



SAPIENZA
UNIVERSITÀ DI ROMA

Statistical Mechanics and Thermodynamics of Small Systems

Scuola di dottorato Vito Volterra
Dottorato di Ricerca in Fisica – XXIX Ciclo

Candidate
Luca Cerino
ID number 1233203

Thesis Advisors
Prof. Angelo Vulpiani
Dott. Andrea Puglisi

A thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics

Ottobre 2016

Thesis defended on February 16, 2016
in front of a Board of Examiners composed by:
Prof. Mauro Carfora (chairman)
Prof. Sergio Caracciolo
Prof. Antonello Scardicchio

Statistical Mechanics and Thermodynamics of Small Systems

Ph.D. thesis. Sapienza – University of Rome

© 2016 Luca Cerino. All rights reserved

This thesis has been typeset by L^AT_EX and the Sapthesis class.

Version: February 14, 2017

Author's email: luca.cerino@roma1.infn.it, cerino.luca@gmail.com

*Dedicata ad
Ilaria, mia compagna di vita.*

Abstract

In this thesis many aspects of the statistical mechanics and thermodynamics of small systems are studied. The very same possibility of defining a thermodynamics for this class of systems, for which the usual properties of the thermodynamic limit do not apply, is discussed by means of general considerations and specific examples. We show that it is possible to preserve most of the features of thermodynamics for a specific class of systems which are, at the same time, far enough from the infinite- N limit to be small, but large enough to be studied with a statistical approach. A review of the necessary mathematical and physical tools to study this particular class of systems is included.

Eventually, a specific system is studied, both from an equilibrium and a non-equilibrium perspective: it is found that this system, composed by a gas included in a container with a moving wall (the piston), has an highly non-trivial dynamics caused by the interplay of the different degrees of freedom of the system, which cannot be easily reproduced by means of coarse-grained equations. At the same time, the smallness of the system is responsible for large fluctuations that strongly characterize the system. We show that this system reproduces the behavior of an heat engine, when the external parameters vary in time: in particular we show that different working regimes (engine, refrigerator, heat pump) can be obtained depending upon the total time of a cycle of the external parameters. We also derive some analytical results reproducing, with a fair degree of approximation, the behavior of the system.

Ringraziamenti

Questa tesi è il frutto del lavoro che ho svolto in tre anni molto intensi della mia vita in cui il lavoro nel campo della fisica è stato indissolubilmente legato ad emozioni, progetti, piani, entusiasmi, illusioni e delusioni che si sono succeduti a ritmo quasi incessante.

La lista delle persone che devo ringraziare è pressoché infinita: comincerò da tutti i colleghi del gruppo TNT che in questi anni mi sono stati vicini in mille modi diversi e che si sono sempre interessati al mio lavoro e alla mia vita. Le chiacchierate, i pranzi, le occasioni speciali da festeggiare, le discussioni sugli esperimenti, le divergenze e/o convergenze politiche, filosofiche e religiose mi hanno fatto crescere e vi sarò sempre grato per questo. In particolar modo vorrei ringraziare Umberto Bettolo, Fabio Cecconi, Massimo Cencini, Massimo Falcioni, Andrea Gnoli, Giorgio Pontuale, Alberto Petri e Alessandro Sarracino per avermi mostrato la passione e l'impegno con cui si può fare questo lavoro. Un pensiero speciale va poi a coloro che, per motivi anagrafici, sono stati i miei più prossimi compagni di viaggio, Antonio Lasanta e Alessandro Manacorda. Un grazie speciale lo devo anche a Dario Villamaina, che, anche se a distanza, si è sempre prodigato in consigli e suggerimenti e con cui rimpiango di non aver potuto lavorare più a stretto contatto.

Non credo di essere in grado di poter esprimere in queste poche righe la gratitudine che provo nei confronti di Angelo ed Andrea: voglio ringraziarli non solo per avermi insegnato il 99,9% della fisica contenuta in questa tesi, ma anche per avermi trasmesso la loro filosofia e il loro modo di vedere il mondo, e per avermi permesso di lavorare su argomenti così stimolanti e interessanti, con una libertà ed un'attenzione che non credo si possa trovare in molti altri gruppi di ricerca.

Un pensiero lo devo anche ai vari compagni che si sono succeduti nella stanza Pi Greco, dove molto del lavoro contenuto in questa tesi è stato fatto: in particolare con Manuel, Martina, Jacopo, Carlo e Matteo abbiamo condiviso tutta la fatica, le frustrazioni, le difficoltà, le preoccupazioni e le soddisfazioni che derivano dal lavoro del ricercatore. Ma soprattutto abbiamo perso tempo in maniera davvero spettacolare.

Un grandissimo ringraziamento lo devo ai miei genitori ed al resto della mia famiglia, che hanno sempre apprezzato il mio lavoro: grazie per aver sempre provato a leggere i miei articoli!

Infine il ringraziamento più importante di tutti ad Ilaria: che, come sempre, mi dona felicità e mi sprona sempre a dare il meglio di me; e che sono sicuro sarà per sempre al mio fianco nel lungo e bellissimo cammino che abbiamo davanti.

Contents

Introduction	xi
I General formulation of the problem	1
1 Foundations of the Statistical Mechanics of Small Systems	3
1.1 The ergodic problem	4
1.2 Thermodynamic Limit and Ensemble Equivalence	10
1.2.1 Microcanonical ensemble	11
1.2.2 Canonical ensemble and equivalence of ensembles	12
1.3 Macroscopic Irreversibility	14
1.3.1 The Boltzmann interpretation of macroscopic irreversibility	15
1.3.2 Can small systems be irreversible?	17
1.3.3 Small systems with an irreversible behavior	20
1.4 Isolated small systems and the problem of the definition of temperature	22
1.4.1 Second law and energy flux between two systems in contact	25
1.4.2 Subsystems	26
1.4.3 The generalised Maxwell-Boltzmann distribution	27
1.4.4 How to measure T_B and T_G	28
1.4.5 Equivalence of ensembles and the equipartition formula	29
1.4.6 Can single-particle “thermodynamics” rule out the Boltzmann temperature?	30
1.5 Conclusions	30
2 Irreversible and Stochastic Thermodynamics: a Brief Review	33
2.1 On the definition of microscopic work and heat	34
2.2 Stochastic Thermodynamics	37
2.2.1 Work and Heat in the SDE formalism	38
2.2.2 Work and heat in the Fokker-Planck formalism	39
2.2.3 Connection with standard thermodynamics and entropy production	40
2.3 Heat engines and linear irreversible thermodynamics	43
2.3.1 Linear regime and Onsager reciprocity relations	46
2.3.2 Efficiency at maximum power	49
2.4 Conclusions	49
2.A Fluctuation theorems and Jarzynski Equality	50

2.A.1	Isolated Hamiltonian systems	52
2.A.2	Thermostatted Hamiltonian systems	53
2.A.3	Thermostatted Hamiltonian systems with a time-dependent temperature	54
II	A case of study: the gas-piston engine prototype	55
3	Equilibrium description and coarse-grained equations of motion	57
3.1	Introduction	57
3.2	The model	58
3.3	Statistical Equilibrium Description	59
3.3.1	Microcanonical	60
3.3.2	Canonical	61
3.3.3	Comparison between microcanonical and canonical ensemble	62
3.4	The interacting case	63
3.5	Langevin equation	67
3.6	Conclusions	68
3.A	Derivation of the Langevin Equation	70
4	The piston engine	73
4.1	Introduction	73
4.2	The Mechanical Model (MM)	74
4.2.1	The engine protocol	75
4.3	The “phase diagram” of the engine	78
4.3.1	Average work and heats as functions of τ	78
4.3.2	Fluctuations	79
4.4	Coarse-grained description	79
4.5	Analytic results in the two-variables (2V) model	81
4.6	Linear regime	83
4.6.1	Analytic expression of the Onsager coefficients in the 2V model	86
4.7	Efficiency at maximum power: linear regime and beyond	87
4.7.1	Linear regime	88
4.7.2	Non-linear regime	90
4.8	An analytic study of fluctuations by means of Large Deviation Theory	92
4.9	Conclusions	95
4.A	Details on the derivation of the Langevin Equation	96
4.B	Details on the analytic solution of the 2V model	98

Introduction

In the last decades an impressive technological progress led to the introduction of techniques and machineries allowing to observe, control and even manipulate systems on the nano/micro-scale, with characteristic size of $1 \text{ nm} \div 1 \mu\text{m}$: typical examples are biological macromolecules operating inside cells.

Many studies that analyze the folding/unfolding cycle of proteins and/or nucleic acids [69], or the performance of molecules like kinesin [14], a molecule that moves on a microtubule and transports other molecules inside a cell, insist on the importance of being “small”: these systems, able to convert the temperature fluctuations of the environment into work and movement, are intrinsically small and can only operate on the nano-scale. Furthermore their behavior is highly non-trivial, displaying unusual dynamics and non-gaussian fluctuations [6, 61, 72, 93].

Their macroscopic counterpart is represented by granular systems, i.e. systems composed by a small number ($100 \div 1000$) of macroscopic spheres with a diameter of few millimeters: indeed, in such a case we are still in the presence of a small system, since the macroscopicity of the system (i.e. a large number of atoms and molecules) is compensated by a small number of constituents, the beads. It is possible to design setups of vibrated granular gases at low density, the vibration reintroducing in the system the energy dissipated in the inelastic collisions, where fluctuations can be rectified into movement in a definite direction [82, 45, 89]. All these systems constitute examples of small engines, i.e. systems with few degrees of freedom that are able to convert thermal energy and temperature fluctuations into useful work.

The advances obtained in experiments pushed towards a theoretical reconsideration of the physics of small systems: in fact, even though a statistical mechanical approach seems necessary in order to understand the physics of this class of systems, it is well known that the most relevant predictions of statistical mechanics (equilibrium equation of states, phase transitions etc. . .) are only recovered in the thermodynamic limit, $N \rightarrow \infty$. In fact, it is immediate to understand that, even at equilibrium, when the number of the components of a system is small, fluctuations of collective observables, i.e. depending upon all the degrees of freedom of the system, have wild fluctuations around their average values. For this reason the predictive ability of equilibrium results is deeply weakened when the considered system is small: the intrinsic meaning of equilibrium should be rethought in this context.

Moreover, most of the system we are interested in operate in non-equilibrium steady states, e.g. by exploiting a temperature gradient in order to move [20], or with non-constant external parameters: these conditions are necessary in order to observe systems that transform thermal energy into work. Standard thermodynamics and statistical mechanics tools are not appropriate to describe and reproduce the physics

of heat engines: for this reason, it is necessary to consider new methodologies. As an important example we mention the progresses in the study of fluctuations and their relation with response functions [71].

