_B T$.

the origin will lie somewhere between $2t_{AB}$ and t_{orb} (note that they share the same order of magnitude)

$$\frac{2t_{AB}}{t_{orb}} = \sqrt{\frac{2de}{\pi}}. \tag{4.64}$$

Campisi *et al.* proposed an experiment with neutral atoms contained in a cubic box with sides of length L (see Figure 4.1). A charged wire passed through the centre of the box generating a two-dimensional electric field felt by an ion, which played the role of the logarithmic oscillator. Picture a dilute gas of 25 argon atoms initially at a temperature $T_0 = 7/3$ K that we wish to bring down to $T = 1$ K. Accordingly, we set the control parameter to $\lambda = 2k_B T$ (4.55). If the thermostat works, then the logarithmic particle must absorb about

$$\Delta E = \frac{3}{2} N k_B \, (T_0 - T) = 50 \; k_B \tag{4.65}$$

units of energy. The turning points should move out at least to r_{max},

$$r_{max} = r_{s,orb}\ e^{\frac{\Delta E}{2k_BT}} > 10^{10}\ r_{s,orb}, \tag{4.66}$$

where $r_{s,orb}$ stands for the initial orbital radius, which should be greater than the cross-sectional radius of the charged wire, r_{cs}. With $r_{cs} \approx 10^{-3}L$, r_{max} would grow to be greater than $10^7 L$ and the logarithmic particle would spend most of the time outside the box. Even if we could reduce r_s to atomic size (ignoring the problem of how to give the wire the desired charge), we would still need a cubic-metre-sized box for only 25 argon atoms, making the probability of an interaction between the argon and the ion completely negligible.

In the experiment, the wire sets a lower bound on the radial coordinate. The computational models have to deal with the singularity at the origin. The higher energies require long simulations, as the particle slowly reduces its velocity and turns around. Lower energy orbits become exponentially faster, but the numerical methods will also need smaller time steps to resolve the rapid changes in the slope for the potential. In the one-dimensional version of the problem, the logarithmic oscillator must always pass through the singularity. When the particle steps over the singularity, we could use equation (4.11) to determine how much time it should have taken the particle to reach the new position, correct the simulation time and the kinetic energy, and continue. Unfortunately, we would have to evaluate the error function numerically twice per period.

An alternative approach simply changes the potential to a new function with a similar shape but no singularity at the origin [99], as we have done in Figure 4.2. Campisi *et al.* used

$$V_1(s) = \frac{k_BT}{2}\ \ln\left(\frac{s^2+b^2}{b^2}\right). \tag{4.67}$$

At the origin, $V(0) = 0$. As for larger values of q, the potential approximates $k_BT\ \ln(s/b)$. The denominator b simply adds a constant term to the energy, so it has no effect on the dynamics. I preferred the simpler

$$V_2(s) = k_BT\ \ln\left(\frac{s}{b}+1\right). \tag{4.68}$$

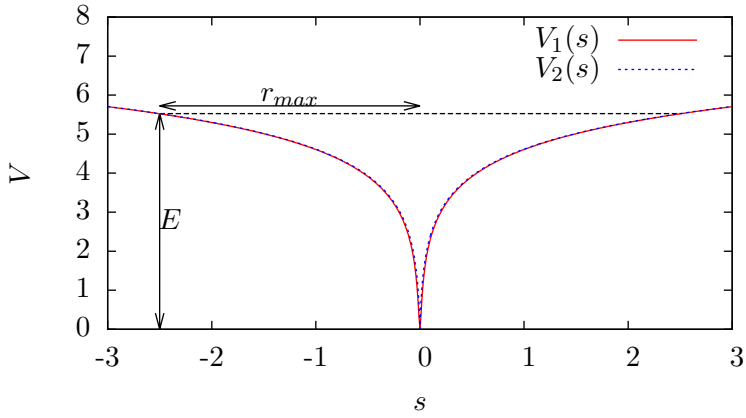

Figure 4.2: Modified potential energy functions V_1 (4.67) and V_2 (4.68) for the logarithmic oscillator simulations ($k_BT = 1$, $b = 10^{-2}$). The distance from the origin to the turning points, r_{max} depends exponentially on the system energy E.

Figure 4.3 displays the phase space trajectories for three different values of the energy.

The modified potentials induce a correction in the density of states that diminishes in importance as the energy rises. Note that, because the modified potentials introduce a lower bound on the energy, the energy range available to the logarithmic oscillator becomes

$$E \in \left[0,\ \frac{k_BT}{2}\ \ln\left(\frac{L^2+b^2}{b^2}\right)\right], \tag{4.69}$$

as the particle's excursions will presumably have the box size as an upper bound $r_{max} < L$. To double the energy range, we need to *square* L. Returning to our argon atoms, suppose that the initial state already has the desired temperature of $k_BT = 1$, but we want the logarithmic thermostat to allow typical fluctuations in the energy. If N equals the number of particles

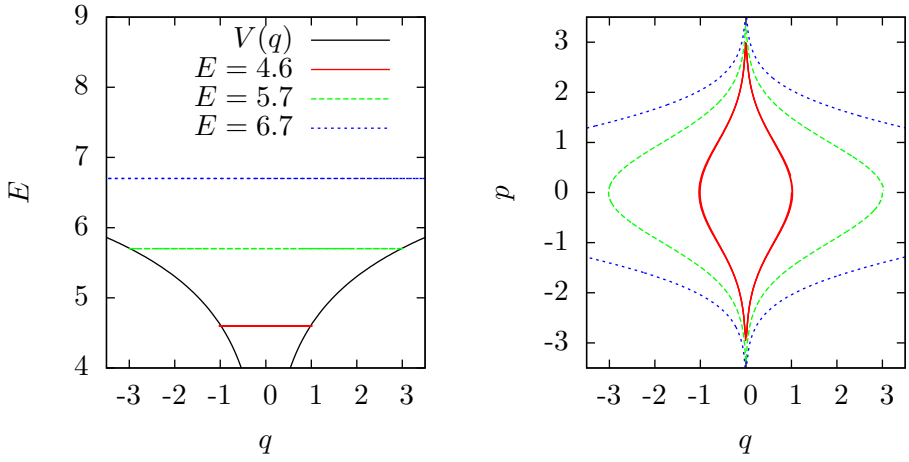

Figure 4.3: Logarithmic potential $V(q) = \ln\left(\frac{q}{b} + 1\right)$, $b = 10^{-2}$ (*left*) and trajectories in phase space (*right*) for the one-dimensional logarithmic oscillator. The figure shows the motion for three different values of the energy E.

in a gas, we expect the variance of the energy to satisfy (see (4.49))

$$\frac{\sqrt{\langle H^2\rangle - \langle H\rangle^2}}{\langle H\rangle} = \frac{k_B T \sqrt{\frac{\partial}{\partial T}\langle H\rangle}}{\langle H\rangle} \sim \frac{1}{\sqrt{\frac{3}{2}N}}. \tag{4.70}$$

We had $N = 25$ atoms, so a thermostat should allow at least for a fluctuation in the energy of about $\Delta E = \pm 0.16\ \langle H\rangle$. We begin with the oscillator at rest in the centre of the box to keep the total energy down. Then the box length must already be greater than 403 b ($L > r_{max} = e^{\Delta E/(2k_B)}$). Remember that the size of the time steps limits how small we can make b, and it probably does not make sense to make b smaller than an atomic diameter anyway.

Despite the drawbacks mentioned over the last few pages, I did not think

it impossible to make a logarithmic thermostat work, so I wrote a short arXiv contribution entitled *On the logarithmic oscillator as a thermostat* [115] pointing out what I believed were a few problems to think about, and then sent a copy to Michele Campisi and William Hoover.

Michele Campisi found my comments upsetting and replied quickly. He wrote that I had an overall negative attitude towards their work because I had not read the manuscript carefully. Further, he claimed that I was misrepresenting the content of their contribution, so it would be best for everyone if I withdrew my preprint from the arXiv before it got a time stamp, and then took some time to write a more accurate contribution.

Hoover wrote a couple of days later to say that he had found my comments helpful. He revealed that he had been asked to evaluate Campisi's paper for Physical Review Letters but, in the end, was left out of the review process because he had already been discussing the preprint with the authors.

Campisi's manuscript was eventually accepted and published in Physical Review Letters as an editor's choice article [99]. William Hoover considered writing a fiery comment article for the journal, but his wife, Carol Hoover, argued that it would bring "more heat than light".

After we exchanged some emails, Hoover and I decided to write a comment avoiding controversy and send it to the American Physical Society. In our manuscript, we drew the readers' attention to two "unusual" properties of the logarithmic thermostat. First, when Hoover connected the ends of a ϕ^4 chain to logarithmic oscillators with different λ parameters, they failed to produce a temperature gradient and heat flow in agreement with Fourier's law [118] (as we will see in the next section). Second, we explained that the time and length scales depended exponentially on the energy, as discussed above.

We also published an arXiv contribution with some supplementary information on our simulations [119]. Unfortunately, I interpreted some of my data incorrectly. The error did not invalidate our argument, but when it was detected by the team at Augsburg, they kindly pointed out to the editors of Physical Review Letters that I had made an "egregious mistake".

4.4 The comment and its aftermath

> While we can find systems with only a few degrees of feedom (less than one hundred, say) for which the time and length scales of the logarithmic oscillator do not become unreasonable, we argue that it does not act like an efficient thermostat even in these cases.

Problem number two in the statistical mechanics exam that we imagined at the beginning of this chapter mentioned an insulated liquid inside a vacuum flask. The students were expected to assume that the liquid's energy remained constant and that its probability distribution in phase space corresponded to the microcanonical ensemble. The annoying student used the canonical ensemble instead, so we marked her answer wrong. At office hours, she walks in and argues that she answered correctly. Vacuum flasks never insulate perfectly. They contain some gas particles that carry energy back and forth between the inner and outer surfaces of the flask. Therefore, the liquid will eventually reach thermal equilibrium with the surrounding air and sample the canonical distribution.

What makes this type of student particularly annoying is the fact that they are very often right in some sense [116]. Of course, there is no point in using a vacuum flask if it does not keep your tea warmer than room temperature, but it only does so for hours, not for months. Nevertheless, some serious research papers sometimes consider isolated systems as limiting cases of very weak thermal contact with a heat reservoir (see Vaikuntanathan and Jarzynski's *Dissipation and lag in irreversible processes* [59] for an example). We have to explain to our student that the use of a vacuum flask assumes times scales over which the liquid's internal energy changes negligibly, or just give her partial credit, which is quicker.

Campisi's thermostat also invokes the question regarding the relevant time scales. In his original paper, the logarithmic thermostat was thought of as a "fully deterministic version of the Andersen thermostat" [99]. Andersen had simulated collisions that changed the velocity of a randomly chosen particle [117]. The logarithmic oscillator samples the Maxwell-Boltzmann distribution for a temperature determined by the λ parameter, indepen-

dently of its energy [87]. This means that collisions with the oscillator at random times look like collisions with an ideal gas particle at the same temperature, supposing (and here is the key observation) that we contemplate time intervals larger than the period of oscillation.

To confirm my prediction about the slowness of the logarithmic thermostat, I replicated the one-dimensional simulation carried out at Augsburg. The logarithmic oscillator interacted through truncated Lennard-Jones potentials with two neutral particles in a box. A fourth-order Runge-Kutta took 2×10^9 time steps to generate a reasonable reproduction of the results shown by Campisi *et al.* (Figure 4.4).

I also simulated a single particle system interacting with the oscillator. Initially, I interpreted the data incorrectly, as I mentioned above. I naively assumed that the results were expected to converge to the same curve as in the case with two particles. Luckily, the mistake was corrected by Campisi and his colleagues.