At a difference with standard thermodynamics, it is clear that, in this regime, transformations occur over a finite interval of time: for this reason the common assumption of quasi-static, or adiabatic, transformations is not appropriate for systems operating at these scales. This problem has been frequently discussed in the recent past, also from an engineering point of view, giving birth to the so-called finite time thermodynamics [28, 2, 99], which focuses on the study of engines working at finite power, i.e. far from the maximum (adiabatic) Carnot efficiency.

The other important difference with respect to the standard setups is, of course, the smallness of the system: from this point of view a prototypical example is represented by the brownian motion, a single colloidal particle moving in a fluid under the influence of random collisions from the molecules of the surrounding liquid. The study of the brownian motion, since the seminal works of Perrin [78], Einstein [34] and Smochulowski [104], covered a central role in the development of the physics of small and stochastic systems. A step forward is represented by stochastic thermodynamics, based on the idea that the thermodynamical concepts of work, heat and entropy can consistently be extended to a single trajectory of a brownian particle [92, 94]: by means of simple stochastic tools is possible to re-define the principles of thermodynamics at the level of the stochastic evolution of single brownian particles moving in time-dependent potentials and with non-constant bath temperatures.

The results of stochastic thermodynamics strongly rely upon the so-called fluctuation theorems [42, 66, 71], i.e. a very wide class of equalities, explaining the true, probabilistic, meaning of the second principle of thermodynamics: indeed, in great generality and under some rather broad hypothesis, it is possible to quantify the ratio of the probability of measuring a given variation of entropy and the probability of measuring the opposite value. Of course, when the system is very large, it is infinitely much more probable to measure an increasing entropy than a decreasing one: on a small scale the probabilities become comparable and the second principle of thermodynamics is only a statement on their averages.

The aim of this Thesis is to combine all the different aspects of the physics of small systems in order to explore and investigate new frontiers of the discipline that have not been analyzed in the last years. In particular, one of our *desiderata* is to extend the considerations of stochastic thermodynamics, that only apply to single brownian particles, to much more complicated small systems, composed by a number of particles larger than one. In order to do this we will study a model, analog to the usual gas-piston system usually encountered in thermodynamics, in which many non-trivial aspects can be observed and analyzed.

The Thesis is divided in two parts:

- In **Part I** we introduce the general aspects of the physics of classical small systems from a Statistical Mechanics point of view. We discuss some rather general topics and introduce most of the tools and instruments that will allow the reader to understand the delicate points, which will be critically discussed.

-
- In **Chapter 1** the possibility of making thermodynamics out of small systems will be questioned. In particular, the foundations of Statistical Mechanics, allowing a connection between the microscopic realm and the everyday life, will be revisited; furthermore a discussion on how these considerations can be extended to systems with a small number of degrees of freedom will be carried on. We will discuss in detail, from a small system perspective, the following topics: the relevance of the ergodic theorem in the context of Statistical Mechanics; how is it possible to obtain irreversible behaviors out of reversible dynamics; what is the correct definition of microscopic entropy. This Chapter contains some of the results already published in Ref. [Cerino et al., 2015a, Cerino et al., 2016a].
 - In **Chapter 2** we review the fundamental aspects of stochastic and irreversible thermodynamics: most of the tools needed to perform an energetic analysis of nonequilibrium processes will be introduced here. We will carry on a deepened discussion on how it is possible to define work and heat in a microscopic context, and we will show some explicit formulas, based on the above considerations, in the context of stochastic processes (underdamped uni-dimensional Langevin equation). Eventually, we will analyze the connections between equilibrium and non equilibrium definitions of entropy and, in the end, we will define the most relevant figures of merit characterizing the performance of an heat engine.
 - In **Part 2** most of the concepts introduced in the first part of the Thesis will be used and applied to a specific model, a gas included in a container with a moving wall, a piston. By letting some external parameters to vary in time we will be able to obtain an heat engine, i.e. a system converting thermal energy extracted from a thermostat into useful mechanical energy.
 - In **Chapter 3** a two-dimensional version of the model will be introduced and its statistical equilibrium properties will be studied and discussed. A peculiar aspect of the system to be deepened is the difference between the fluctuations of the position of the piston when the system is isolated or when it is attached to a constant temperature thermostat. We will show that this inequality corresponds to a non-trivial dynamic that, in addition, is different in the constant energy and in the constant temperature regime: by means of a coarse grained Langevin equation we will try to explain this interesting behavior and explore the different regimes. This Chapter contains most of the results already published in Ref. [Cerino et al., 2014].
 - In **Chapter 4** a uni-dimensional simplification of the model introduced in the previous Chapter will be discussed. Our principal aim is to investigate the behavior of the system when the temperature of the thermostat and the external pressure are varied in time with a cyclical protocol. We will see that, depending on the velocity with which this parameters are varied in time, the system may either act as an engine (i.e. transforming heat into work) or as a refrigerator (i.e. consuming work and transferring heat from a cold source to a hot sink). By means of an improved Langevin equation we will try to understand this behavior

with analytical predictions. This Chapter contains the results already published in Ref. [Cerino et al., 2015b] and [Cerino et al., 2016b].

Part I

General formulation of the problem

Chapter 1

Foundations of the Statistical Mechanics of Small Systems

At its origins, Statistical Mechanics was conceived as a tool to derive and justify thermodynamics from the atomistic hypothesis, i.e. from the assumption that all systems, e.g. gases, are made of a humongous number of elementary particles, the atoms. At the beginning of the twentieth century such an hypothesis was proved by the pioneering works of A. Einstein (who found a connection between the diffusion coefficient of Brownian particles and the Avogadro number \mathcal{N}_A) and J. Perrin (who measured the value of \mathcal{N}_A with extreme precision). After this success, Statistical Mechanics proved to be a very powerful theory able to interpret and explain most of the features of the physics of macroscopic systems, as the irreversible evolution towards equilibrium or the phenomena occurring in the proximity of a phase transition. Most of this results rely on the assumption that the systems we are dealing with are composed by a very large number of elementary constituents (of the order of magnitude of \mathcal{N}_A): from a mathematical point of view this consists in taking the thermodynamic limit $N \rightarrow \infty$. In fact, even if every physical system is composed of a finite number of particles, such a number, for macroscopic systems, is so large, $\mathcal{O}(10^{23})$, that the thermodynamic limit is the most effective representation of these systems.

In the recent past, the development of experimental techniques able to observe and even manipulate systems on mesoscopic scales shifted the attention of part of the scientific community towards the physics of systems composed by a small number of elementary constituents. In fact it is now possible to build and then control mesoscopic particles, i.e on the scale of micrometers up to tenths of millimeters, in which fluctuations, an important component of statistical mechanics that can not be observed on macroscopic systems, can be directly measured.

Many results, mostly numerical, obtained in the last decades proved that the tools and the concepts introduced in statistical mechanics can predict with success the behaviour of small systems. This new range of applicability exceeds the original intent of the theory, i.e. the justification of thermodynamics from the atomic hypothesis, and introduces new perspectives in its development.

In this transition from the macroscopic to the mesoscopic, numerical simulations played a crucial role. In fact, realistic macroscopic systems with $N \sim \mathcal{O}(10^{20})$

degrees of freedom are far beyond the limit of current computers and, for this reason, all the systems analyzed in numerical experiments are, in some sense, “small”. On the other hand on a computer we can track the motion of every single particle of the system and observe other details and collective behaviours that cannot be measured in actual experiments. This possibility has urged many physicists, since the availability of the first computers in the 1950s, to verify some of the hypothesis underlying statistical mechanics. For instance Fermi, Pasta and Ulam [39] in the 1950s challenged the ergodic hypothesis (one of the pillars of statistical mechanics, which will be discussed in depth later), highlighting for the first time the differences between small and macroscopic systems.

In this chapter we will try to persuade the reader that extending statistical mechanics to small systems is a very delicate, though fascinating, operation, and it can lead to misleading conclusions, when performed without the required care.

For this reason we will discuss the importance of the presence of a large number of degrees of freedom to validate some results of statistical mechanics. These considerations will be fundamental to determine the classes of small systems for which a statistical approach is still meaningful: in particular we will show that it is possible to construct a coherent “small system thermodynamics” only for stochastic systems or for isolated *many-particle small systems* (MPSS), i.e. systems composed by $\mathcal{O}(10^2)$ particles. We restrict our attention, for the sake of consistency and simplicity, to classical systems: in fact, the description of quantum small systems, even with some similarities, must be developed independently, since it presents a very different phenomenology and a huge number of technical and conceptual subtleties [96].

1.1 The ergodic problem

The fundamental idea supporting the methods of statistical mechanics consists in identifying measurements of macroscopic observables like pressure or temperature, with averages of specific functions defined over the phase space. From a physical point of view, a measurement consists in the observation of the effects of the interaction of an instrument with the system of interest for a certain time \mathcal{T} , which is typically much longer than the characteristic times of the microscopic motion of atoms and molecules. For this reason, the results of the measurement starting at $t = t_0$ of a certain state-dependent quantity $A(\mathbf{X})$ does not give the instantaneous value $A(t_0) = A(\mathbf{X}(t_0))$ but, instead, the time average of the same quantity over an “infinitely long” time interval \mathcal{T} :

$$\bar{A}(\mathbf{X}(t_0)) = \lim_{\mathcal{T} \rightarrow \infty} \frac{1}{\mathcal{T}} \int_{t_0}^{t_0 + \mathcal{T}} A(\mathbf{X}(t)) dt, \quad (1.1)$$

where $\mathbf{X}(t)$ is the microscopic state of the system that, for Hamiltonian systems, is identified by the position and the momentum of each particle of the system. Since for a deterministic dynamics the state of the system at time t is uniquely determined by its initial state, the average will only depend upon the initial condition of the system $\mathbf{X}(t_0)$. The aim of statistical mechanics is to predict the result of the measurement of \bar{A} in Eq. (1.1), without the knowledge of the trajectory of the system $\{\mathbf{X}(t)\}_{t=t_0}^{t_0+\mathcal{T}}$:

this can be done by assuming that there exists a certain measure $\rho(\mathbf{X})$ ¹, which is invariant with respect to the dynamics of the system, and such that

$$\bar{A}(\mathbf{X}(t_0)) = \langle A \rangle = \int A(\mathbf{X})\rho(\mathbf{X})d\mathbf{X}, \quad (1.2)$$

where the integral on the rightmost side must be performed over all the possible available microscopic configurations (i.e. the phase-space for Hamiltonian systems). The assumption expressed in Eq. (1.2) goes under the name of *ergodic hypothesis*. It is interesting to note that when the ergodic hypothesis holds, since the right-hand side of Eq. (1.2) does not depend upon $\mathbf{X}(t_0)$, the time average of any observable must be independent of the initial condition of the trajectory.

The ergodic hypothesis, if true, allows then to predict the result of the measure of any macroscopic or thermodynamic quantity from microscopic considerations without the necessity of computing the whole trajectory of the system for a very long time: when the invariant distribution ρ is known, the result of the measure is simply the average value of the observable over the space of microscopic configurations of the system.

It is interesting to note that even if, from an historical point of view, the ergodic hypothesis was the conceptual bridge allowing the crossing between thermodynamics and microscopic dynamics, such an assumption is not necessary *nor* sufficient to justify a statistical approach in the description of mechanical systems.

The hypothesis is not necessary because many systems for which it is not possible to prove the ergodic hypothesis are, at a practical level, indeed ergodic, at least if $N \gg 1$ for a vast class of observables A . It must be stressed that the proof of the ergodic hypothesis in Eq. (1.2) is currently available for a very narrow class of systems: part of the difficulty of the proof relies in the fact that the hypothesis, as expressed in Eq. (1.2) is a very strict condition, since the equivalence between the two averages must hold for every observable $A(\mathbf{X})$.

On the other hand, the hypothesis is not even a sufficient condition to guarantee the significance of statistical mechanics: in fact, even when the two sides of Eq. (1.2) are equal, no condition about the time of convergence of the time average, which can be really long, is required. A simple calculation, elaborated by Boltzmann as an answer to his detractors, shows that, for certain observables, the time of convergence can be much longer than the age of the universe.

Take, as an observable, the characteristic function of a given subset \mathcal{M} of the space of all the microscopic configurations of the system

$$\chi(\mathbf{X}) = \begin{cases} 1 & \text{if } \mathbf{X} \in \mathcal{M} \\ 0 & \text{if } \mathbf{X} \notin \mathcal{M}. \end{cases} \quad (1.3)$$

From the definition of $\chi(\mathbf{X})$ it is trivial to prove that its average value with respect to the invariant ergodic distribution is simply the measure of the region \mathcal{M} (from a probabilistic point of view it is the probability of the event \mathcal{M} to occur)

$$\langle \chi \rangle = \int_{\mathcal{M}} \rho(\mathbf{X})d\mathbf{X} \quad (1.4)$$

¹Such a measure is not, in general, e.g. in non-Hamiltonian systems, absolutely continuous with respect to the Lebesgue measure. For this reason with the adopted notation is implicit that $\rho(\mathbf{X})$ is not a function but rather a *generalized function*.

From a dynamic point of view, the time average of the observable $\chi(\mathbf{X})$ corresponds to the fraction of time that the system passes in the region \mathcal{M} . It is then quite clear that, in order to have the time average to reproduce the “spatial” average (i.e. the measure of the region \mathcal{M}), the system must visit the region for an amount of time that, in proportion to the total time of the evolution, must be equal to the measure of the region: i.e. it must evolve inside the region for a certain amount of time, then leave the region and visit the region again at a successive time. If the system is ergodic, the average time of return $\tau(\mathcal{M})$ or *average recurrence time*, i.e. the average time spent by the system outside a given region \mathcal{M} between two subsequent visits, can be computed with the so-called *Kac’s Lemma* [58],

$$\tau(\mathcal{M}) = \frac{\tau_0}{\int_{\mathcal{M}} \rho(\mathbf{X}) d\mathbf{X}}, \quad (1.5)$$

with τ_0 a characteristic time i.e. it is the inverse of the measure, according to the invariant distribution, of the region of interest. From this formula is clear that the time τ can be made arbitrarily big by focusing on regions with very small measure, e.g. by measuring the time spent by the system in a very narrow region around one specific configuration of all its constituents.

Moreover, Boltzmann proved that, in an ideal gas of N particles, for large regions of phase-space whose measure is smaller than unity (but not necessarily vanishing), the recurrence time is exponential in N : therefore, if the number of particles is “thermodynamic” $\mathcal{O}(10^{20})$, the recurrence time can be several orders of magnitude larger than the age of the universe (for a throughout discussion see Ref. [38]). It is then clear that, in this case, the ergodic hypothesis, even if true from a mathematical point of view, does not allow to interpret macroscopic measures as time averages of microscopic observables.

Note that the ergodic hypothesis does not require any restriction about the size of the system: for example, the ergodic hypothesis is trivially true for one-dimensional Hamiltonian systems, since their trajectory coincides with the set of all the points² at a given energy E . For multi-dimensional systems, as already anticipated, proving the ergodic hypothesis can be a very difficult task and, as we emphasized, it may not be the solution of the problem.

From the above discussion, it may appear that the ergodic hypothesis, in addition to the apparent impossibility of obtaining a proof of Eq. (1.2) for realistic physical systems, is not a relevant aspect in the foundations of statistical mechanics. Nevertheless a solution of this apparent problem was exhibited by Khinchin in his celebrated book on the mathematical foundations of statistical mechanics [59].

He primarily observed that the ergodic hypothesis is a too strong condition to require: on one hand, it is clear that in thermodynamics one is only interested in measuring some specific quantities as temperature, volume, pressure, internal energy, etc. . . On the other hand, in thermodynamics, one only deals with macroscopic objects, i.e. with systems composed of a very large number of degrees of freedom. Khinchin proved that, by limiting the range of applicability of statistical mechanics to 1) a limited set of “interesting” observables; 2) systems with an enormously large

²For the sake of simplicity we assume that the constant- E surface is a simply connected set for every value of E .

number of constituents, it is possible to have a physical formulation of the ergodic hypothesis, whose proof is far more accessible and without the contradictions that we exhibited in the previous paragraphs.

Khinchin focused his attention on d -dimensional Hamiltonian systems with N particles and $2dN$ degrees of freedom $\mathbf{X} = \{\mathbf{q}_i, \mathbf{p}_i\}_{i=1}^N$ with separable hamiltonian i.e. such that

$$H = \sum_{i=1}^N h_i(\mathbf{q}_i, \mathbf{p}_i) \quad (1.6)$$

and on a special class of observables, the so-called *sum functions*, that can be expressed as the sum of N one-particle-dependent contributions

$$f(\mathbf{X}) = \sum_{i=1}^N f_i(\mathbf{q}_i, \mathbf{p}_i). \quad (1.7)$$

It was proved in [59] that, by denoting with $\langle f \rangle$ the phase space average of the observable f

$$\langle f \rangle = \int f(\mathbf{X}) \rho(\mathbf{X}) d\mathbf{X}, \quad (1.8)$$

one has

$$\text{Prob} \left(\left| \frac{f}{\langle f \rangle} - 1 \right| > \frac{K_1}{N^{\frac{1}{4}}} \right) < \frac{K_2}{N^{\frac{1}{4}}}, \quad (1.9)$$

where K_1 and K_2 are two positive constants $\mathcal{O}(1)$ and the probability must be considered with respect to the invariant measure $\rho(\mathbf{X})$. The physical interpretation of Eq. (1.9) is immediate when we identify the value assumed by a sum function $f(\mathbf{X})$ with one of the thermodynamic observables (average kinetic energy, local density etc...). If the system is large, the vast majority of the microscopic configurations of the system are points for which the value assumed by the thermodynamic observable is not very far from the average, or *equilibrium*, value $\langle f \rangle$. In other words, the time average appearing in the ergodic hypothesis gives the correct value for reasonable value of \mathcal{T} , only by considering sum functions and large systems, and, in conclusion, only because the value of $f(\mathbf{X})$ is almost always close to $\langle f \rangle$.

We want to stress that, even if the proof by Khinchin was conceived for sum functions in the simple case of non-interacting systems with separable hamiltonian, it is reasonable to expect that similar results hold also in the case of more realistic systems and observables. For instance, Mazur and van der Linden [73] proved that an equivalent version of Eq. (1.9) holds for systems with short range interactions. The key of the Khinchin interpretation of ergodicity is the fact that macroscopic observables are many-to-one functions, i.e. that large regions of the space of microscopic configurations corresponds to the same value of the observable. In the limit of large number of particles, some of these regions (those corresponding to values close to equilibrium) become significantly larger in measure with respect to the others and, therefore, it is increasingly more unlikely to find the system outside these regions. In conclusion, the role played by the dynamics in this context is negligible with respect to the presence of a large number of components.

Eq. (1.9) explains the correspondence between time averages and ensemble averages in the case of very large systems (thermodynamic limit) for a certain class

of observables. Different kinds of ergodicity may exist, e.g. ergodic systems where the time average converges only on astronomic timescales, but, in those cases, it is not possible to link statistical mechanics and thermodynamics.

It is quite clear that this reasoning cannot be straightforwardly extended to small systems. Nevertheless, it is possible to divide systems far from the thermodynamic limit into three different categories for which an interpretation to the ergodic problem can be found.

Single (or very few) particle systems. An Hamiltonian system composed by a single particle in one dimension that moves in an external potential, as already stated, is, by definition, ergodic. In fact, in the phase space of such a system (which is two-dimensional) the “constant energy surface”, where the system evolves, trivially coincides with the trajectory of the system. This is not necessarily true when we increase the number of particles, since it is possible to exhibit system composed by two constituents which are no longer ergodic (e.g. two coupled harmonic oscillators). Nevertheless there are several examples of few-dimensional ergodic systems and it is clear that in this case, the aforementioned objections, in particular the one regarding the time of convergence of the time average, are not relevant. For this reason, several authors [31, 50] believe that a notion of thermodynamic is possible also for this class of systems. It is very interesting to note that the first historical efforts to derive thermodynamics from statistical mechanics focused on isolated single particles systems for which H. von Helmholtz [103] was able to derive the so-called *heat theorem*: in Sec. 1.4 this topic will be discussed in detail. On the other hand, it should be quite clear that for such a class of systems, the equivalence between dynamics and thermodynamics is sometimes possible, but only at a formal level, because the essence of this equivalence, at a difference with the genuine one, is not the presence of a large number of particles.