Meanwhile, Hoover observed a similar behaviour with his ϕ^4 chain simulations [119]. Under propitious conditions, with the system temperature starting off near the desired temperature, the system equilibrated to the right temperature given a sufficiently long simulation (10^9 time steps). Beginning with a more distant initial temperature, the system showed no tendency at all towards equilibration, as the logarithmic oscillator seemed unable to absorb much energy during the simulation. Furthermore, connecting the chain to logarithmic oscillators with different λ parameters failed to produce heat flow.

Campisi had vehemently protested that I had misrepresented his work. He argued they had never claimed that logarithmic thermostats were useful in simulations because they were well aware of the limitations, and they had insisted on their application to (very) small systems. Under the circumstances, I was most surprised by their reply in the arXiv [120], as they completely ignored the central issue of the exponential time and length scales! They simply corrected my mistake and disregarded Hoover's results as off-topic and not sufficiently well documented.

At this point, Hoover and I joined forces with my advisor, P. Español.

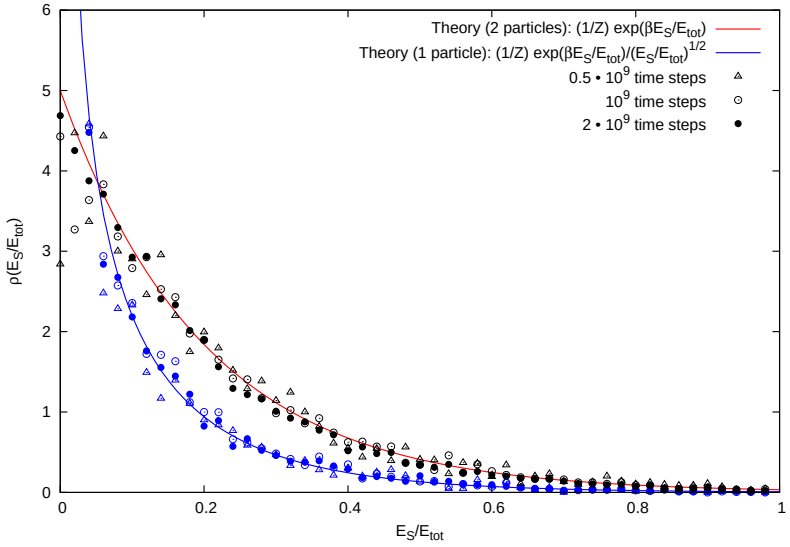

Figure 4.4: Probability distribution for E_S/E_{tot} in the original numerical experiment proposed by Campisi *et al.* [99]. The system interacting with the logarithmic oscillator has an energy E_S, while E_{tot} denotes the total energy of the combined system. The black points correspond to the numerical results for $t = 0.5 \times 10^6$, 10^6 and 2×10^6 for a system of two particles, as in the article (the time step was set to $\Delta t = 0.001$). The blue points correspond to a system of only one thermostated particle. The solid line follows the theoretical prediction for a system in contact with a heat bath (red for a system of two particles, blue for only one).

We rewrote the comment, reconstructing our argument from direct quotations of Campisi's paper.

Campisi *et al.* chose the size of the box equal to $L = 2\sigma\sqrt{e^{2E_{tot}/(k_BT)} - 1}$, which guaranteed that the ion never left the box. The energy E_{tot} should have a "large" value to compensate for the effect of truncation in the modified potential. They estimated that, if the number N of atoms in the box increased, then E_{tot} should increase accordingly, $E_{tot} \propto 3Nk_BT/2$ (for two and three particles, the values chosen for E_{tot} were $5k_BT$ and $8k_BT$). For a system of just 26 atoms, this meant that the length L of the box was larger than the diameter of the Earth (setting $\sigma = 10^{-10}$, as in the article, and $k_BT = 1$). Moreover, the 26 atoms plus the ion would take extraordinarily long to equilibrate. The mean free time τ for the logarithmic oscillator in the most propitious case (a flat square box of side L and height σ) would exceed the age of the universe ($m = 1$amu)

$$\tau \sim \sqrt{\frac{m}{k_BT}}\frac{L^2\sigma}{4\pi N\sigma^2} \sim 10^{20} \text{ s.} \tag{4.71}$$

The comment was published in Physical Review Letters [121] together with a reply by Campisi *et al.* [122], who provided a table of values that was supposed to show the possibility of an experimental setup with realistic time and length scales in the context of present day cold-atom technology [123]. They described an implementation with N rubidium atoms ($m = 85.4678$ amu) in which the number of degrees of freedom was cut down to one third by forcing all the particles to move along a single dimension. Table 4.5, copied from their reply, illustrates the exponential growth of mean free times τ and box lengths L as the required precision H_{KS} or the number of particles N increase.

Having no prior experience with cold-atom physics, we contacted Prof. I. Bloch, who kindly lent us some of his time and confirmed that such a precise one-dimensional setup, though "challenging", should be feasible in principle.

In addition to admiring the wonderful many-digit precision of modern calculators, I would like to dwell on an important detail regarding the

Table 4.5: Total energy, box lengths and mean free times for the logarithmic oscillator experiment as a function of the number of degrees of freedom, N, and the required precision, H_{KS}, measured as a Kolmogorov-Smirnov distance (from Campisi *et alii* [122]).

N	H_{KS}	E_{tot}/k_B	$L\ [m]$	$\tau\ [s]$
20	0.005	16.45	2.78724×10^{-1}	1.41295×10^{-3}
20	0.01	14.8	5.35289×10^{-2}	2.71358×10^{-4}
20	0.02	13.1	9.77885×10^{-3}	4.95726×10^{-5}
20	0.03	11.9	2.94533×10^{-3}	1.4931×10^{-5}
20	0.04	11.05	1.25888×10^{-3}	6.38172×10^{-6}
10	0.02	7.75	4.64314×10^{-5}	4.70756×10^{-7}
20	0.02	13.1	9.77885×10^{-3}	4.95726×10^{-5}
30	0.02	18.1	1.45131×10^{0}	4.95726×10^{-3}
40	0.02	23.1	2.15393×10^{2}	5.45955×10^{-1}
50	0.02	28	2.89251×10^{4}	5.86529×10^{1}

magnitudes in Table 4.5: they assume that the rubidium atoms start off at a temperature very close to the desired thermostat temperature. With an initial temperature off by ΔT degrees, the logarithmic thermostat would have to absorb at least $\Delta E = Nk_BT/2$ units of energy. For $N = 20$ and $\Delta T = 5$ K, for instance, the energy absorbed would be about $\Delta E = 50k_B$. Compare this value to those displayed in the table, where the total energy of the whole system plus the oscillator never exceeds $30k_B$. Thus, we can describe the logarithmic oscillator as a "heat bath simulator", but not as a thermostat. While it cannot effectively absorb excess heat, it does induce the canonical ensemble over long time scales under the right conditions, when the initial system temperature lies close to the oscillator's average kinetic temperature.

In their reply, Campisi and his coworkers disregarded the criticism related to the absence of heat flow with two different thermostat temperatures. They simply replied that their letter had suggested that logarithmic oscillators would illuminate the study of time-varying temperatures and

not temperature gradients in space, and added that the criticism was "neither sufficiently documented nor conclusive" (although without explaining why)[6].

Campisi and Hänggi felt that they had provided a "response which dispels a criticism raised by Hoover and coworkers" [87]. We were not convinced.

Shortly after, Hoover noticed that the logarithmic oscillator Hamiltonian (4.30) had the wrong configurational temperature (4.9),

$$\frac{\left(\frac{\partial H_{log}}{\partial r_s}\right)^2}{\frac{\partial^2 H_{log}}{\partial q^2}} = -k_B T, \tag{4.72}$$

that is, it had the same magnitude as the average kinetic temperature, but the opposite sign! J. Horowitz suggested that we examine what happens when you start off a ϕ^4 chain heat flow simulation with initial velocities assigned in accordance with the linear temperature gradient between the two logarithmic thermostats. When we carried out the simulation, we saw that the logarithmic particles responded too slowly, so the temperature profile quickly became flat, as the chain equilibrated along its length, ignoring the "thermostats".

More recently, Sponseller and Blaisten-Barojas have reported their results with simulations of small clusters of four and five Lennard-Jones particles and of rubidium atoms interacting weakly with a logarithmic oscillator [124]. They observe that the kinetic energy of interatomic vibrations was not significantly affected by the coupling to the logarithmic thermostat, so they interpret this as a failure of the logarithmic particle to act as a thermostat.

[6]In their arXiv reply, they had added that it was "not possible to infer whether the simulations were done in the proper parameter regime and to draw any conclusions from them" [122].

4.5 Equilibrating FPUT chains

We illustrate the presence of memory effects by analysing the equilibration of two anharmonic FPUT chains in weak contact.

Before we turn to some final general thoughts on memory and time, we will try to get some insight from a different problem: the equilibration of Fermi-Pasta-Ulam-Tsingou (FPUT) chains[7] [126].

The statistical mechanical image of an ideal gas neglects the interaction among the particles. If this interaction were truly inexistent, there would be no way for the gas to equilibrate. Similarly, when we represent a solid crystal as an array of harmonic oscillators on a lattice, we find no coupling between different normal modes of oscillation. This means that the energy associated with each mode remains in that mode forever and that the crystal cannot approach equilibrium, where the law of equipartition states that each mode should have a fair share of the total internal energy.

In the early 1950s, Fermi proposed the use of computers to reveal the mathematical properties of systems that he could not study analytically, becoming a pioneer in the field we now refer to as computational physics. The numerical experiments were supposed to demonstrate mixing and thermalisation. The team studied a one-dimensional chain of harmonic oscillators with a small nonlinear perturbation. To their surprise, the energy stored in the first mode migrated towards higher modes but then, after a very long time, returned almost completely to the first mode [125]. The unexpected outcome, known as the Fermi-Pasta-Ulam paradox, demonstrated that nonlinearity did not necessarily imply the equipartition of energy. A conclusion strengthened by the work of Kolmogorov, Arnold and Moser, who together managed to prove that slightly perturbed integrable Hamiltonian systems remain quasiperiodic [50]. By increasing the nonlinear perturbation, we can destroy the quasiperiodic orbits [126].

We wish to study the disorganised transmission of energy between FPUT

[7] Traditionally called Fermi-Pasta-Ulam chains. Thierry Dauxois has persuasively argued that we should also recognise Mary Tsingou's contribution [127].

chains. The Hamiltonian of a single chain,

$$H(q,\ p) = \sum_{i=1}^{N} \left(\frac{p_i^2}{2m_i} + \frac{1}{2}k\ (q_i - q_{i-1})^2 + \frac{1}{4}\kappa\ (q_i - q_{i-1})^4 \right), \tag{4.73}$$

gives rise to complex trajectories, with solitons in the continuum limit for some values of k and κ [126]. We need to make sure that our chains redistribute the energy in one normal mode among the other modes. With all the energy stored in the nth mode, when the oscillation reaches its maximum amplitude and the kinetic energy vanishes, the energy $e(n)$ equals

$$e(n) = \frac{k}{2}\sum_{i=1}^{N}(q_i - q_{i-1})^2 + \frac{\kappa}{4}\sum_{i=1}^{N}(q_i - q_{i-1})^4. \tag{4.74}$$

The original article by Fermi *et al.* kept the ends of the chain fixed, making the coordinates q_i for a chain oscillating in the nth mode

$$q_i = A\ \sin\left(2\pi n\ \frac{i}{N}\right). \tag{4.75}$$

Substituting the expression for the coordinates in the energy above,

$$\begin{aligned} e(n) = {} & \frac{k}{2}A^2\sum_{i=1}^{N}\left(\sin\left(2\pi n\ \frac{i}{N}\right) - \sin\left(2\pi n\ \frac{i-1}{N}\right)\right)^2 \\ & + \frac{\kappa}{4}A^4\sum_{i=1}^{N}\left(\sin\left(2\pi n\ \frac{i}{N}\right) - \sin\left(2\pi n\ \frac{i-1}{N}\right)\right)^4. \end{aligned} \tag{4.76}$$

Chains with many links ($N \gg 1$) allow us to replace the difference of sines with a cosine as long as the coordinate values do not change abruptly from one link to the next. The lower modes satisfy this condition.

$$\frac{1}{N}\ \left(\sin\left(2\pi n\ \frac{i}{N}\right) - \sin\left(2\pi n\ \frac{i-1}{N}\right)\right) \approx 2\pi n \cos\left(2\pi n \frac{i}{N}\right). \tag{4.77}$$

Inserting the approximation into the expression for the energy,

$$\begin{aligned} e(n) =& \frac{k}{2}A^2\frac{(2\pi n)^2}{N}\sum_{i=1}^{N}\frac{1}{N}\ \cos^2\left(2\pi n\frac{i}{N}\right) \\ &+ \frac{\kappa}{4}A^4\frac{(2\pi n)^4}{N^3}\sum_{i=1}^{N}\frac{1}{N}\ \cos^4\left(2\pi n\frac{i}{N}\right). \end{aligned} \tag{4.78}$$

Now we can use integrals to approximate the sums,

$$\sum_{i=1}^{N}\frac{1}{N}\ f\left(\frac{1}{N}\right) \approx \int_0^1 f(x)\ dx. \tag{4.79}$$

Calculating the resulting integrals for the lower modes, $n = 1,\ 2,\ 3,\ \ldots$,

$$\int_0^1 \cos^2(2\pi nx)\ dx = \frac{1}{2}, \tag{4.80}$$

$$\int_0^1 \cos^4(2\pi nx)\ dx = \frac{3}{8}, \tag{4.81}$$

we are left with a biquadratic equation,

$$\frac{\kappa}{4}A^4\frac{(2\pi n)^4}{N^3}\frac{3}{8} + \frac{k}{2}A^2\frac{(2\pi n)^2}{N}\frac{1}{2} - e(n) = 0, \tag{4.82}$$

which we can easily solve to obtain

$$A(e(n)) \approx \frac{N}{\pi n}\sqrt{\frac{-1+\sqrt{k^2+\frac{6\kappa e}{N}}}{3\kappa}}. \tag{4.83}$$

Equation (4.83) tells us how to store any given amount of energy in the nth normal mode. We then need to make sure that the energy drifts away to other modes. We can do this by keeping track of the amplitudes of the Fourier components. We mentioned before that a sufficiently large nonlinear perturbation could make the dynamics ergodic. Equivalently, we

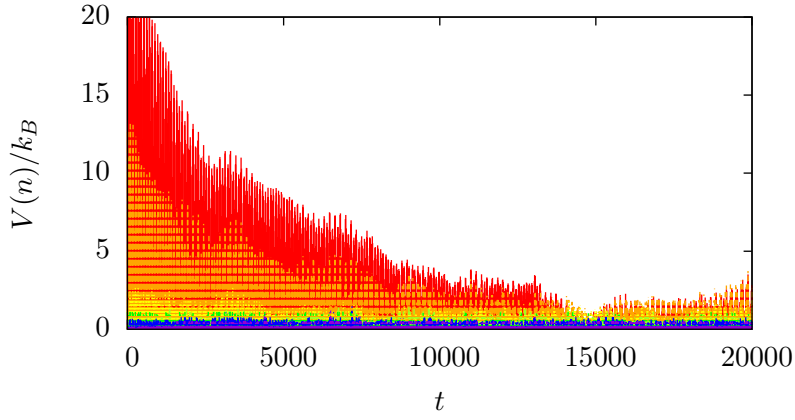

Figure 4.6: Potential energy associated with the first nine normal modes versus time. The colours correspond to different modes, with the lower frequencies closer to red, starting at $n = 1$, and the higher frequencies shifted towards blue, ending in violet, $n = 9$. The slight rise at the right of the plot did not continue. The modes plotted kept below $5k_B$ energy units for the following 20000 time units.

can increase the internal energy, for this makes the oscillations wider and the nonlinear contribution more significant. By dumping enough energy into the first mode, $n = 1$, I observed that the system approached equipartition for long times (see Figure 4.6). The numerical integration was carried out with a velocity Verlet algorithm with periodic boundary conditions, to avoid unwanted edge effects.

I argued in Section 3.2 that an appropriate choice of macroscopic variables should render the past history of a process relevant for a very short time. Mathematically, this condition amounts to stating that $\delta\rho = \rho - \bar{\rho}$ tends to vanish quickly when we view the phenomenon on the macroscopic time scale. We will see that this premise imposes some limits on how we should set up our experiments and simulations.

A simple model should make the point clear (see Figure 4.7). We have

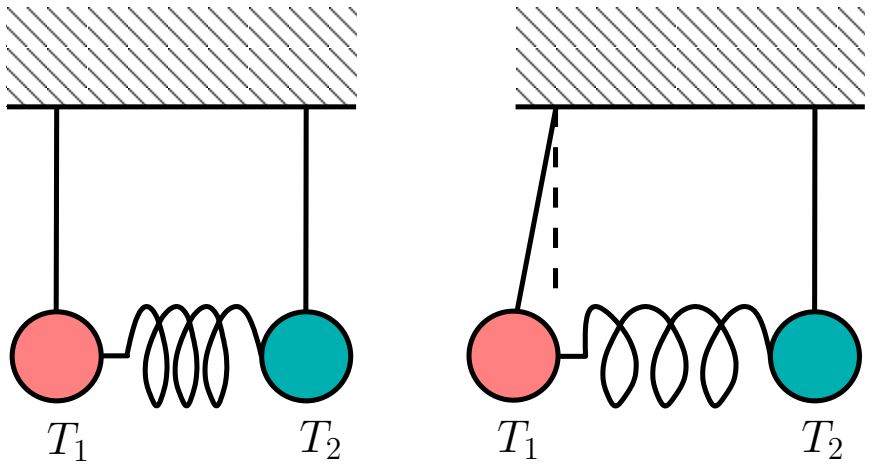

Figure 4.7: Isolated system composed of two metal spheres at different temperatures connected with a weak metallic spring in a state of mechanical equilibrium (*left*), and with the left sphere displaced (*right*).

two identical metal spheres insulated at different temperatures and we bring them into contact by means of a thin conducting wire with negligible mass. Heat flow will ensue, but the spheres will not have a single temperature during the process. Rather, a heat gradient will appear between the points closer to the wire and those further away. If we had only picked the two internal energies as our relevant variables, then the relevant distribution $\bar{\rho}$ would correspond to a flat energy density for both spheres, in the sense that it would assign a small probability of finding some part of the sphere with its internal energy departing significantly from the average. Therefore, the probability density, ρ, would spend most of the time far from the relevant distribution during the relaxation towards equilibrium.

Now suppose that energy flows slowly between the two spheres, compared to the time it takes the spheres to equilibrate on their own. Macroscopic work could still keep the combined system from following memoryless laws. Imagine the wire as a weak spring. If we displace one of the spheres

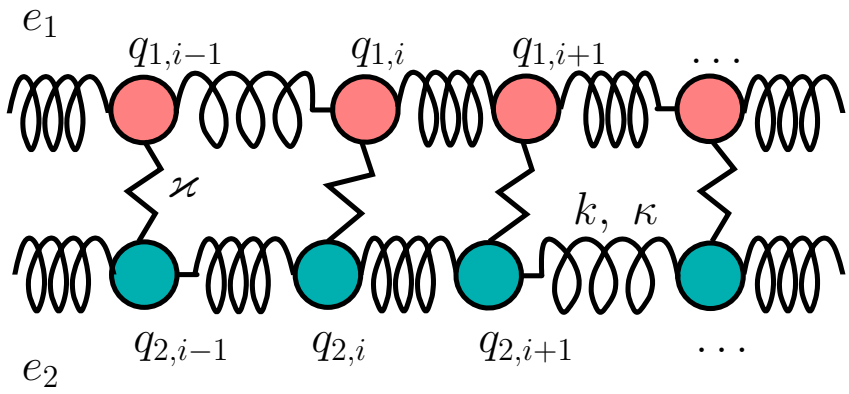

Figure 4.8: Two Fermi-Pasta-Ulam-Tsingou chains in contact through weak linear springs ($\varkappa \ll k$). The coordinates $q_{j,i}$ represent the position of the ith link in chain j (1 or 2) relative to its corresponding lattice node. The strength of the quadratic and quartic potentials for the FPUT models were tuned with k and κ, respectively.

from its equilibrium position, it will start to oscillate and we will observe an ebb and flow of mechanical energy between the spheres that does not take into account which of the spheres is hotter. The final state will evidently also coincide with mechanical and thermal equilibrium, but the energies will no longer follow a monotonic approach like the one plotted in figures 3.1 and 3.7.

Returning to the FPUT models, we cannot simply connect the ends of two chains together, because we want to avoid the formation of heat gradients. We need to bring them into global contact, as in the hard-sphere gas example in the previous chapter. With that in mind, we connect each link in one chain to a corresponding link in the other through a weak linear spring with constant $\varkappa \ll k$ (weak because we would like to neglect the interaction energy compared to the energy of each of the chains). The diagram in Figure 4.8 displays the setup.

Although the chains did approach equilibrium in the long run, the tran-

sient process differed from the monotonic relaxation that we would expect from equation (3.90). If there really were a separation of time scales, then we should observe the same relaxation in the FPUT simulations as in the hard-sphere gas example (apart from a scaling factor on the time axis). Figure 4.9 shows a typical evolution of the energies of two 256-particle chains. Note that when the gases approached equilibrium in Figure 3.3, the energy trajectories curved towards the equilibrium temperature, instead of away from it, as in Figure 4.9.

If we were wrong about the separation of time scales in this problem, then that meant that the memory kernel in equation (3.77) did not decay rapidly. Checking the time correlation function $\mathrm{C}_{\hat{\rho}}[\hat{L}\mathrm{i}E_1,\ \hat{L}\mathrm{i}E_1(t-s)]$ confirmed this intuition (see Figure 4.10). The origin of this phenomenon lies in a subtle form of macroscopic work.

Let us disregard the anharmonic perturbation for the moment. The chains behave in each normal mode as if they were two coupled oscillators, each with the characteristic frequency of the mode, connected with the weak spring [128] (see Figure 4.11). The equations of motion for the ith pair of links when they are vibrating in the nth mode are

$$\left.\begin{aligned} \ddot{q}_{1,i} &= -\omega_n^2 q_{1,i} - \frac{\varkappa}{m}\,(q_{1,i} - q_{2,i}), \\ \ddot{q}_{2,i} &= -\omega_n^2 q_{2,i} - \frac{\varkappa}{m}\,(q_{2,i} - q_{1,i}). \end{aligned}\right\} \tag{4.84}$$

We have used ω_n to denote the frequency of the nth mode. A change of variables to the new $x_1 = q_{1,i} + q_{2,i}$ and $x_2 = q_{1,i} - q_{2,i}$ transforms the equations above into

$$\left.\begin{aligned} \ddot{x}_1 &= -\omega_n^2 x_1, \\ \ddot{x}_2 &= -(\omega_n^2 + 2\frac{\varkappa}{m})\ x_2. \end{aligned}\right\} \tag{4.85}$$

This system has an easy solution,

$$\left.\begin{aligned} x_1 &= A\ e^{\mathrm{i}\omega_n t}, \\ x_2 &= B\ e^{\mathrm{i}t\ \sqrt{\omega_n^2 + 2\kappa/m},} \end{aligned}\right\} \tag{4.86}$$

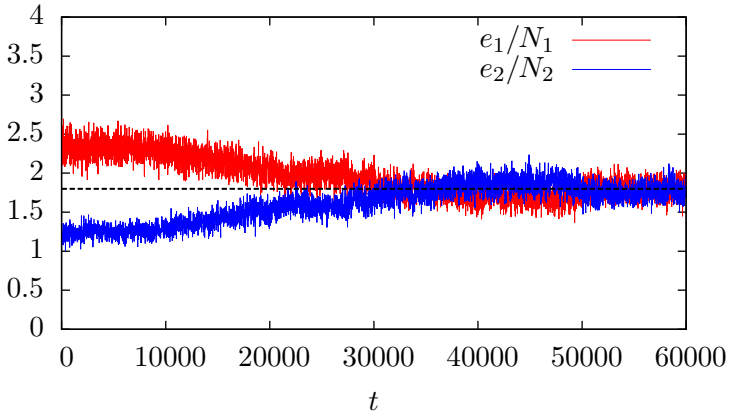

Figure 4.9: Energies per particle stored in two Fermi-Pasta-Ulam-Tsingou chains ($N_1 = N_2 = 256$) interacting weakly versus time. The chains were allowed to equilibrate separately before activating the coupling between them. The dashed line indicates the average energy per particle at equilibrium.