Small systems with stochastic dynamics. With this term we refer to systems that have a non-deterministic, or probabilistic, dynamics: examples are Markov chains and systems described by stochastic differential equations (*Langevin Equation* of the kind

$$\ddot{x} = -\frac{dV(x)}{dx} - \gamma\dot{x} + \sqrt{2D}\eta, \quad (1.10)$$

where η represents white noise, $\langle \eta(t)\eta(t') \rangle = \delta(t-t')$, $V(x)$ is an external potential and γ and D are constants. The proof of the ergodic hypothesis, as in Eq. (1.2), for stochastic systems is far more accessible than its counterpart for deterministic systems: in particular it can be proven that every irreducible and non-periodic Markov process is ergodic [43, 74] and that, in addition, it also satisfies the so-called *mixing* condition, i.e. the fact that the initial probability distribution of the system converges toward an asymptotic stationary state:

$$\rho_0(\mathbf{X}, t = 0) \xrightarrow{t \rightarrow \infty} \rho(\mathbf{X}). \quad (1.11)$$

Of course, in complete analogy with the deterministic case, the time of convergence can be very long for stochastic large systems ($N \gg 1$), the solution to this problem consisting in focusing only on many-to-one observables $A(\mathbf{X})$. Again, focusing on

small stochastic systems, e.g. single-particle systems like the one described by Eq. (1.10), does not present particular difficulties regarding the ergodic problem: the ergodicity of the system is granted by the theorem presented above and the typical time of the system are small because the system is composed by few degrees of freedom. This is one of the (many) reasons why single particles stochastic systems have covered a central role in the most recent developments on small systems thermodynamics and stochastic thermodynamics: furthermore, thanks to their simplicity and accessibility many interesting analytical results can be obtained. One of the aims of our work is to extend and test the results obtained for single particle stochastic systems to more complex examples of small systems: the entire second part of the thesis will be devoted to this task.

Another conceptual advantage of considering small stochastic systems over deterministic ones is the fact the, in general, the aleatory dynamics can (and should) be interpreted as the effect of the interaction of the system of interest with a larger environment. This consideration, on one hand, increase the “realism” of this models, allowing to describe a larger fraction of actual experimental situation. On the other hand, in view of all this, the system we are investigating is a subsystem of a larger isolated system, and its state is a many-to-one function: $(X, Y) \in \mathbb{R}^{2d(N_1+N_2)} \rightarrow Y \in \mathbb{R}^{2dN_1}$, with $N_1 \ll N_2$, where Y is the variable representing the system we are studying, N_1 and N_2 are, respectively the number of particles of the system and the environment, and d is the dimension of the space. This means that, when the system is smaller than the environment, many different configurations of the environment correspond to the same value of Y : this allows, at least heuristically, to interpret the state of a small subsystem as a “thermodynamical observable”. Of course, in order to have a stochastic description of the subsystem, we also must assume that fluctuations from the average effect produced by the environment on the systems should not be too large: this condition is realized, for instance, when the mass of the particles of the system is larger than that of the particles of the environment.

Many-Particle Small Systems (MPSS). We emphasized the importance of the $N \gg 1$ limit in order to grant the validity of the ergodic hypothesis for macroscopic object. We also stressed the fact that, for large systems, the ergodicity cannot be valid for all the observables, but only for a subset composed of “interesting” many-to-one functions. On the other hand we have also seen that for very small systems, especially the ones composed by a single particle, the ergodic hypothesis is trivially true. It should be evident to the reader that single particle systems do not present most of the interesting behaviors produced by the interplay of the many different components of a system. For this reason one may be interested in investigating medium size systems, i.e. systems composed of $\mathcal{O}(10 - 10^2)$ particles. In this intermediate range, spanning from tenths up to several hundreds of particles, we have systems presenting, on one hand, all the interesting collective behaviors typical of large systems and that, in addition, are far enough from the infinite- N limit to exhibit several characteristics of small systems (large fluctuations, non-standard equilibrium states, etc...). For the sake of simplicity we will refer to systems belonging to this category with the name *Many-Particle Small System* (MPSS). The question regarding the ergodicity of these systems may appear more difficult to

answer with respect to single-particle or stochastic systems: on one hand we have complex systems, for which a proof of ergodicity cannot, in general, be obtained. On the other hand we have systems with a finite number of particles, with N large enough to have the convergence times of the time averages to explode (we recall that such a time is exponential in N), but much smaller than the Avogadro number, so that the $N \rightarrow \infty$ limit seems illegitimate.

Nevertheless, in the above discussion, we did not specify a minimum value of N for our considerations to be valid, since in macroscopic systems we always have an enormous number of particles $N \sim N_A$. Nevertheless, from a physical and practical point of view, it is important to stress that the Khinchin explanation of ergodicity applies to MPSS. In fact, for MPSS, even if N is not infinite, many large- N approximations (e.g. Laplace approximation of integrals . . .) can be carried on safely, and discrepancies with the asymptotic behavior are only quantitative, and do not inficiate the overall scenario.

It is interesting to note that, even though this not generally acknowledged, this is the class of systems usually studied in numerical simulations: in fact, typical computer simulations involve systems with a number of particles spanning from the few hundreds up to some tens of thousands of particles. This numbers are clearly much smaller (of several orders of magnitude) than the typical size of macroscopic systems $\mathcal{O}(10^{20})$, nevertheless they are commonly considered large enough to study the properties of real systems on everyday scale.

From the point of view of the ergodic theory we can safely assume that the Khinchin interpretation of ergodicity holds for this class of systems: i.e. that collective variables, that depend upon a large fraction of degrees of freedom of the system, never assume value too far from their average. On the other hand, since N is finite, the fluctuations of such variables are visible and non-negligible, and their study will be one of the main objectives of our work.

1.2 Thermodynamic Limit and Ensemble Equivalence

From a thermodynamic point of view, the state of a system is identified by a set of macroscopic observable like pressure P , volume V , internal energy U , entropy S , temperature T , . . . When an equation of state exists, i.e. an equation connecting two or more of these variables, any subset of independent macroscopic variable gives an equivalent description of the system: that is to say that, for example, it is possible to describe the equilibrium state of an ideal gas via its pressure P and volume V , via its volume V and its temperature T , or via its volume V and the internal energy U .

In addition, in equilibrium thermodynamics, the interaction of a system with the environment does not play any role: the physics of an isolated system, i.e. that does not exchange energy with its surroundings, with a given internal energy U and temperature T does not change when it is put in contact with a thermostat at the same temperature T . Differences between the two different physical situations emerge only when those systems are taken away from the equilibrium.

When turning from thermodynamics to statistical mechanics a different situation emerges. In fact it can be understood quite easily that the physics of an isolated mechanical system can, in principle, be quite different from the physics of a closed

or open system. From a statistical mechanics point of view this difference emerges in the different ensembles (*microcanonical* and *canonical*) corresponding to the aforementioned physical situations.

1.2.1 Microcanonical ensemble

Consider an isolated system composed of N particles living in a d -dimensional space, with hamiltonian $\mathcal{H}(\mathbf{X}, \boldsymbol{\lambda})$, being $\mathbf{X} \in \mathbb{R}^{2dN}$ the phase-space variable and $\boldsymbol{\lambda}$ a set of external parameters. An ensemble consists in a probability distribution function defined over the phase space such that the ergodic hypothesis, as discussed in the previous Section, holds. For an isolated system, denoting with E its conserved energy, this probability distribution reads

$$\rho_{\{E, \boldsymbol{\lambda}, N\}}(\mathbf{X}) = \frac{\delta(\mathcal{H}(\mathbf{X}, \boldsymbol{\lambda}) - E)}{\omega(E, \boldsymbol{\lambda}, N)}, \quad (1.12)$$

with $\delta(\cdot)$ the Dirac delta and $\omega(E, \boldsymbol{\lambda}, N)$ the density of states,

$$\omega(E, \boldsymbol{\lambda}, N) = \int d\mathbf{X} \delta(\mathcal{H}(\mathbf{X}, \boldsymbol{\lambda}) - E). \quad (1.13)$$

In this context the total energy E can be readily identified with the internal energy of the system U . In order to obtain the other thermodynamic observables it is necessary to relate the thermodynamic entropy S with a microscopic quantity, via the celebrated Boltzmann tombstone formula

$$S(E, \boldsymbol{\lambda}, N) = k_B \ln \omega(E, \boldsymbol{\lambda}, N). \quad (1.14)$$

As we will discuss in the following sections, the choice of Eq. (1.14) over the different possible definitions of S has generated a very heated scientific dispute in the last few years. From the definition of entropy the temperature T simply reads:

$$T(E, \boldsymbol{\lambda}, N) = \left(\frac{\partial S(E, \boldsymbol{\lambda}, N)}{\partial E} \right)^{-1}. \quad (1.15)$$

Analogously, all the generalized pressure p_i , i.e. the derivatives of the internal energy with respect to the external parameters can be expressed as

$$p_i(E, \boldsymbol{\lambda}, N) = T \frac{\partial S(E, \boldsymbol{\lambda}, N)}{\partial \lambda_i} \quad (1.16)$$

When the number of particles N is very large, i.e. for macroscopic objects, the asymptotic properties of intensive quantities must be analyzed. A very natural problem concerns the existence of the so-called *thermodynamic limit*,

$$s(e, \boldsymbol{\lambda}, \rho) = \lim_{\substack{N \rightarrow \infty \\ N/V = \rho}} \frac{1}{N} S(Ne, \boldsymbol{\lambda}, N), \quad (1.17)$$

where the number density $\rho = N/V$, and V is the average volume occupied by the system; the specific energy $e = E/N$ must be kept constant in the limiting operation. For systems with pair interaction among its constituents, i.e. with hamiltonian

$$\mathcal{H}(\{\mathbf{p}_i, \mathbf{q}_i\}_{i=1, \dots, N}) = \sum_{i=1}^N \frac{|\mathbf{p}_i|^2}{2m} + \sum_{i < j} V(|\mathbf{r}_i - \mathbf{r}_j|), \quad (1.18)$$

the existence of the thermodynamic limit can be proved [86], provided the potential $V(r)$ is *stable* (the total potential energy is limited) and *short-range* (i.e. it decrease faster than r^{-d} for large r). A fundamental property that can be proven for this class of systems is the convexity of the function $s(e, \boldsymbol{\lambda})$ i.e. the fact that, for every e_1, e_2 and $0 \leq \lambda \leq 1$,

$$s(\lambda e_1 + (1 - \lambda)e_2) \geq \lambda s(e_1) + (1 - \lambda)s(e_2). \quad (1.19)$$

Systems with unstable potentials or long-range interactions may also admit a thermodynamic limit, but, in some cases, they lose the convexity property.