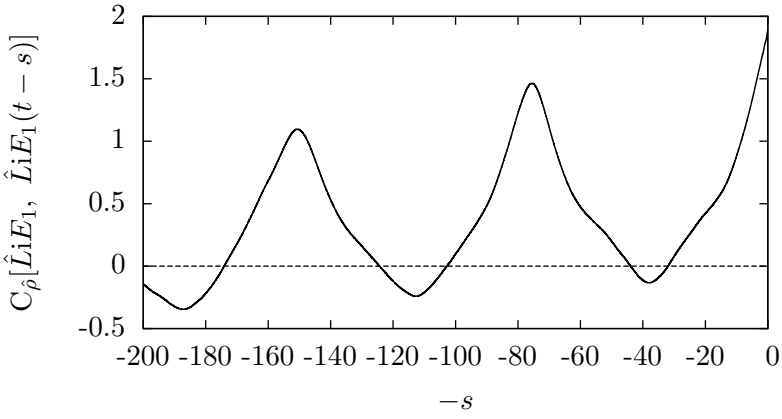

Figure 4.10: Time correlation function $C_{\hat{\rho}}[\hat{L}iE_1,\ \hat{L}iE_1(t-s)]$ versus $-s$ for interacting FPUT chains with initial energies $e_1 = 512$ and $e_2 = 256$. Clearly, the correlation does not rapidly decay to zero. Note that the conventional representation of time correlation functions is a mirror image of the plot shown here.

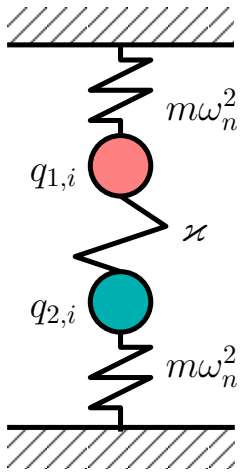

Figure 4.11: The interaction between harmonic chains behaves in each mode like a pair of coupled oscillators exchanging energy through a weak linear spring. We can represent the characteristic frequency of oscillation as if it came from the action of a linear spring with constant equal to $m\omega_n^2$.

that we can convert back to the original variables $q_{1,i}$ and $q_{2,i}$,

$$\left.\begin{aligned} q_{i,1} = \frac{x_1 + x_2}{2} &= e^{\mathrm{i}\omega_n t}\left(\frac{A}{2} + \frac{B}{2}\, e^{\mathrm{i}t\,\sqrt{1+2\omega_\varkappa^2/\omega_n^2}}\right), \\ q_{2,1} = \frac{x_1 - x_2}{2} &= e^{\mathrm{i}\omega_n t}\left(\frac{A}{2} - \frac{B}{2}\, e^{\mathrm{i}t\,\sqrt{1+2\omega_\varkappa^2/\omega_n^2}}\right). \end{aligned}\right\} \tag{4.87}$$

The frequency of the weak spring was abbreviated to $\omega_\varkappa = \varkappa/m$. We see now that the coordinates oscillate with frequency ω_n, with the amplitude oscillating with frequency $\sqrt{\omega_n^2 + 2\omega_\varkappa^2}$.

The slower modes transfer energy more effectively between the chains because the same amount of energy leads to larger amplitudes when n is smaller (4.83), creating a tighter pull on the springs connecting the chain. The energy stored in the lower modes moves periodically back and forth between the chains affecting the total energy in the chain. If we wish e_1 to

behave like a slow variable, then the ebb and flow of energy in the lower modes should be slow compared to the time τ it takes the energy to drift away into higher modes. We will not be able to use the slow variable approximation unless

$$\frac{2\pi}{\sqrt{\omega_n^2 + 2\omega_\varkappa^2}} \ll \tau, \tag{4.88}$$

or equivalently,

$$\omega_n^2 + 2\omega_k^2 \ll \frac{4\pi^2}{\tau^2}. \tag{4.89}$$

Our simulations did not satisfy this condition. However, although we can easily make the left side smaller by adding more links to the chains (as this reduces the frequency of the lower modes), the main problem lies in the large value of τ (of order 10^3 in our computations), which increases with the number of particles. If we insisted on shortening the memory, one possibility would be to make the nonlinear contribution stronger. This would amount to simulating a different kind of system, no longer a harmonic chain with a small nonlinear perturbation. An altogether different approach would include the amplitudes of the first few modes as additional relevant variables.

I do not wish to argue here that it is impossible to set up FPUT simulations that follow a slow variable relaxation to equilibrium. Rather, my point is that, for subtle reasons, we would be wrong to postulate a separation of time scales for the kind of setup explained above. The FPUT chains (or a system in contact with a logarithmic oscillator) may well display ergodic dynamics. If they do, this provides information about average values *in the long run*, but the values observed over the time scales of interest could depart from this prediction.

For example, we could eliminate the weak springs between the FPUT chains and add a few hard-spheres moving around. Every once in a while, a sphere would collide with one of the links. I suppose the chains would have enough time to equilibrate between collisions, eliminating the correlations. The chains would then inch towards equilibrium according to the laws for slow variables. Similarly, if we keep the interaction with the logarithmic

oscillator down to a single collision every few periods, then mixing dynamics will end up sampling the canonical distribution. But how long do we have to wait?

4.6 Memory and dissipation

The irrelevant part of the distribution, which we can estimate from the energy dissipated into a heat bath, helps us quantify the significance of memory effects in a nonequilibrium process.

We have finally come to the moral of our story: *if we wish to infer memoryless macroscopic equations from microscopic laws, then we need enough information to determine a relevant distribution that closely approximates the solution of Liouville's equation,* (2.2) *or* (3.124). The time integral in our generalised Fokker-Planck equations disappears when $\delta\rho$ vanishes.

In the particular case of heat transfer, systems should forget their previous states very quickly, so that we can treat subsequent microscopic interactions as independent of what happened before. Interestingly, we can simulate this behaviour and build systems that "pretend" not to remember, in much the same way that a pseudorandom number generator "pretends" that the numbers it produces have no relation to each other. Nosé-Hoover dynamics are a prominent example of this kind of setup.

Horowitz and Vaikuntanathan have shown that systems with feedback satisfy a generalised version of equation (2.39) [129],

$$\mathrm{D}(\rho\|\rho') = \frac{1}{k_B T}\langle W_{diss}\rangle + \langle I\rangle\,. \tag{4.90}$$

The mutual information $\langle I\rangle$ represents the change in uncertainty about the microstate as a consequence of the measurement employed by the feedback mechanism. Nosé-Hoover dynamics, for instance, "measure" the total kinetic energy to determine how to modify the friction. Even without a heat reservoir ($W_{diss} = 0$), we can produce an equivalent change in the relative entropy through feedback.

The logarithmic oscillator behaves differently in the experiments proposed by Campisi *et alii*. Their proof relies on the oscillator moving independently, except for a negligible interaction, so $\langle I \rangle \approx 0$, which means that the oscillator does not "pretend" to be a memoryless system.

This chapter cautions against a thoughtless application of the techniques of coarse-graining without carefully considering the relevant variables and time scales. To avoid memory effects, we need the irrelevant part of the distribution to decay rapidly. Can we estimate the importance of memory effects *a priori*?

We always begin with a relevant ensemble which evolves into a distribution that we can divide into two parts, $\rho(t) = \bar{\rho}(t) + \delta\rho(t)$. We take a small value of t, when $\bar{\rho}$ still approximates ρ well, with $\delta\rho$ consisting of a small undulation superimposed on ρ, that is, $\delta\rho/\bar{\rho} \ll 1$. Section 2.8 explained that the dissipated work establishes an upper bound on how much ρ can depart from the relevant distribution $\bar{\rho}$, assuming the latter is an equilibrium ensemble. We estimate the value of the logarithm $\rho/\bar{\rho}$,

$$\ln\left(\frac{\rho(z)}{\bar{\rho}(z)}\right) = \ln\left(1 + \frac{\delta\rho(z)}{\bar{\rho}(z)}\right) \approx \frac{\delta\rho(z)}{\bar{\rho}(z)} \tag{4.91}$$

and insert it into the relative entropy,

$$\begin{aligned}\Delta S(\rho(t)\|\bar{\rho}(\lambda(t))) &\approx -k_B \int \delta\rho(z) \left(1 + \frac{\delta\rho(z)}{\bar{\rho}}\right) dz \\ &\approx -k_B \int \delta\rho(z)\, dz = -k_B \,\mathrm{Tr}[\delta\rho].\end{aligned} \tag{4.92}$$

We allow some large values of $\delta\rho$, as long as they do not contribute significantly to the integral. We can pick a given microstate along the trajectory followed by the system and use it as an initial state in a different simulation. With the external parameters fixed, we let the system equilibrate with a reservoir at temperature T equal to the initial temperature of the system and determine the average dissipated work $\langle W_{diss} \rangle$ (2.39).

$$\Delta S(\rho(t)\|\bar{\rho}(\lambda)) = -\frac{\langle W_{diss} \rangle}{T} \approx -k_B \,\mathrm{Tr}[\delta\rho]. \tag{4.93}$$

We can repeat the process for different values of t to get an idea of how the dissipated work $\langle W_{diss} \rangle$ varies in time. Because $\mathrm{Tr}[\delta\rho] \approx \langle W_{diss} \rangle / (k_B T)$, if we observe a more or less constant amount of energy dissipated as time goes by, then this means that the relevant distribution $\bar{\rho}$ still approximates ρ reasonably closely. But if the dissipated work starts to grow uncontrollably, then we expect ρ and $\bar{\rho}$ to drift away from each other (in the sense of the Kullback-Leibler divergence between them). If this evolution away from the relevant distribution continues over macroscopic time scales, then the slow variable approximation will no longer be justified and we will have to incorporate memory effects into our descriptions.

One of Zwanzig's early claims was that his equations made it possible to study memory effects in irreversible thermodynamics [73]. The exact evolution equations for relevant variables unveiled in chapter 3 are better suited to this task because they are easier to solve numerically. However, remember that we need to determine time correlation functions over the whole time interval that contributes to the memory. If we have the computational power to carry out this calculation with a long simulation, then it will usually be easier to solve the problem directly with molecular dynamics.

4.7 Summary

There are many possible definitions of temperature for nonequilibrium states, which all coincide at equilibrium. Deterministic thermostats monitor the value of one of these temperatures and use some kind of feedback to modify the trajectories in phase space in such a way that the system behaves as if in contact with a heat reservoir. We presented a plausible reconstruction of Nosé's thought process leading up to his canonical dynamics and a short explanation of Nosé-Hoover equations as a prelude to the discussion about the logarithmic thermostat proposed by Campisi *et alii*.

We discussed the dynamical and statistical properties of the logarithmic oscillator and demonstrated that, while it does induce the canonical ensemble when brought into weak contact with another system (assuming

ergodicity), it differs from a heat reservoir in that it has no tendency to warm or cool the system in contact with it. The logarithmic potential leads to orbital radii and oscillation periods that grow exponentially with the energy. Furthermore, it only behaves as if it were at the reservoir temperature over time scales much larger than the period of oscillation. Therefore, when it manages to induce a canonical distribution, it does so extremely slowly.

A heat bath should act as if it had quickly "forgotten" its previous interactions with a system. To analyse this idea, we turned to Fermi-Pasta-Ulam-Tsingou chains and showed how the energy trapped in the lower modes of vibration drive the behaviour away from the monotonic approach to thermal equilibrium described in Chapter 3.

I concluded with some comments on how to estimate the importance of memory effects, bringing my part to an end, and leaving you, dear reader, with the task of pondering over the worth of my contribution. I hope my work has contributed to change your opinions on some of these matters. When I think back to the beginning of this project, I see that it has irreversibly changed mine.

Conclusions

Reversible memoryless laws predict the emergence of irreversibility and memory when we coarse-grain our descriptions. Within the general framework of Markov processes, we have demonstrated that reversible laws generally give rise to irreversible transitions between macrostates. Furthermore, we have used a bit-reversible scheme to verify this fact numerically in the particular case of Hamiltonian mechanics. The algorithm, though inspired by the work of Levesque and Verlet, differs from their proposal in two important aspects, as it is both memoryless and symplectic. Because the numerical trajectories lie on the exact solution of a Hamiltonian problem (related to our original setup through a slight perturbation), our simulations prove that irreversible phenomena may follow from reversible laws. We can retrace the trajectories backwards to the very last bit, so we know that the observed irreversibility is not an artefact due to truncation.

We have proposed a way to measure the irreversibility of transitions between macrostates as a ratio between direct and reverse transition probabilities. We saw that this quantity was related to the difference of the Boltzmann entropies for the macrostates. From this realisation we derived the Jarzynski equality, the work inequality and the dissipated work for a system in contact with a heat bath.

We expressed dissipation in terms of relative entropy. Interestingly, the maximisation of relative entropy constitutes a distinct variational principle from which we can derive relevant distributions, which constitute a central object in the theory of coarse-graining. We proved the equivalence between the relative entropy principle and Jaynes's maximum entropy for-

malism when the equilibrium distribution determines the probabilities of the dynamical invariants. In spite of this equivalence, the relative entropy approach simplifies the sampling of distributions with unknown or cumbersome dynamical invariants. Furthermore, we should be able in principle to use the results of a simulation to calculate the averages for *different* values of the simulation parameters, especially near equilibrium. We have presented some tentative results along these lines, but future research should determine the method's efficiency.

The third chapter derives the exact dynamical equations for the evolution of macroscopic variables and the corresponding probability distributions, following a process similar to Zwanzig's projection operator technique. We have presented the theory without resorting to projection operators. While this might be interesting from the pedagogical point of view, the central contribution of this thesis consists in the absence of projected dynamics in the coarse-grained equations. This avoids the need for approximations to describe the projected dynamics, as in Zwanzig's theory, and simplifies the analytical treatment of the equations and their connection to simulation results. We also expect it will ease the way when examining dynamical processes in which memory effects play a relevant part.

We can disregard memory effects as long as the relevant distribution approximates the evolution of the initial distribution closely, in the sense of a small Kullback-Leibler divergence between them. We have suggested how to estimate the importance of memory effects, but have left the detailed investigation to further research.

Classical statistical mechanics determines the macroscopic properties of equilibrium states, which do not change in time. When we consider macrosopic *processes*, we must consider the relevant time scales carefully. Properties calculated by averaging over the whole set of accessible microstates may not adequately represent the behaviour of a system over the time scales of interest, as we showed in the case of the logarithmic thermostat. Although the logarithmic oscillator does cause a system to sample the canonical distribution through weak coupling (assuming ergodicity), it does so over such vast time and length scales that it cannot be used as an efficient thermostat.

Bibliography

[1] Carroll, L.: *Sylvie and Bruno*, in *The Complete Illustrated Works of Lewis Carroll*. Chancellor Press (1993).

[2] Loschmidt, J. *Über den Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft*. Sitzungsber. Kais. Akad. Wiss. Wien Math. Naturwiss. **73**: 128–142 (1876).

[3] Boltzmann, L.: *Further Studies on the Thermal Equilibrium of Gas Molecules*, in *The Kinetic Theory of Gases*. World Scientific (2003).

[4] Boltzmann, L.: *On the Relation of a General Mechanical Theorem to the Second Law of Thermodynamics*, in S. Brush: *Kinetic Theory. Volume 2. Irreversible Processes*. Pergamon (1966).

[5] Lebowitz, J. L.: *Boltzmann's Entropy and Time's Arrow*. Physics Today Online **46**, 9: 32–38 (1993).

[6] Maguire, G.: *Out of Oz*. William Morrow (2011).

[7] Kelly, F. P. *Reversibility and Stochastic Networks*. Wiley (1979).

[8] Jaynes, E. T.: *Information Theory and Statistical Mechanics*, in *Brandeis University Summer Institute Lectures in Theoretical Physics*, Volume 3: *Statistical Physics*. W. A. Benjamin INC. (1963).

[9] Shannon, C. E.: *The Mathematical Theory of Communication*. University of Illinois Press, Urbana (1949).

[10] Jaynes, E. T.: *Gibbs vs Boltzmann entropies*. American Journal of Physics **33**, 5: 391–398 (1965).

[11] Lewis, G. N.: *A New Principle of Equilibrium*. Procedures of the National Academy of Sciences of the United States of America **11**, 3: 197–183 (1925).

[12] Español i Garrigós, J.: *Initial ensemble densities through the maximization of entropy*. Physics Letters A **146**, 1, 2: 21–24 (1990).

[13] Papoulis, A., and Pillai, S. U.: *Probability, Random Variables and Stochastic Processes*, fourth edition. McGraw Hill (2002).

[14] Jarzynski, C.: *Hamiltonian derivation of a detailed fluctuation theorem*. Journal of Statistical Physics **98**, 1-2: 77–102 (2000).

[15] Van den Broeck, C., and Esposito, M.: *Ensemble and Trajectory Thermodynamics: A Brief Introduction*. arXiv: 1403.1777v1 [cond-mat.stat-mech] (2014).

[16] Crooks, G. E.: *Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences*. Physical Review E **60**, 3: 2721–2726 (1999).

[17] Laplace, P. S.: *A philosophical essay on probabilities*. John Wiley & Sons, London (1902).

[18] Lanczos, C.: *The variational principles of mechanics*, fourth edition. Dover (1986).

[19] Hoover, W. G., and Hoover, C. G.: *Time Reversibility, Computer Simulation, Algorithms, Chaos*, 2nd Edition. World Scientific (2012).

[20] Keynes, J. M.: *A tract on Monetary Reform*. Macmillan and Co. (1924).

[21] Alder, B. J., and Wainwright, T. E.: *Studies in Molecular Dynamics. I. General Method*. The Journal of Chemical Physics, **31**, 2: 459–466 (1959).

[22] Born, M.: *Problems of Atomic Dynamics.* Massachusetts Institute of Technology (1926).

[23] Yoshida, H.: *Recent progress in the theory and application of symplectic integrators.* Celestial Mechanics and Dynamical Astronomy, **56**: 27–43 (1993).

[24] Yoshida, H.: *Construction of higher order symplectic integrators.* Physics Letters A, **150**, 5, 6, 7: 262–268 (1990).

[25] Yoshida, H.: *Non-existence of the modified first integral by symplectic integration methods.* Physics Letters A, **282**: 276–283 (2001).

[26] Verlet, L.: *Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules.* Physical Review, **159**: 98–103 (1967).

[27] Levesque, D., and Verlet, L.: *Molecular Dynamics and Time Reversibility.* Journal of Statistical Physics, **72**: 519–537 (1993).

[28] Lebowitz, J. L., and Penrose, O.: *Modern ergodic theory.* Physics Today **26**, Februrary 1973: 23–29 (1976).

[29] Jaynes, E. T.: *Information Theory and Statistical Mechanics.* The Physical Review, **106**, 4: 620–630 (1957).

[30] Feynman, R. P.: *Statistical Mechanics. A set of Lectures.* Westview Press (1998).

[31] Schrödinger, E.: *Statistical Thermodynamics.* Dover (1989).

[32] Carnot, N. L. S.: *Reflections on the Motive Power of Heat.* John Wiley & Sons (1897).

[33] Clausius, R.: *The Mechanical Theory of Heat, with its Applications to the Steam-Engine and to the Physical Properties of Bodies.* Taylor and Francis (1867).

[34] Meléndez, M. and Español, P.: *Gibbs-Jaynes Entropy Versus Relative Entropy*. Journal of Statistical Physics **155**, 1: 93–105 (2014).

[35] Jaynes, E. T.: *Information Theory and Statistical Mechanics. II*. The Physical Review, **108**, 2: 171–190 (1957).

[36] Fermi, E.: *Thermodynamics*. Dover (1956).

[37] Knight, D.: *This way to the Regress*. Galaxy Science Fiction, August 1956: 48-57. Galaxy Publishing Corp. (1956).

[38] Rolleri, J. L.: *Probabilidad, causalidad y explicación*. Universidad Autónoma de Querétaro (2009).

[39] Meléndez, M.: *José Luis Rolleri: Probabilidad, causalidad y explicación*. Theoria **26**, 1: 109–112 (2011).

[40] Callen, H. B.: *Thermodynamics and an Introduction to Thermostatistics. Second edition*. John Wiley & Sons (1985).

[41] Callen, H. B., and Welton, T. A.: *Irreversibility and Generalized Noise*. The Physical Review **83**, 1: 34–40 (1951).

[42] Green, M. S.: *Markoff Random Processes and the Statistical Mechanics of Time-Dependent Phenomena*. The Journal of Chemical Physics **20**, 8: 1281–1295 (1952).

[43] Kubo, R.: *Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems*. Journal of the Physical Society of Japan, **12**, 6: 570–586 (1957).

[44] Bellman, R. E.: *Dynamic Programming*. Princeton University Press (1972).

[45] Evans, D. J., Cohen, E.G.D., and Morriss, G. P.: *Probability of second law violations in shearing steady states*. Physical Review Letters **71**, 15: 2401-2404. (1993).

[46] Evans, D. J., and Searles, D. J.: *Equilibrium microstates which generate second law violating steady states.* Physical Review E **50**, 2: 1645-1648 (1994).

[47] Wang, G. M., Sevick, E. M., Mittag E., Searles, D. J., and Evans, D. J.: *Experimental Demonstration of Violations of the Second Law of Thermodynamics for Small Systems and Short Time Scales.* Physical Review Letters **89**: 050601 (2002).

[48] Crooks, G. E.: *Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems.* Journal of Statistical Physics **90**, 5/6: 1481–1487 (1998).

[49] Jarzynski, C.: *Nonequilibrium Equality for Free Energy Differences.* Physical Review Letters **78**, 14: 2690–2693 (1996).

[50] Weisstein, E. W.: *CRC Concise Encyclopedia of Mathematics.* Chapman & Hall/CRC (1999).

[51] Kawai, R., Parrondo, J. M. R., and Van den Broeck, C.: *Dissipation: The Phase-Space Perspective.* Physical Review Letters **98**, 080602 (2007).

[52] Kullback, S. and Leibler, R. A.: *On Information and Sufficiency.* The Annals of Mathematical Statistics **22**, 1: 79–86 (1951).

[53] Jaynes, E. T.: *The Gibbs Paradox*, in Smith, C. R., Erickson, G. J., and Neudorfer, P. O. (editors): *Maximum Entropy and Bayesian Methods.* Kluwer Academic Publishers, p. 1–22 (1992).

[54] Rock, P. A.: *Chemical Thermodynamics.* University Science Books (1983).