1.2.2 Canonical ensemble and equivalence of ensembles

Consider a system \mathcal{S} , composed of two subsystems \mathcal{S}_1 , with N_1 particles, and \mathcal{S}_2 of $N = N_2 - N_1$ particles. The hamiltonian \mathcal{H} of the system is the sum of three terms,

$$\mathcal{H}(\mathbf{X}) = \mathcal{H}_1(\mathbf{X}_1) + \mathcal{H}_2(\mathbf{X}_2) + \mathcal{H}_{int}(\mathbf{X}_1, \mathbf{X}_2), \quad (1.20)$$

where \mathbf{X} represents the phase space variable of the system and \mathbf{X}_1 and \mathbf{X}_2 , respectively, the phase space variables of \mathcal{S}_1 and \mathcal{S}_2 . We restrict our attention to the situation where the system \mathcal{S}_1 is much smaller than \mathcal{S} and \mathcal{S}_2 , i.e. when $N_1 \ll N$. We also assume that the interaction term \mathcal{H}_{int} in the hamiltonian \mathcal{H} is negligible when compared to the other two terms, this last hypothesis being satisfied by systems with short-range interaction. Physically, we imagine that the subsystem \mathcal{S}_1 , i.e. our system of interest, is attached via short range interactions to a much larger system, a *thermostat* or *reservoir*, whose energy is so large that it is not affected by the energy exchanges with \mathcal{S}_1 . With this assumptions, is quite easy to prove [51] that, if the the total isolated system $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$ has energy E and it is described by the microcanonical probability distribution function, Eq. (1.12), the probability distribution of \mathcal{S}_1 reads

$$\rho_{\{\beta, \boldsymbol{\lambda}, N\}}(\mathbf{X}_1) = \frac{1}{\mathcal{Z}(\beta, \boldsymbol{\lambda}, N)} e^{-\beta \mathcal{H}_1(\mathbf{X}_1)}, \quad (1.21)$$

where $\beta = 1/T$ is the inverse temperature of the total system \mathcal{S} obtained from Eq. (1.15), and \mathcal{Z} is a normalization constant, the so-called partition function. The connection with thermodynamics is given by the formula

$$F(\beta, \boldsymbol{\lambda}, N) = -\frac{1}{\beta} \ln \mathcal{Z}(\beta, \boldsymbol{\lambda}, N), \quad (1.22)$$

where F is the Helmholtz free energy, i.e. the Legendre transform of the thermodynamic internal energy. Again, when considering large systems, one is, in general, interested in studying the intensive properties of the system, i.e. the thermodynamic limit:

$$f(\beta, \boldsymbol{\lambda}, \rho) = \lim_{\substack{N \rightarrow \infty \\ N/V \rightarrow \rho}} \frac{1}{N} F(\beta, \boldsymbol{\lambda}, N) \quad (1.23)$$

The existence of the thermodynamic limit in the canonical ensemble and the convexity of the specific free energy f , can be proven under the assumptions of stability and

short-range of the potential. Moreover, it is very easy to prove that, in this limit, the canonical and the microcanonical ensemble describe the same thermodynamic state: i.e. that it is possible to connect the free energy $f(\beta, \boldsymbol{\lambda}, \rho)$ and $s(e, \boldsymbol{\lambda}, \rho)$ through a Legendre-Fenchel transform [51]:

$$f(\beta, \boldsymbol{\lambda}, \rho) = \inf_e \{\beta e - s(e, \boldsymbol{\lambda}, \rho)\}. \quad (1.24)$$

This equation represents the facts that the two thermodynamic functions, that were derived in different ensembles, are describing the same situation with different thermodynamic variables (energy in the microcanonical and temperature in the canonical). The proof of Eq. (1.24) consists in computing the partition function

$$\begin{aligned} \mathcal{Z} = e^{-\beta N f} &= \int e^{-\beta \mathcal{H}(\mathbf{X})} d\mathbf{X} = \int \delta(\mathcal{H}(\mathbf{X}) - E) e^{-\beta(E)} d\mathbf{X} dE = \\ &= \int e^{-\beta N e} \omega(E) dE, \end{aligned} \quad (1.25)$$

where, in the last passage, we used Eq. (1.13) to recover the density of states $\omega(E)$. If N is large, it is possible to write $\omega(E) = \exp\{Ns(e)\}$, and then perform a steepest descent computation to obtain the value of the integral in the last term of the equation. Then, by comparing the second and the last term of the chain of equalities one simply gets Eq. (1.24).

This conclusion, as already observed, covers a central role in statistical mechanics: it is of fundamental importance to recover, from microscopic considerations, that, with respect to the thermodynamic properties, there is no difference between isolated and closed systems. Moreover it is very important to stress that one of the most remarkable successes of statistical mechanics, i.e. the possibility of a microscopic interpretation of phase transitions, can be achieved only in the thermodynamic limit: the literature on this topic is huge, the interested reader can refer to common textbooks.

In spite of this, one may still be interested in the physics of systems far from the thermodynamic limit, where the two ensembles may produce different results. In this regime, when N is finite, fluctuations of macroscopic observables, that in general decrease as $1/N$, may be detectable and measurable. In particular, it is necessary to stress that the ensemble equivalence is an equality relating thermodynamic quantities, i.e. the average of some microscopic observable that depend upon a large number of degrees of freedom. Therefore, there may exist observables whose fluctuations (e.g. their standard deviations) are different in the canonical and microcanonical ensemble: this fact, which is usually neglected since, for large N , fluctuations vanish, is of central importance in small systems.

This is particularly relevant from a dynamical point of view: consider a system for which a certain observable \mathcal{A} has different fluctuations in the canonical and the microcanonical ensemble $\sigma_A^{can}(T) \neq \sigma_A^{micro}(e(T))$, where $e(T)$ is the value of the energy per particle that minimizes the r.h.s. of Eq. (1.24). Then, if we denote with $\dot{\mathcal{A}}$ the time derivative of \mathcal{A} ,

$$\mathcal{A}(t) - \mathcal{A}(0) = \int_0^t \dot{\mathcal{A}}(t') dt', \quad (1.26)$$

we immediately get

$$\sigma_{\mathcal{A}}^2 = \int_0^t \int_0^t dt' dt'' \langle \mathcal{A}(t') \mathcal{A}(t'') \rangle, \quad (1.27)$$

which means that, when the variances differ, the correlation functions are different and, therefore, also the dynamics from which they are generated. In conclusion, macroscopic observables with the same average equilibrium values may have different evolutions in the two considered ensembles.

In the second part of this Thesis we will discuss a system composed of a small number of degrees of freedom, in which the above situation can be directly observed and studied. In particular we will see how the differences between the two dynamics, which should survive in the thermodynamic limit, can be observed only when N is small: this is due to the fact that the timescales on which the differences appear diverge when N increases, and, therefore, the discrepancies are no longer observable in the $N \rightarrow \infty$ limit.

1.3 Macroscopic Irreversibility

Everyday experience demonstrates that many natural processes are intrinsically irreversible at the *macroscopic level*. Think of a gas initially confined by a septum in one half of a container, that spontaneously fills the whole available volume as soon as the separator is removed. Or, closer to daily experience, consider the evolution of an ink drop into water [67]. We would be astounded and incredulous while observing the reverse processes to occur spontaneously: a gas self-segregating in one half of the container, or an ink drop emerging from a water-and-ink mixture. In thermodynamics, the second principle amounts to a formalization of this state of “incredulity”. From Newtonian (and quantum) mechanics, we know that at the *microscopic level* the dynamics is reversible. How can we reconcile macroscopic irreversibility with microscopic reversibility of the dynamics ruling the elementary constituents of macroscopic bodies?

A solution to this riddle was proposed more than 140 years ago, when Boltzmann laid down the foundation of statistical mechanics. At the beginning, Boltzmann’s ideas on macroscopic irreversibility elicited a heated debate mainly due to the *recurrence paradox*, formulated by Zermelo, and the *reversibility paradox* by Loschmidt (a detailed discussion on the historical and conceptual aspects of the Boltzmann’s theory can be found in Refs. [23, 24]).

According to Boltzmann, irreversibility is well defined only for systems with a very large number of degrees of freedom. It should be observed in the vast majority of the individual realizations of a macroscopic system starting far from equilibrium: “vast majority” is usually referred to as *typicality* in the literature [47, 107]. Hence, there is no need to repeat the experiment many times to understand that the free-gas expansion or the spreading of an ink drop in the water are irreversible processes, a single observation is enough. In this Section, following [Cerino et al., 2016a] we will briefly explain the Boltzmann interpretation of irreversibility and, at the light of the main interest of the Thesis, i.e. the thermodynamics of small systems, we will discuss the problem of irreversibility in systems composed of few degrees of freedom.

1.3.1 The Boltzmann interpretation of macroscopic irreversibility

Consider an Hamiltonian system whose microstate is denoted with $\mathbf{X} \in \mathbb{R}^{2dN}$. At any time t the state of the system is completely determined by the position \mathbf{x}_i and the momentum \mathbf{p}_i of its components (particles), $\mathbf{X}(t) \equiv (\mathbf{x}_1(t); \dots; \mathbf{x}_N(t)) \equiv (\mathbf{q}_1(t), \mathbf{p}_1(t); \dots; \mathbf{q}_N(t), \mathbf{p}_N(t))$. The whole set of admissible microscopic configurations, $\{\mathbf{X}\}$, defines the phase space, or Γ -space. The evolution of a macroscopic system from an initial state $\mathbf{X}(0)$ at time 0 up to a specified time $T > 0$, $\{\mathbf{X}(t)\}_{t=0}^T$, constitutes a “forward” trajectory. The time “reversed” trajectory is obtained by applying the time reversal transformation \mathcal{R} , i.e. considering as initial state the one with particles at the positions reached at time T but with reversed velocities, i.e. $\mathbf{X}^{\mathcal{R}}(0) = \mathcal{R}(\mathbf{X}(T)) \equiv (\mathbf{q}_1(T), -\mathbf{p}_1(T); \dots; \mathbf{q}_N(T), -\mathbf{p}_N(T))$. When the system is evolved from $\mathbf{X}^{\mathcal{R}}(0)$, thanks to the invariance of Newton’s equations under time reversal, it traces back the forward trajectory (with reversed velocities) as if the evolution movie were played backwards, i.e. given $0 \leq t \leq T$, $\mathbf{X}^{\mathcal{R}}(t) = \mathcal{R}(\mathbf{X}(T-t))$.