[55] Patra, P. K., Melendez, M., and Bhattacharya, B.: *Approximating the entire spectrum of nonequilibrium steady state distributions using relative entropy: An application to thermal conduction.* Forthcoming.

[56] Patra, P. K., Melendez, M., and Bhattacharya, B.: *Approximating the entire spectrum of nonequilibrium steady state distributions using relative entropy: An application to thermal conduction.* arXiv:1409.6141 [cond-mat.stat-mech] (2014).

[57] Shore, J. E., Johnson, R. W.: *Properties of Cross-Entropy Minimization.* IEEE Transactions on Information Theory, **27**, 4: 472–482 (1981).

[58] Grabert, H.: *Projection Operator Techniques in Nonequilibrium Statistical Mechanics.* Springer-Verlag (1982).

[59] Vaikuntanathan, S., and Jarzynski, C.: *Dissipation and lag in irreversible processes.* Europhysics Letters **87**, 6: 60005 (2009).

[60] Hoover, Wm. G., and Holian, B. L.: *Kinetic moments method for the canonical ensemble distribution.* Physics Letters A **211**, 5: 253–257 (1996).

[61] Augustine of Hippo: *Confessiones Sancti Patris nostri Augustini* (397-398).

[62] Hoy, R. C.: *Parmenides' Complete Rejection of Time.* The Journal of Philosophy **91**, 11: 573–598 (1994).

[63] Newton, I.: *Newton's Principia: The Mathematical Principles Of Natural Philosophy (1846)* (translated by Andrew Motte). Daniel Adee (1846).

[64] Aristotle: *Physics*, in *The Works of Aristotle.* Volume I. Encyclopædia Britannica, Inc. (1952).

[65] Zwanzig, R.: *Ensemble Method in the Theory of Irreversibility.* The Journal of Chemical Physics **33**, 5: 1338–1341 (1960).

[66] Mahajan, S.: *Street-Fighting Mathematics: The Art of Educated Guessing and Opportunistic Problem Solving.* MIT Press (2010).

[67] Zwanzig, R.: *Nonequilibrium Statistical Mechanics*. Oxford University Press (2001).

[68] Goldenfeld, N.: *Lectures on phase transitions and the renormalization group*. Perseus Books Publishing, L.L.C. (1992).

[69] Glauber, R. J.: *Time-Dependent Statistics of the Ising Model*. Journal of Mathematical Physics **4**, 2: 294–307 (1963).

[70] Pawula, R. F.: *Approximation of the Linear Boltzmann Equation by the Fokker-Planck Equation*. Physical Review **162**, 1: 186–188 (1967).

[71] Risken, H., and Vollmer, H. D.: *On the application of generalized Fokker-Planck equations*. Zeitschrift für Physik B Condensed Matter **35**, 3: 313–315 (1979).

[72] Plilmak, L. I., Olsen, M. K., Fleischhauer, M., and Collett, M. J.: *Beyond the Fokker-Planck Equation: Stochastic simulation of complete Wigner representation for the optical parametric oscillator*. Europhysics Letters **56**, 3: 372 (2001).

[73] Zwanzig, R.: *Memory Effects in Irreversible Thermodynamics*. Physical Review **124**, 4: 983–992 (1961).

[74] Melendez, M., and Español, P.: *El intercambio de energía entre sistemas Hamiltonianos* (Poster).

[75] Krauth, W.: *Statistical Mechanics. Algorithms and Computations*. Oxford University Press (2006).

[76] Gerald, C. F., and Wheatley, P. O.: *Applied numerical analysis*. Addison-Wesley (1994).

[77] Krauth, W.: *Introduction to Monte Carlo Algorithms*. arXiv:cond-mat/9612186 [cond-mat.stat-mech] (1996).

[78] Whewell, W.: *History of the Inductive Sciences. Volume 2. Book X: Thermotics and Atmology*. Appleton (1866).

[79] Eisberg, R., and Resnick, R.: *Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles*. Second Edition. John Wiley & Sons (1985).

[80] McPhie, M. G., Daivis, P. G., Snook, I. K., Ennis, J., and Evans, D. J.: *Generalized Langevin equation for nonequilibrium systems*. Physica A **299**, 3–4: 412–426 (2001).

[81] Maxwell, J. C.: *Theory of Heat*. Longman's, Green and Co. (1902).

[82] Rugh, H. H.: *Dynamical Approach to Temperature*. Physical Review Letters, **78**, 5: 772–774 (1997).

[83] Brańka, A. C., and Pieprzyk, S.: *Configurational Temperature and Monte Carlo Simulations*. Computational Methods in Science and Technology **16**, 2: 119–125 (2010).

[84] Rickayzen, G., and Powles, J. G.: *Temperature in the classical microcanonical ensemble*. The Journal of Chemical Physics **114**, 9: 4333 (2001).

[85] Butler, B. D., Ayton, G., Jepps, O. G., and Evans, D. J.: *Configurational temperature: Verification of Monte Carlo simulations*. The Journal of Chemical Physics **109**, 16: 6519 (1998).

[86] Landau, L., and Lifshitz, E.: *Curso abreviado de física teórica*. Tercera Edición. Editorial Mir (1982).

[87] Campisi, M., and Hänggi, P.: *Thermostated Hamiltonian Dynamics with Log-oscillators*. The Journal of Physical Chemistry B **117** 42: 12829-12835 (2013).

[88] von Helmholtz, H.: *Principien der Statik monocyclischer Systeme*. Journal für die reine und angewandte Mathematik **97**: 111–140 (1884).

[89] Campisi, M.: *On the mechanical foundations of thermodynamics: The generalized Helmholtz theorem*. Studies in History and Philosophy of Modern Physics **36**: 275–290 (2005).

[90] Jungnickel, C., and McCormmach, R.: *Intellectual Mastery of Nature. Theoretical Physics from Ohm to Einsten, Volume 2: The Now Mighty Theoretical Physics, 1870 to 1925*. University of Chicago Press (1990).

[91] Hoover, Wm. G., Ladd, A. J. C., and Moran, B.: *High-Strain-Rate Plastic Flow Studied via Nonequilibrium Molecular Dynamics*. Physical Review Letters **48**, 26: 1818–1820 (1982).

[92] Evans, D., Hoover, Wm. G., Failor, B. H., Moran, B., and Ladd, A. J. C.: *Nonequilibrium Molecular Dynamics via Gauss's principle of least constraint*. Physical Review A **28**, 2: 1016–1021 (1983).

[93] Hoover, Wm. G.: *Molecular Dynamics. Lecture Notes in Physics*, Volume 258. Springer-Verlag (1986).

[94] Dettmann, C. P., and Morriss, G. P.: *Hamiltonian formulation of the Gaussian isokinetic thermostat*. Physical Review E **54**, 3: 2495–2500 (1996).

[95] Hoover, Wm. G., and Hoover, C. G.: *Hamiltonian dynamics of thermostated systems: Two-temperature heat-conducting ϕ^4 chains*. The Journal of Chemical Physics **126**, 16: 164113 (2007).

[96] Nosé, S.: *A unified formulation of the constant temperature molecular dynamics methods*. The Journal of Chemical Physics **81**: 511–519 (1984).

[97] Nosé, S.: *A molecular dynamics method for simulations in the canonical ensemble*. Molecular Physics **52**, 2: 255–268 (1983).

[98] Ramsey, N. F.: *Thermodynamics and Statistical Mechanics at Negative Absolute Temperatures*. Physical Review **103**, 1: 20–28 (1956).

[99] Campisi, M., Zhan, F., Talkner, P., and Hänggi, P.: *Logarithmic Oscillators: Ideal Hamiltonian Thermostats*. Physical Review Letters **108**: 250601 (2012).

[100] Hoover, Wm. G.: *Canonical dynamics: Equilibrium phase-space distributions*. Physical Review A **31**, 3: 1695–1697 (1985).

[101] Hoover, Wm. G., and Hoover, C. G.: *Hamiltonian thermostats fail to promote heat flow*. Communications in Nonlinear Science and Numerical Simulation **18**, 12: 3365–3372 (2013).

[102] Sprott, J. C., Hoover, Wm. G., and Hoover, C. G.: *Heat conduction, and the lack thereof, in time-reversible dynamical systems: Generalised Nosé-Hoover oscillators with a temperature gradient*. Physical Review E **89**, 4: 042914 (2014).

[103] Dettmann, C. P., and Morriss, G. P.: *Hamiltonian reformulation and pairing of Lyapunov exponents for Nosé-Hoover dynamics*. Physical Review E **55**: 3693–3696 (1997).

[104] Martyna, G. J., Klein, M. L., and Tuckerman, M.: *Nosé-Hoover chains: The canonical ensemble via continuous dynamics*. The Journal of Chemical Physics **97**, 4: 2635–2643 (1992).

[105] Patra, P. K., and Bhattacharya, B.: *A deterministic thermostat for controlling temperature using all degrees of freedom*. The Journal of Chemical Physics **140**, 6: 064106 (2014).

[106] Sánchez Quesada, F., Sánchez Soto, L. L., Sancho Ruiz, M., and Santamaría Sánchez-Barriga, J.: *Fundamentos de electromagnetismo*. Editorial Síntesis (2000).

[107] Campisi, M., Zhan, F., and Hänggi, P.: *On the origin of power-laws in equilibrium*. EuroPhysics Letters **99**: 60004 (2012).

[108] Lynden-Bell, D., and Wood, R.: *The gravo-thermal catastrophe in isothermal spheres and the onset of red-giant structure for stellar systems*. Monthly Notices of the Royal Astronomical Society **138**: 495–525 (1968).

[109] Lynden-Bell, D.: *Negative specific heat in astronomy, physics and chemistry*. Physica A: Statistical Mechanics and its Applications **263**, 1: 293–304 (1999).

[110] Thirring, W.: *Systems with Negative Specific Heat*. Zeitschrift für Physik **235**, 4: 339–352 (1970).

[111] Hoover, Wm. G.: *Another Hamiltonian "Thermostat" – Comments on arXiv Contributions 1203.5968 and 1204.0312*. arXiv: 1204.0312 [cond-mat.stat-mech] (2012).

[112] Campisi, M., Zhan, F., Talkner, P., and Hänggi, P.: *Reply to W. G. Hoover [arXiv:1204.0312v2]*. arXiv:1204.4412 [cond-mat.stat-mech] (2012).

[113] Hoover, Wm. G., Posch, H. A., Holian, B. L., Gillan, M. J., Mareschal, M., Massobrio, C.: *Dissipative irreversibility from Nosé's reversible mechanics*. Molecular Simulation **1**: 79–86 (1987).

[114] Hoover, Wm. G.: *Liouville's theorems, Gibbs' entropy, and multifractal distributions for nonequilibrium steady states*. The Journal of Chemical Physics **109**, 11: 4164–4170 (1998).

[115] Meléndez, M.: *On the logarithmic oscillator as a thermostat*. arXiv:1205.3478 [cond-mat.stat-mech] (2012).

[116] Calandra, A.: *Angels on a Pin*. The Saturday Review, December 21: 60 (1968).

[117] Andersen, H. C.: *Molecular dynamics simulations at constant pressure and/or temperature*. Journal of Chemical Physics **72**, 4: 2384–2393 (1980).

[118] Aoki, K., and Kusnezov, D.: *Nonequilibrium Steady States and Transport in the Classical Lattice ϕ^4 Theory*. Physics Letters B **477**, 348–354 (2000).

[119] Meléndez, M., and Hoover, Wm. G.: *Comment on "Logarithmic Oscillators: Ideal Hamiltonian Thermostats"*. arXiv:1206.0188v2 [cond-mat.stat-mech] (2012).

[120] Campisi, M., Zhan, F., Talkner, P., and Hänngi, P.: *Reply to M. Meléndez and W. G. Hoover [arXiv:1206.0188v2]*. arXiv:1207.1859v1 [cond-mat.stat.mech] (2012).