We already know that the macrostate of a large system ($N \gg 1$) is specified by a small number of macroscopic observables, $M_\alpha(t) = M_\alpha(\mathbf{X}(t))$ with $\alpha = 1, \dots, k \ll N$. The observables M_α to be qualified as “macroscopic” must depend on a large number of the system degrees of freedom. In general, we have that many microscopic configurations correspond to the same value of the observables, in other terms the relation between micro and macrostate is many to one. As already observed in the previous sections, some examples are the energy of a subsystem composed of many particles, the number density in specific (not too small) regions, or the number of particles with velocity in a given interval. At equilibrium the macroscopic observables assume specific values $M_\alpha^{eq} \equiv \langle M_\alpha \rangle_{eq}$, where $\langle \cdot \rangle_{eq}$ denotes the equilibrium average with respect to, e.g., the microcanonical distribution (in principle, other ensembles can be used, we use here the microcanonical one as it is the appropriate one for discussing isolated systems). We can define a state to be far from equilibrium when the observables deviate from their equilibrium values well beyond the equilibrium fluctuations, in other terms when $|M_\alpha - M_\alpha^{eq}| \gg \sigma_M^{eq} \equiv \sqrt{\langle M_\alpha^2 \rangle_{eq} - (M_\alpha^{eq})^2}$. Conversely, whenever $|M_\alpha - M_\alpha^{eq}| \approx \sigma_M^{eq}$ we speak of close-to-equilibrium states.

Macroscopic irreversibility refers to the fact that when starting from far-from-equilibrium states, the (macroscopic) system evolves toward equilibrium, i.e. at times long enough we have that $M_\alpha(t) \rightarrow M_\alpha^{eq}$, while we never observe the opposite, i.e. that starting close to equilibrium the system approaches (spontaneously) a far from equilibrium state, in spite of the fact that such reversed trajectories would be perfectly compatible with the microscopic dynamics.³

Boltzmann explained the asymmetry in the time evolution of macroscopic systems in term of a probabilistic reasoning. He realized that the number of microscopic configurations corresponding to the equilibrium state, i.e. \mathbf{X} such that $M_\alpha(\mathbf{X}) \approx M_\alpha^{eq}$ is, when the number of degrees of freedom N is very large, astronomically (i.e. exponentially in N) larger than those corresponding to non-equilibrium states: as explained in Sec. 1.1, an analogous observation was exploited by Khinchin to

³Obviously, weakly interacting particles, in an empty infinite space, can spontaneously leave the region where they were initially released and never return there [95]. This form of irreversibility is quite trivial, so we shall only consider systems evolving in a bounded region of Γ .

solve the ergodic problem in macroscopic objects. The essence of the Boltzmann reasoning is that, somehow “intuitively”, it is overwhelmingly “more probable” to see a system evolving from a very “non-typical” state, i.e. which can be obtained with (relatively to equilibrium) a negligible number of microscopic configurations, toward an equilibrium state, which represents a huge number of microscopic states, than to see the opposite. This “intuitive”⁴ notion of “more probable” can be formalized in terms of the Boltzmann’s entropy of a given macrostate, which is the log of the number of microstates corresponding to that macrostate, one of the greatest contributions of Boltzmann was to identify such entropy with the thermodynamic entropy when in equilibrium. These entropic aspects have been (beautifully and thoroughly) discussed in other articles [10, 46, 67], to which we refer to.

In the case of very dilute (monoatomic) gases, Boltzmann was even able to do more, with his celebrated H -theorem, by demonstrating the irreversible dynamics of the one-particle empirical distribution function⁵ $f_1(\mathbf{x}, t) = \frac{1}{N} \sum_{i=1}^N \delta(\mathbf{x} - \mathbf{x}_i(t))$, where $\mathbf{x} = (\mathbf{q}, \mathbf{p})$ denotes the position and momenta of a single particle, i.e. the so-called μ -space. The interesting aspects about the empirical distribution are that f_1 is a well defined macroscopic observable and can be, in principle, measured in a single system, e.g. in numerical simulation.

With the term *typicality* [47], we refer to the fact that the outcome of an experiment on a macroscopic system takes a specific (typical) value overwhelmingly often. In statistical mechanics typicality holds in the thermodynamic limit (and thus for $N \gg 1$). It is in such an asymptotics that the ratio between the set of typical (equilibrium) states and non-typical ones goes to zero extremely rapidly (i.e. exponentially in N), thus it is only when N is large that the probability to see the irreversible dynamics of initially far-from equilibrium macrostates toward equilibrium ones becomes (at any practical level) one. The concept of typicality is not only at the basis of the second law, but (possibly at a more fundamental level) in the very possibility to have reproducibility of results in experiments (on macroscopic objects) or the possibility to have macroscopic laws [10]. Consider a system with N particles, and a given macroscopic observable $\mathcal{A}_N(\mathbf{X})$. Let us assume an initial well behaving⁶ phase-space density $\rho(\mathbf{X}, 0)$ prescribing a given macroscopic state. From a physical point of view we can assume, e.g. $\rho(\mathbf{X}, 0) = 0$ if $\mathcal{A}_N(\mathbf{X}) \notin [\mathcal{A}_0 : \mathcal{A}_0 + \Delta\mathcal{A}]$, for some \mathcal{A}_0 (usually chosen far from equilibrium) with $\Delta\mathcal{A}/\mathcal{A}_0 \ll 1$, that is we consider that one or more (macroscopic) constraints on the dynamics are imposed. Then we consider the ensemble of the microstates compatible with that constraint. Common examples are, e.g., a gas at equilibrium confined in a portion of the container by some separator (see next Section for some numerical examples). At time $t = 0$ such constraints are released and we monitor the evolution of the system by looking at the macroscopic observable $\mathcal{A}_N(\mathbf{X}(t))$: we denote with $\langle \mathcal{A}_N(t) \rangle$ the average over all the possible initial conditions weighted by $\rho(\mathbf{X}, 0)$. If $N \gg 1$ and the initial state is far from equilibrium $\|\mathcal{A}_0 - \mathcal{A}^{eq}\| \gg \sigma_{\mathcal{A}}^{eq}$, according to the “Boltzmann’s interpretation” of irreversibility, the time evolution of $\mathcal{A}_N(t)$ must be typical i.e. apart from a set

⁴Intuitive only a posteriori and in a very subtle way indeed.

⁵Here, we define it through Dirac-deltas from a mathematical point of view we should always think to some regularization via, e.g., some coarse-graining.

⁶From a mathematical point of view this means that it has to be absolutely continuous with respect to the Lebesgue measure.

of vanishing measure (with respect to $\rho(\mathbf{X}(0), 0)$), most of the initial conditions originate trajectories over which the value of $\mathcal{A}_N(\mathbf{X}(t))$ is very close to its average $\langle \mathcal{A}_N(t) \rangle$ at every time t .⁷ In other terms, if N is large, behaviors very different from the average one (e.g. a ink drop *not* spreading in water) *never* occur:

$$\text{Prob}\{ \mathcal{A}_N(t) \approx \langle \mathcal{A}_N(t) \rangle \} \approx 1 \quad \text{when} \quad N \gg 1. \quad (1.28)$$

The rigorous proof of the above conjecture is very difficult and, of course, it is usually required to put some restrictions. It is remarkable that it is possible to show the validity of this property in some stochastic systems, like the Ehrenfest model [33].

1.3.2 Can small systems be irreversible?

In the previous subsection we emphasized the importance of a large number of degrees of freedom to have irreversible behavior. Nonetheless, it is worth reporting that some authors have a different opinion. For example, in his comment to a Lebowitz paper, [67] Driebe [3] states that irreversible processes can be observed in systems with few degrees of freedom, such as the baker transformation or other reversible, low-dimensional chaotic systems, under the form of relaxation of phase space probability density. If a dynamical system exhibits “good chaotic properties”, more precisely, it is *mixing*, a generic probability density distribution of initial conditions, the *ensemble*, $\rho(\mathbf{X}, 0)$, relaxes (in a suitable technical sense) to the invariant distribution for large times t

$$\rho(\mathbf{X}, t) \rightarrow \rho^{inv}(\mathbf{X}) . \quad (1.29)$$

It is worth remarking that in systems satisfying Liouville theorem, the relaxation to the invariant distribution must be interpreted in a proper mathematical sense: for every $\epsilon > 0$ and for every \mathbf{X} , one has

$$\int_{|\mathbf{X}-\mathbf{Y}|<\epsilon} \rho(\mathbf{Y}, t) d\mathbf{Y} \rightarrow \int_{|\mathbf{X}-\mathbf{Y}|<\epsilon} \rho^{inv}(\mathbf{Y}) d\mathbf{Y} . \quad (1.30)$$

We want to make clear here that the property (1.29) or, equivalently, (1.30) is a form of irreversibility completely unrelated to the second law of thermodynamics. In such low-dimensional chaotic systems, irreversibility due to the mixing property is observed only by considering ensembles of initial conditions, while single realizations do not show a preferential direction of time. If that was the essence of irreversibility, e.g., we should repeat the experiment of an ink drop spreading many times and, only by averaging over the different realizations, we would notice a preference in the direction on time. This occurs also in macroscopic systems when we monitor the evolution of an observable that is not macroscopic, e.g. a single molecule property either in the gas or in the ink drop. In that case, nothing astounding happens by looking at the forward or reversed trajectory, as we cannot decide the direction of

⁷Such property does not hold for all the observables in all situations, for instance one has to exclude situations in which the macroscopic dynamics is unstable. In this case the transient to equilibrium may vary from realization to realization though the final equilibrium state will be reached by almost all the realizations.

the process. On the other hand, in thermodynamics, an irreversible behavior must be typical, i.e. it must be observed every time the experiment is repeated.