[121] Meléndez, M., Hoover, Wm. G., Español, P.: *Comment on "Logarithmic Oscillators: Ideal Hamiltonian Thermostats"*. Physical Review Letters **110**: 028901 (2013).

[122] Campisi, M., Zhan, F., Talkner, P., and Hänggi, P.: *Campisi et al. Reply*. Physical Review Letters **110**: 028902 (2013).

[123] Bloch, I., Dalibard, J., Zwerger, W.: *Many-body physics with ultracold gases* Reviews of Modern Physics **80**: 885–964 (2008).

[124] Sponseller, D., and Blaisten-Barojas, E.: *Failure of logarithmic oscillators to serve as a thermostat for small atomic clusters*. Physical Review E **89**: 021301(R) (2014).

[125] Fermi, E., Pasta, J., and Ulam, S.: *Studies of non linear problems*. Document LA-1940 (1955)

[126] Berman, G. P., and Izrailev, F. M.: *The Fermi-Pasta-Ulam problem: 50 years of progress*. arXiv:nlin/0411062 (2005).

[127] Dauxois, T.: *Fermi, Pasta, Ulam and a mysterious lady*. Physics Today **6**, 1: 55–57 (2008).

[128] Girifalco, L. A.: *Statistical Mechanics of Solids. Monographs on the Physics and Chemistry of Materials*. Oxford University Press (2000).

[129] Horowitz, J. M., and Vaikuntanathan, S.: *Nonequilibrium detailed fluctuation theorem for repeated discrete feedback*. Physical Review E **82**: 061120 (2010).

[130] Davis, H. F.: *Fourier series and orthogonal functions*. Dover (1989).

[131] Melendez, M., and Español, P.: *The theory of coarse-graining without projected dynamics*. Forthcoming.

[132] Melendez, M., Hoover, Wm. G., and Español, P.: *Sobre el oscilador logarítmico como termostato* (*On the logarithmic oscillator as a thermostat*, Poster). XVIII Congreso de Física Estadística FisEs 2012, Palma, 18^{th}–20^{th}October (2012).

[133] Melendez, M., and Español, P.: *The Microscopic Origin of Boundary Conditions in Heat Conduction* (Poster). Fluid-Structure Interactions in Soft-Matter Systems: From the Mesoscale to the Macroscale (school cum workshop), Prato, 26^{th}–30^{th}November (2012).

[134] Meléndez, M. and Español, P.: *Gibbs-Jaynes Entropy Versus Relative Entropy*. arXiv:1402.2205 [math-ph] (2014).

[135] Melendez, M., and Español, P.: *The microscopic origin of thermal boundary conditions*. Forthcoming.

[136] Meléndez, M., and Schofield, D. (transl.): *Problemas de Dinámica Atómica* (*Problems of Atomic Dynamics*, by M. Born. Published in Hawking, S. (ed.): *Los Sueños de los que está hecha la materia* (*The Dreams that Stuff is Made of*). Crítica (2011).

[137] Meléndez, M., and Schofield, D. (transl.): *Lecciones sobre Mecánica Cuántica (Selección)* (*Excerpts from Lectures on Quantum Mechanics*), by P. A. M. Dirac. Published in Hawking, S. (ed.): *Los Sueños de los que está hecha la materia* (*The Dreams that Stuff is Made of*). Crítica (2011).

[138] Meléndez, M., and Schofield, D. (transl.): *Matemáticas Cotidianas para Dummies* (*Everyday Math for Dummies*), by Ch. Seiter, adaptation by M. Meléndez, including a new chapter on probability and statistics. Grupo Planeta (2012).

[139] Meléndez, M.: *Mandelbrot, domador de fieras matemáticas.* fronterad (2010).
http://www.fronterad.com/?q=mandelbrot-domador-fieras-matem
aticas

Appendix A

Projection Operators

The theory of linear spaces [130] states that we can represent the projection of a function A onto the subspace spanned by a set of mutually orthogonal functions, $\{\varphi_m\}$, by the action of a projection operator P applied to A,

$$\mathrm{P}A = \sum_m \frac{\langle A,\ \varphi_m \rangle}{\|\varphi_m\|}\ \varphi_m. \tag{A.1}$$

The angle brackets indicate the inner product, and $\|\varphi_m\| = \langle \varphi_m,\ \varphi_m \rangle$. The projection onto the orthogonal subspace results from the action of the complementary operator $\mathrm{Q} = \mathrm{I} - \mathrm{P}$, where we have indicated the identity transformation with I. In the general case in which the functions φ_m are not orthogonal [67],

$$\mathrm{P}A = \Big(\langle A,\ \varphi \rangle \left(\langle \varphi,\ \varphi \rangle \right)^{-1} \Big) \cdot \varphi. \tag{A.2}$$

We use $\langle A,\ \varphi \rangle$ to denote a row vector with components $\langle A,\ \varphi_j \rangle$. The $\langle \varphi,\ \varphi \rangle$ bracket refers to the $k \times k$ square matrix with components $\langle \varphi_j,\ \varphi_m \rangle$. Any linear operator L separates into a projected and an orthogonal part,

$$\mathrm{L} = \mathrm{PL} + \mathrm{QL}. \tag{A.3}$$

In Chapter 3, we treated the probability density as a time-dependent phase function. This point of view, known as the *Schrödinger picture*, has a dual representation, the so-called *Heisenberg picture* [58, 67], that takes the probability density as a fixed function and considers that the relevant phase functions change in time according to

$$\frac{\partial F}{\partial t} = \hat{L}\mathrm{i}F. \tag{A.4}$$

We write the formal solution of this equation as

$$F(t) = e^{\hat{L}\mathrm{i}t}F, \tag{A.5}$$

with $F(t)$ representing $F(z;\ t)$ and A standing for $F(z) = F(z;\ 0)$. The Hermitian character of the Liouvillian operator constitutes a manifestation of the dual relation between the two points of view, $\mathrm{Tr}[\rho\ \hat{L}\mathrm{i}F] = -\mathrm{Tr}[(\hat{L}\mathrm{i}\rho)F]$.

We wish to rewrite the time evolution of $F(t)$ as a sum of a projected (relevant) part and an orthogonal part. We view the time evolution operator as a composition of two transformations

$$e^{\hat{L}\mathrm{i}t} = e^{(\mathrm{Q}+\mathrm{P})\hat{L}\mathrm{i}t} = e^{\mathrm{Q}\hat{L}\mathrm{i}t}e^{\mathrm{P}\hat{L}\mathrm{i}t}, \tag{A.6}$$

first a projected evolution in the subspace spanned by F, and then a transformation in the orthogonal direction in phase space. In this context, we view the evolution of our phase functions as if they were contained in the projected subspace, and we interpret the effects in the orthogonal direction as the action of a stochastic force, $\psi(t)$,

$$\psi(t) = e^{\mathrm{Q}\hat{L}\mathrm{i}t}\mathrm{Q}\hat{L}\mathrm{i}F. \tag{A.7}$$

The time derivative of (A.5) is

$$\frac{\partial}{\partial t}F(t) = e^{\hat{L}\mathrm{i}t}\hat{L}\mathrm{i}F. \tag{A.8}$$

We break the Liouvillian into projected and orthogonal parts with (A.3).

$$\frac{\partial}{\partial t}F(t) = e^{\hat{L}\mathrm{i}t}\mathrm{P}\hat{L}\mathrm{i}F + e^{\hat{L}\mathrm{i}t}\mathrm{Q}\hat{L}\mathrm{i}F. \tag{A.9}$$

The projection operators do not depend on time, so we use the properties of the exponential of $\hat{L}it$ to rewrite the first term on the right as

$$e^{\hat{L}\mathrm{i}t}\mathrm{P}\hat{L}\mathrm{i}F = \mathrm{P}e^{\hat{L}\mathrm{i}t}\hat{L}\mathrm{i}F = \mathrm{P}\hat{L}\mathrm{i}e^{\hat{L}\mathrm{i}t}F = \mathrm{P}\hat{L}\mathrm{i}F(t). \tag{A.10}$$

The second term we transform with a decomposition of the exponential operator,

$$e^{\hat{L}\mathrm{i}t} = e^{\mathrm{Q}\hat{L}\mathrm{i}t} + \int_0^t e^{\hat{L}\mathrm{i}s}\mathrm{P}\hat{L}\mathrm{i}e^{\mathrm{Q}\hat{L}\mathrm{i}(t-s)}ds, \tag{A.11}$$

which can be checked directly by (clever) differentiation. Now we can apply (A.11) to the last term in (A.9).

$$e^{\hat{L}\mathrm{i}t}\mathrm{Q}\hat{L}\mathrm{i}F = e^{\mathrm{Q}\hat{L}\mathrm{i}t}\mathrm{Q}F + \int_0^t e^{\hat{L}\mathrm{i}s}\mathrm{P}\hat{L}\mathrm{i}e^{\mathrm{Q}\hat{L}\mathrm{i}(t-s)}\mathrm{Q}F. \tag{A.12}$$

Together with the definition of $\psi(t)$ (A.7), and (A.10), the equation above transforms (A.9) into

$$\frac{\partial}{\partial t}F(t) = \mathrm{P}\hat{L}\mathrm{i}F(t) + \int_0^t e^{\hat{L}\mathrm{i}s}\mathrm{P}\hat{L}\mathrm{i}\psi(t-s)\ ds + \psi(t), \tag{A.13}$$

representing the effect of the projected evolution, the accumulated effect of the stochastic force, and its contribution at time t.

We have expressed (A.4) in the new mathematical form (A.13), but we have yet to specify the meaning of the inner products in the definition of P. If we wish (A.13) to behave like a generalised Langevin equation, then the stochastic force should vanish when averaged over the relevant distribution. By an appropriate choice of the inner product, we can make this average vanish automatically[1].

$$\langle A,\ B\rangle = \mathrm{Tr}[\rho^{eq}A\ B] \tag{A.14}$$

[1]We will not justify this choice here. The argument is not difficult to understand, but it is a bit too lengthy for this appendix. See Zwanzig's *Nonequilibrium statistical mechanics*, Chapter 8 [67].

makes $\mathrm{Tr}[\bar{\rho}\ \psi(t)] = 0$.

We average (A.13) over the relevant distribution to obtain a closed equation for the expected value $f = \mathrm{Tr}[\bar{\rho}\ F]$.

$$\frac{\partial f}{\partial t} = \left(\left\langle \hat{L}\mathrm{i}F,\ F\right\rangle (\langle F,\ F\rangle)^{-1}\right)\cdot F + \mathrm{Tr}\left[\bar{\rho}\left(\left\langle \hat{L}\mathrm{i}\psi(t-s),\ F\right\rangle (\langle F,\ F\rangle)^{-1}\right)\cdot F(s)\right]. \quad \text{(A.15)}$$

We did not arrive at our own exact equation for the time evolution of f (3.32), derived without reference to projection operators, by simply rewriting (A.15). Of course, being exact expressions for $\partial f/\partial t$, they must state the same proposition with different symbols, in a sense. However, (3.32) comes from an alternative decomposition of the time evolution operator.

Let $\mathcal{P}^\dagger$ be the name of the projection operator that transforms a probability density function into the relevant distribution,

$$\mathcal{P}^\dagger\rho = \bar{\rho}. \quad \text{(A.16)}$$

Accordingly, $\mathcal{Q}^\dagger = \mathrm{I} - \mathcal{P}^\dagger$. We will express the time evolution of $\mathcal{Q}^\dagger\rho$ in terms of $\mathcal{P}^\dagger\rho$.

$$\frac{\partial}{\partial t}\mathcal{Q}^\dagger\rho = \frac{\partial}{\partial t}(\mathrm{I}-\mathcal{P}^\dagger)\rho = -\hat{L}\mathrm{i}\mathcal{P}^\dagger\rho - \hat{L}\mathrm{i}\mathcal{Q}^\dagger\rho - \frac{\partial}{\partial t}\mathcal{P}^\dagger\rho. \quad \text{(A.17)}$$

The last term on the right represents the time rate of change of the relevant distribution, that is,

$$\frac{\partial\bar{\rho}}{\partial t} = \frac{\partial}{\partial t}\mathcal{P}^\dagger\rho = \lim_{\tau\to 0}\frac{\mathcal{P}^\dagger(\rho - \tau\hat{L}\mathrm{i}\rho) - \mathcal{P}^\dagger\rho}{\tau} = -\mathcal{P}^\dagger\hat{L}\mathrm{i}\rho. \quad \text{(A.18)}$$

We get a differential equation for $\mathcal{Q}^\dagger\rho$,

$$\frac{\partial}{\partial t}\mathcal{Q}^\dagger\rho = -\hat{L}\mathrm{i}\mathcal{P}^\dagger\rho - \hat{L}\mathrm{i}\mathcal{Q}^\dagger\rho - \mathcal{P}^\dagger\hat{L}\mathrm{i}\rho, \quad \text{(A.19)}$$

that we can solve formally,

$$\mathcal{Q}^\dagger \rho = e^{\hat{L}\mathrm{i}t}\mathcal{Q}^\dagger\rho(0) + \int_0^t e^{\hat{L}\mathrm{i}(t-s)}[\hat{L}\mathrm{i},\ \mathcal{P}^\dagger]\ \rho\ ds. \tag{A.20}$$

The square brackets stand for the commutator,

$$[\hat{L}\mathrm{i},\ \mathcal{P}^\dagger] = \hat{L}\mathrm{i}\mathcal{P}^\dagger - \mathcal{P}^\dagger\hat{L}\mathrm{i}. \tag{A.21}$$

as in Chapter 3, we assume that $\mathcal{Q}^\dagger\rho(0) = \delta\rho(0) = 0$. Now we can interpret (A.20) as a definition of the action of $\mathcal{Q}^\dagger$ on a distribution. Therefore, we can decompose the time evolution operator into a relevant and an orthogonal part.

$$\begin{aligned} e^{-\hat{L}\mathrm{i}t} &= e^{-\hat{L}\mathrm{i}t}\mathcal{P}^\dagger + e^{-\hat{L}\mathrm{i}t}\mathcal{Q}^\dagger \\ &= e^{-\hat{L}\mathrm{i}t}\mathcal{P}^\dagger + e^{-\hat{L}\mathrm{i}t}\int_0^t e^{-\hat{L}\mathrm{i}(t-s)}[\hat{L}\mathrm{i},\ \mathcal{P}^\dagger]\ ds. \end{aligned} \tag{A.22}$$

The projection and time evolution operators above act on distributions, corresponding to the Schrödinger picture, while Zwanzig's decomposition (A.11) applies to the Heisenberg picture. Pep Español has shown that the dual of (A.22) reads[2]

$$e^{\hat{L}\mathrm{i}t} = e^{\hat{L}\mathrm{i}t}\mathrm{P} + \mathrm{Q}e^{\hat{L}\mathrm{i}t} + \int_0^t e^{\hat{L}\mathrm{i}s}[\mathrm{P},\ \hat{L}\mathrm{i}]e^{\hat{L}\mathrm{i}(t-s)}\ ds \tag{A.23}$$

for operators in the Heisenberg picture (compare (A.23) to (A.11)).

The operator identities (A.22) and (A.23) generate a theory of coarse-graining equivalent to our third Chapter, but in the language of projection operators. In contrast to Zwanzig's result, though, our decompositions of the time evolution operators do not include projection operators in the exponents. This makes the expressions functions of the real dynamics, which presumably makes it easier to calculate them numerically from simulations.

[2]See our forthcoming paper [131]

The logic in going from (A.16) to (A.22) does not rely on the specific properties of the Liouvillian. It only uses the fact that

$$\frac{\partial \rho}{\partial t} = -\hat{L}\mathrm{i}\rho. \tag{A.24}$$

Therefore, we can replace the Liouvillian with the more general operator L (3.126) defined in Section 3.8, and the decomposition (A.22) will still work.

This research project has emphasised memoryless macroscopic laws, that is, laws expressed as partial differential equations. We saw how to derive such equations for slow macroscopic variables. Nonetheless, Zwanzig's original motivation also included the study of memory effects from the point of view of irreversible thermodynamics [73]. It was very difficult to analyse these effects in part because we had to work out what the projected dynamics look like. We hope our new equations in terms of the real dynamics will contribute to simplifying the task.

Appendix B

Contributions

This thesis contains some previously unpublished results, including the main contribution in Chapter 3. We have already presented others as peer-reviewed articles:

[**121**] Meléndez, M., Hoover, Wm. G., and Español, P.: *Comment on "Logarithmic Oscillators: Ideal Hamiltonian Thermostats"*. Physical Review Letters **110**: 028901 (2013).

[**34**] Meléndez, M. and Español, P.: *Gibbs-Jaynes Entropy Versus Relative Entropy*. Journal of Statistical Physics **155**, 1: 93–105 (2014).

Currently, we are preparing the following manuscripts:

[**55**] Patra, P. K., Melendez, M., and Bhattacharya, B.: *Approximating the entire spectrum of nonequilibrium steady state distributions using relative entropy: An application to thermal conduction*. Forthcoming.

[**131**] Melendez, M., and Español, P.: *The theory of coarse-graining without projected dynamics*. Forthcoming.

[**135**] Melendez, M., and Español, P.: *The microscopic origin of thermal boundary conditions*. Forthcoming.

In addition, we worked on three posters which were accepted and presented at conferences.

[74] Melendez, M., and Español, P.: *El intercambio de energía entre sistemas Hamiltonianos* (*The transfer of energy among Hamiltonian systems*, Poster). XVIII Congreso de Física Estadística FisEs 2012, Palma, 18^{th}–20^{th}October (2012).

[132] Melendez, M., Hoover, Wm. G., and Español, P.: *Sobre el oscilador logarítmico como termostato* (*On the logarithmic oscillator as a thermostat*, Poster). XVIII Congreso de Física Estadística FisEs 2012, Palma, 18^{th}–20^{th}October (2012).

[133] Melendez, M., and Español, P.: *The Microscopic Origin of Boundary Conditions in Heat Conduction* (Poster). Fluid-Structure Interactions in Soft-Matter Systems: From the Mesoscale to the Macroscale (school cum workshop), Prato, 26^{th}–30^{th}November (2012).

Furthermore, the reader may find additional details in our arXiv contributions.

[115] Meléndez, M.: *On the logarithmic oscillator as a thermostat.* arXiv:-1205.3478 [cond-mat.stat-mech] (2012).

[119] Meléndez, M., and Hoover, Wm. G.: *Comment on "Logarithmic Oscillators: Ideal Hamiltonian Thermostats".* arXiv:1206.0188 [cond-mat.stat-mech] (2012).

[134] Meléndez, M. and Español, P.: *Gibbs-Jaynes Entropy Versus Relative Entropy.* arXiv:1402.2205 [math-ph] (2014).

[56] Patra, P. K., Melendez, M., and Bhattacharya, B.: *Approximating the entire spectrum of nonequilibrium steady state distributions using relative entropy: An application to thermal conduction.* arXiv:1409.6141 [cond-mat.stat-mech] (2014).

During the research project, I was invited to present my findings on three occasions.

- Meléndez, M.: *On the logarithmic oscillator as a thermostat.* Group meeting. Departamento de Física Atómica, Molecular y Nuclear (Universidad Complutense de Madrid). 12 December (2012).

- Meléndez, M.: *Conducción térmica y el termostato logarítmico.* Seminario. Departamento de Fśica Fundamental (Universidad Nacional de Educación a Distancia). 23 January (2013).

- Meléndez, M.: *Gibbs-Jaynes Entropy versus Relative Entropy.* Group meeting. Departamento de Física Atómica, Molecular y Nuclear (Universidad Complutense de Madrid). 5 March (2014).

Finally, I also worked on translations, talks and articles in the fields of popular science and philosophy of science which are not directly relevant to the contents of this thesis (See Refs. [136, 137, 138, 139, 39]).

List of Figures and Tables

List of Symbols

Δ	finite difference
ΔS	relative entropy, page 73
Γ	region of phase space
$\Gamma(s)$	Euler's gamma function
Λ	phase space compression factor
Ω	number (or density) of states
Ψ_f	$\prod_i \delta(F_i(z) - f_i)$
$\bar{\rho}$	relevant part of ρ
δA	variation of A
$\delta\rho$	$\rho - \bar{\rho}$
$\delta(x)$	Dirac delta function
δ_{ij}	Kronecker delta function
$\epsilon,\ \sigma$	Lennrad-Jones potential parameters
$\eta,\ \xi$	friction variables
$\lambda,\ \lambda_1,\ \lambda_2,\ \ldots$	Lagrange multipliers, parameters

$\omega_n,\ \omega_\varkappa$ angular frequencies

$\psi(t)$ stochastic force

$\psi_{i,n}(q)$ wave function

$\rho,\ \rho',\ \rho^0$ probability distributions

ρ^{eq} equilibrium ensemble

τ time step

$\varkappa$ harmonic oscillator spring constant

ζ damping constant

$\{u,\ v\}$ Poisson bracket of u and v

A action, page 18

$\mathrm{C}_{\bar{\rho}}$ time correlation function, page 108

D Kullback-Leibler divergence

$d,\ đ$ exact and inexact differentials

$D_{ij}(f)$ diffusion coefficient

$\mathbf{e}$ vector of energy values, $(e_1,\ e_2,\ \ldots,\ e_k)$, page 113

$\langle F \rangle$ expected value of F

$\mathbf{F}_i$ force on the ith particle

$\mathcal{F}$ free energy, page 66

$F,\ F_1,\ F_2,\ \ldots$ phase functions

$f,\ f_1,\ f_2,\ \ldots$ expected values of $F,\ F_1,\ F_2,\ \ldots$

FPUT Fermi-Pasta-Ulam-Tsingou, page 187

G virial, page 171

$G(q,\ p)$ generating function, page 28

H Boltzmann's H function

$H,\ E_i$ Hamiltonian functions

$\langle I \rangle$ mutual information

i imaginary unit, $\sqrt{-1}$

I dynamical invariant

$i,\ j,\ k,\ l,\ N,\ M$ integers

$\mathrm{Int}(x)$ integer part of x, page 31

J_T Jacobian matrix for T

$k,\ \kappa$ FPUT chain parameters

$\hat{L}\mathrm{i}$ Liouvillian operator, page 104

L compressible flow operator, page 145

$m(z)$ measure

$m_1,\ m_2,\ \ldots$ masses, or number of measurements

P, Q, $\mathcal{P}^\dagger,\ \mathcal{Q}^\dagger$ projection operators, page 217

$p(z,\ t)$ probability of state z at time t

$P_F(f;\ t)$ probability of $F(z) = f$ at time t

p_{ji} transition probability from z_i to z_j

$(q,\ p)$ coordinates and momenta

$\dot{q},\ \dot{p}$ time derivatives of $q,\ p$

Q heat

r distance

S entropy functional, page 10

S_B Boltzmann entropy, page 12

S_H Hertz entropy, page 160

S_S Shannon entropy, page 11

T temperature

$T(z),\ \tilde{T}(z)$ direct and reverse transformations

$t,\ t_1,\ t_2,\ \ldots$ times

T_k kinetic energy

Tr trace, page 37

V potential energy

$v(f)$ flux density, page 87

$v,\ \mathbf{v}_i,\ v_i$ velocities

$v_i(f)$ organised drift, page 104

W work

W_{diss} dissipated work, page 71

$x,\ x'$ positions

$\tilde{z}$ $(q,\ -p)$, reverse of $z = (q,\ p)$, page 18

Z partition function

$z,\ z',\ z_1,\ z_2,\ \ldots$ microstates

Index