In order to highlight the difference between the “genuine” macroscopic irreversible behavior of large systems and the fictitious one induced by the mixing property, we introduce an idealized simple model, which is an idealization of the spreading of an ink drop into water. It consists in a discrete-time symplectic map (akin to an Hamiltonian system) involving $2N$ degrees of freedom, and 2 auxiliary variables. We consider a special case of the system proposed in Ref. [7], in particular

$$\begin{cases} y_i(t+1) &= y_i(t) + \epsilon \cos[x_i(t) - \theta(t)] \\ x_i(t+1) &= x_i(t) + y_i(t+1) \\ J(t+1) &= J(t) - \epsilon \sum_{j=1}^N \cos[x_j(t) - \theta(t)], \\ \theta(t+1) &= \theta(t) + J(t+1). \end{cases} \quad (1.31)$$

Each pair (x_i, y_i) identifies a “particle” ($i = 1, \dots, N$),⁸ and periodic boundary conditions on the two-dimensional torus $\mathbb{T}_2 = [0, 2\pi] \times [0, 2\pi]$ are assumed. For $\epsilon = 0$, the particles do not interact, while when $\epsilon > 0$ (in our numerical examples we use $\epsilon = 1$) particles interact (the “collisions” of water molecules) via a mean-field-like interaction, mediated by the variables θ and J . We emphasize that θ and J do not have a precise physical meaning, they represent a simple mathematical expedient to introduce the interaction among particles in a symplectic manner. System (1.31) can be shown to be time-reversible. We used a system with interacting particles to

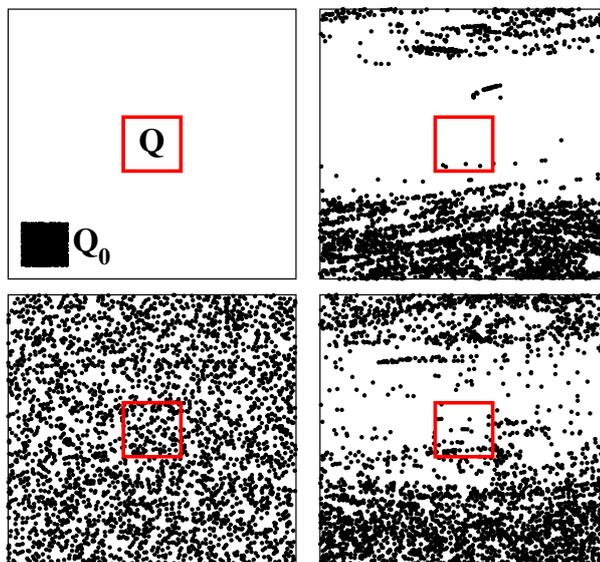


Figure 1.1. Irreversible spreading of an ink drop of $N_I = 3.2 \cdot 10^3$ particles on the Torus \mathbb{T}_2 at $t = 0, 4 \cdot 10^3, 2.9 \cdot 10^4, 2.33 \cdot 10^5$ (in clockwise order from top left). The N_I ink particles start uniformly distributed in $Q_0 \equiv [0.3 : 1.3] \times [0.3 : 1.3]$, while the $N_W = 10^7$ solvent ones have been thermalized in a previous time integration. The instantaneous occupation $n(t)$ is monitored in the (red) box Q centered in (π, π) with side $\pi/5$.

⁸Notice that x_i and y_i can be interpreted as the position and momentum of the i -particle, respectively.

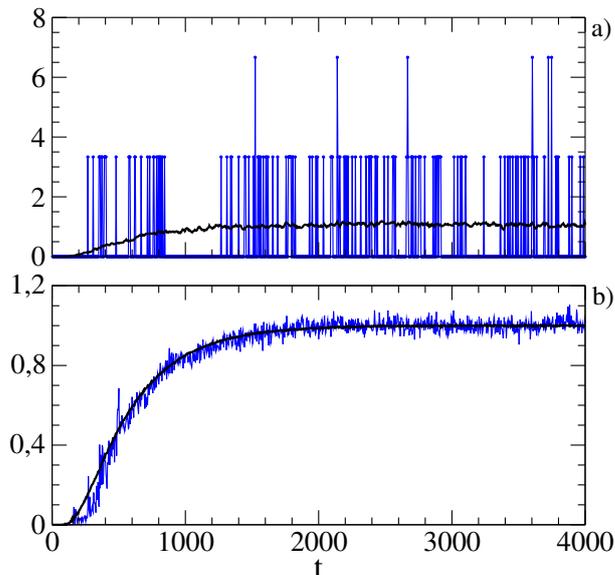


Figure 1.2. Instantaneous occupation $n(t)/n_{eq}$ of the set Q (blue, fluctuating curve) and its average $\langle n(t) \rangle / n_{eq}$ (black, smooth curve) over 500 independent initial conditions starting from Q_0 : (a) $n_{eq} = 0.3$ (drop with very few particles, $N_I = 8$ and $N_W = 2500$) and (b) $n_{eq} = 10^3$ (drop with many particles $N_I = 2.5 \times 10^4$ and $N_W = 10^6$).

avoid confusion between the genuine thermodynamic irreversibility and the mixing property, Eq. (1.29). Since our system is composed of N interacting elements it should be clear that we are dealing with a single large system and not with a collection of different initial conditions as if the particles were non-interacting and evolving according to a generic mixing map of the torus. After several iterations, the system (1.31) reaches an “equilibrium” dynamical state characterized by a uniform distribution of particles on \mathbb{T}_2 . To mimic the spreading of a cloud of “ink”, we split the N particles into N_W particles of solvent (water) and N_I particles of solute (ink), with $N = N_W + N_I$ and $N_I \ll N_W$. Then, we prepare the initial condition of the system with the N_W particles at equilibrium (e.g. after a long integration with N_W particles only), and the solute particles uniformly distributed in a small region Q_0 of \mathbb{T}_2 (top left panel in Fig. 1.1). During the evolution, to measure the degree of mixing, we monitor the number of ink particles, $n(t)$, which at time t reside in a given set $Q \subseteq \mathbb{T}_2$ (the red box Q in Fig. 1.1). At equilibrium, when ink is well mixed, the N_I particles will also distribute uniformly over \mathbb{T}_2 , and thus $n(t)$ will fluctuate around $n_{eq} = N_I \mathcal{A}(Q) / \mathcal{A}(\mathbb{T}_2)$, where $\mathcal{A}(Q)$ is the area of the subset Q .

We now compare (see Fig. 1.2) the behavior of $n(t)$ for a single trajectory with the average $\langle n(t) \rangle$, computed over an ensemble of many independent releases of the ink drop, with the water in different (microscopic) initial conditions arbitrarily chosen in the equilibrium state. Moreover, we study the difference between the case $N_I \sim O(1)$ (Fig. 1.2a) and $N_I \gg 1$ with $N_I \ll N_W$ (Fig. 1.2b). It is important to realize that while the latter case ($N_I \gg 1$) the ink drop can be considered a macroscopic object, in the former ($N_I \sim O(1)$) it cannot. In both cases, we observe that $\langle n(t) \rangle / n_{eq}$ increases monotonically with t , asymptotically approaching 1. However, a dramatic difference emerges if we look at the single realization. For a (macroscopically well

defined) drop with $N_I \gg 1$, the single trajectory closely follows the average one (Fig. 1.2b), and we can define an irreversible behavior for the individual drop. Conversely, when $N_I \sim O(1)$, the single trajectory is indistinguishable from its time reverse one (Fig. 1.2a) and strongly differs from the average one. The latter apparently shows a form of irreversibility, but it is thus a mere artifact of the average over the initial distribution and the special initial condition. We stress that, the lack of irreversibility in this case is due to the fact that, being N_I small, $n(t)$ cannot be considered a macroscopic observable even if the system water plus drop is large ($N \gg 1$), as $n(t)$ depends only on the few “molecules” of ink.

1.3.3 Small systems with an irreversible behavior

From the above discussion it appears clear that genuine macroscopic irreversibility is only a property of large (macroscopic) systems. Isolated single (or few) particle systems, even when chaotic, thanks to the invariance of Newton equations under time reversal, cannot be irreversible at the level of single trajectory.

Nevertheless, irreversibility can still be observed in the two remaining categories of small systems introduced in Sec. 1.1, namely stochastic systems and many-particle small systems (MPSS).

Stochastic Systems. As already anticipated in the previous Sections, a stochastic dynamics describes a system that, during its evolution, interacts with an hidden external environment (the *thermostat*) whose size is much larger than the one of the system of interest: the interaction of the system with the thermostat, whose complete description is usually not accessible, is described by the stochastic terms appearing in the equations of motion of the system. It is important to remark that in the great majority of cases it is not possible to perform all the analytic calculations that connect the microscopic description to the coarse-grained equations. For this reason, one is generally forced to impose *a fortiori* the properties that the stochastic terms must satisfy: in particular it is customary to require the thermostat to drive the system towards equilibrium, hence being responsible for an irreversible evolution. This property can be formalized into the form of *detailed balance* condition,

$$p(\mathbf{X}_2, t | \mathbf{X}_1, 0) p^{eq}(\mathbf{X}_1) = p(\mathcal{R}\mathbf{X}_1, t | \mathcal{R}\mathbf{X}_2, 0) p^{eq}(\mathcal{R}\mathbf{X}_2), \quad (1.32)$$

where $p(\mathbf{X}, t; \mathbf{Y}, t')$ indicates the probability distribution function of finding the system in the state \mathbf{X} at time t conditioned to finding it in the state \mathbf{Y} at a previous time $t' < t$, $p^{eq}(\mathbf{X})$ is the invariant probability distribution of the system, and \mathcal{R} is the time-reversal operator.

Now consider a *non-equilibrium* state \mathbf{X}_1 and an *equilibrium* one \mathbf{X}_2 , i.e such that $p^{eq}(\mathbf{X}_1) \ll p^{eq}(\mathbf{X}_2)$: the terms equilibrium and non-equilibrium refer to fact that the state \mathbf{X}_2 is often visited by the system in its stationary state, while the opposite is true for the state \mathbf{X}_1 . It is immediate to see that, if Eq. (1.32) holds, the probability of evolving from \mathbf{X}_1 towards \mathbf{X}_2 is much larger than the one of covering the opposite trajectory: this means that, when initializing the system in a non-equilibrium condition \mathbf{X}_1 , we will observe with probability close to one a trajectory driving the system towards equilibrium configurations. On the other hand,

when initializing the system in an equilibrium condition \mathbf{X}_2 we will (almost) never register a trajectory towards a non-equilibrium state.

It is clear, then, that a system that satisfies the detailed balance condition is intrinsically irreversible: this may appear rather artificial to the reader since this is an *a fortiori* imposed condition. Nevertheless, a closer inspection reveals that Eq. (1.32) truly reflect the essence of thermodynamic irreversibility. In fact, since the variable \mathbf{X} is a macrostate (i.e. a projection on a low-dimensional space) of the isolated structure composed by the system and the thermostat, non-equilibrium (i.e. low-probability) states are the ones corresponding to a small number of microscopic configurations: the detailed balance condition, then, is a probabilistic statement of the fact that most of the trajectories starting in a non-equilibrium macrostate correspond to an evolution towards equilibrium.

Many-particle small systems (MPSS). We now discuss why small systems with an intermediate number of particles ($\mathcal{O}(1) \ll N \ll \mathcal{O}(\mathcal{N}_A)$) may display irreversible behaviors. In the first place, for these systems, collective observables \mathcal{A} , that depend upon a large fraction of the degrees of freedom of the system, can be considered true “macroscopic” observables already for $N \sim 10$: indeed, at equilibrium these functions have relative fluctuations (i.e. the ratio of the standard deviation and the average value $\sigma_{\mathcal{A}}/\langle\mathcal{A}\rangle$) decreasing as $1/N$. This number, for the considered values of N , is usually small enough to distinguish between fluctuations of equilibrium and true non-equilibrium conditions, i.e. microstates \mathbf{X} such that $|\mathcal{A}(\mathbf{X}) - \langle\mathcal{A}\rangle| \gg \sigma_{\mathcal{A}}$. This is a signature of the fact that region in the phase space corresponding to close to equilibrium values of \mathcal{A} , are much larger in volume with respect to the other regions.

Unlike real macroscopic systems with $N \sim \mathcal{N}_A$, we do not expect Eq. (1.28) to hold in a strict sense: nevertheless, in complete analogy with the Khinchin equation, Eq. (1.9), it is plausible to assume, at least heuristically, that the differences between the average trajectory and the single realization decrease as some inverse power of the number of degrees of freedom,

$$\sigma_{\mathcal{A}}(t) \sim N^{-\alpha} \quad \alpha > 0, \forall t > 0, \quad (1.33)$$

where $\sigma_{\mathcal{A}}(t) = \langle(\mathcal{A}(\mathbf{X}(t)) - \langle\mathcal{A}\rangle(t))^2\rangle$, and the averages are carried out with respect to the initial distribution probability $\rho(\mathbf{X}, t = 0)$. With small values of $N > 1$, $\sigma_{\mathcal{A}}$ is, on one hand, small enough to identify an irreversible behavior of the system that is also typical, i.e. that is repeated, at least at some degree of approximation, at every realization of the experiment. On the other hand, we have measurable and interesting fluctuations with respect to the average trajectory that can actually be observed and studied, and that in true macroscopic objects may be too small to be detected (the expected scenario is an intermediate situation between the top and the bottom panel of Fig. 1.2).

Again, this last category of systems includes the vast majority of situations that, in the last decades, scientists have observed in numerical simulations. Nevertheless, many times, one is interested to reproduce thermodynamic results and, for this reason, is induced to interpret the characteristic features of these intermediate regime as “undesirable”.

From our point of view, on the other hand, it is very interesting to focus on MPSS, analyzing their characteristic nature, in the attempt of obtaining some general results.

1.4 Isolated small systems and the problem of the definition of temperature

We conclude the Chapter by showing an example discussed in [Cerino et al., 2015a] that demonstrates how trying to extend thermodynamic considerations to single or few particles systems may lead to confusing results. In particular, we want to emphasize that single particle systems cannot be used as counterexamples to well established statistical concepts and definitions: in fact, most of the important results of statistical mechanics, as showed in the previous sections, hold only because thermodynamic systems must be composed of a large (but possibly not infinite) number of particles.

Two different definitions of temperature in equilibrium statistical mechanics have been recently the subject of an intense debate [12, 16, 31, 40, 50, 91, 101, 106], after the publication of experimental measurements of a negative absolute temperature [9, 19]. In [9] it was demonstrated the possibility to prepare a state where the observed distribution of the modified kinetic energy *per atom* appeared to be inverted, i.e. with the largest population in the high energy states, yielding a *de facto* negative absolute temperature.

The possibility of a negative absolute temperature is well known since the theoretical work by Onsager on the statistical hydrodynamics of point vortices [76] and the experimental and theoretical results on nuclear spin systems by Pound, Ramsey and Purcell (see [64, 65, 85] for a review and discussion). In those investigations, it was clear that an inverse temperature parameter β ranging in the full infinite real line $(-\infty, \infty)$ did not lead to any inconsistency or paradox. Ramsey in 1956 already realised that “the Carathéodory form of the second law is unaltered.” [85]

A negative absolute temperature appears whenever the microcanonical entropy is non-monotonic in the energy, a condition which can be realized when the total energy has a global maximum, which may happen when the phase space is bounded. There are also cases where the phase space is bounded but the energy diverges: again this may lead to a non-monotonic entropy, an important example is given by point vortices [4, 5, 75, 76]. It is crucial to highlight that the lack of monotonicity (for entropy vs. energy) is realised if one adopts the simplest definition of microcanonical entropy, which is related to the logarithm of the number of states with a given energy. Such a definition appears in the so-called “tombstone formula” written on Boltzmann’s grave, “ $S = k \ln W$ ”, and it is the equivalent of the entropy defined in Sec. 1.2, Eq. (1.14): for a system with hamiltonian $H(\mathbf{q}, \mathbf{p})$ – where \mathbf{q} and \mathbf{p} are vectors in \mathbb{R}^{dN} , being d the dimension of the system, the Boltzmann entropy reads

$$S_B(E, N) = \ln \omega(E), \quad (1.34)$$

being $\omega(E)$ the density of states, i.e.

$$\omega(E) = \int \delta(H - E) d^{dN} Q d^{dN} P = \frac{\partial \Sigma(E)}{\partial E}, \quad (1.35)$$

1.4 Isolated small systems and the problem of the definition of temperature 23

and $\Sigma(E)$ the total “number” of states with energy less or equal then E , that is

$$\Sigma(E) = \int_{H < E} d^{dN} Q d^{dN} P. \quad (1.36)$$

In definition (1.34) we have ignored an additive constant which is not relevant in our discussion. Propagating the denomination, it is customary to define the “Boltzmann temperature” through

$$\beta_B = \frac{1}{T_B} = \frac{\partial S_B(E, N)}{\partial E}. \quad (1.37)$$

Some authors [31, 50] have argued that a different definition of microcanonical entropy, proposed by Gibbs, has to be used in statistical mechanics, in order to be consistent with a series of “thermodynamic” requirements and avoid unpleasant paradoxes found in single-particle systems. The Gibbs entropy, which is always monotonically increasing, reads

$$S_G(E, N) = \ln \Sigma(E), \quad (1.38)$$

and leads to the Gibbs temperature definition, which is always positive:

$$\beta_G = \frac{1}{T_G} = \frac{\partial S_G(E, N)}{\partial E} \geq 0. \quad (1.39)$$

Let us note that, since T_B is defined directly on the surface of interest (i.e. that at constant energy E), from the point of view of the ergodic approach its use appears rather natural. The Gibbs temperature, on the other side, enters through an ensemble average in the equipartition formula of textbooks [51]:

$$\left\langle x_i \frac{\partial H}{\partial x_j} \right\rangle = \delta_{ij} T_G, \quad (1.40)$$

where x_i is any of the components of vector (\mathbf{q}, \mathbf{p}) and the average is done in the microcanonical ensemble.

In standard systems, i.e. the one with a quadratic kinetic term in the hamiltonian, it is possible to prove [51] that the two entropies provide the same results for $N \gg 1$. This equivalence breaks down in small systems and in non-standard systems, i.e. those that have a non-monotonically increasing entropy and negative Boltzmann temperatures: in fact, since the Gibbs temperature is always positive, whenever a negative Boltzmann temperature appears, there must be a difference between the results with the two different formalizations.

One of the arguments of the supporters of the Gibbs entropy (the complete reasoning is discussed in depth in Ref. [50]) is that T_G appears in the theory of Helmholtz monocycles (which had an important role in the development of the Boltzmann’s ideas for the ergodic theory), for one-dimensional systems [17, 38, 103] which, for the sake of consistency, we revisit briefly.

Consider an unidimensional mechanical system with hamiltonian

$$H(q, p, \lambda) = \frac{p^2}{2m} + V(q, \lambda), \quad (1.41)$$

where λ is an external control parameter. We take a potential V such that the trajectories are periodic for any value of λ and E : we denote by $\tau(E, \lambda)$ the period of the trajectory and with $q^+(E, \lambda)$ and $q^-(E, \lambda)$, respectively the maximum and minimum value of q . We can calculate the following time averages:

$$\frac{T_G}{2} = \frac{1}{\tau(E, \lambda)} \int_0^\tau \frac{p(t)^2}{2m} dt, \quad (1.42)$$

$$P = \frac{1}{\tau(E, \lambda)} \int_0^\tau \frac{\partial V(q(t), \lambda)}{\partial \lambda} dt, \quad (1.43)$$

where the symbols T_G and P are chosen to represent the mechanical equivalent of temperature and pressure. The Helmholtz monocycle theorem states that exists a function, the ‘‘Gibbs’’ entropy,

$$S_G(E, \lambda) = \ln 2 \int_{q^-(E, \lambda)}^{q^+(E, \lambda)} \sqrt{2m[E - V(q, \lambda)]}, \quad (1.44)$$

such that

$$\frac{\partial S_G}{\partial \lambda} = \frac{P}{T}, \quad \frac{\partial S_G}{\partial E} = \frac{1}{T}. \quad (1.45)$$

It is clear that the r.h.s. of Eq. (1.44) represents the volume of the portion of the two-dimensional phase space enclosed into the periodic trajectory (which is also the one-dimensional constant- E surface), in analogy with the definition of the Gibbs entropy, Eq. (1.38). This important example suggests that, at least for unidimensional systems, the connection between thermodynamics and mechanics must involve the Gibbs entropy and not the Boltzmann equivalent: let us remark that all the systems discussed in [50], from which the authors try to show that only T_G is the ‘‘good’’ temperature, are small ($N = \mathcal{O}(1)$) and/or with long interactions. Nonetheless, as already anticipated, we will show that, when turning the attention to larger systems (macroscopic systems, MPSS, ...), in order to preserve many thermodynamic aspects, it is necessary to adopt the Boltzmann convention. In particular, we present a line of reasoning where Boltzmann temperature T_B (positive or negative) is the (unique) proper parameter which is relevant for the statistical properties of the energy fluctuations, as well as in determining the flux of energy between two systems at different temperatures, in addition it is measurable, without the appearance of any evident inconsistency.

In the following we will also exclude long-range potentials or mean-field models: it should be understood that long-range interactions certainly widen the phenomenology of statistical mechanics and may lead to complicate functional dependences for $S_B(E, N)$, e.g. with several maxima or minima, even for large N . Nevertheless they are not necessary for the discussion of negative temperature and, most importantly, they represent quite a peculiar case where even thermodynamics is not obvious: for instance, it is not evident that the typical *Gedankenexperiment* of putting in contact two – previously isolated – systems can be realized, as the isolation condition is prevented by the long-range interaction.

We also assume that $S_B(E, N)$ is always convex, i.e. $d^2 S_B(E, N)/dE^2 \leq 0$. This is certainly true in the limit of vanishing interaction and in short-range-interacting systems for large N , since S_B is strictly related to the large deviation function

associated to the density of states ⁹. Later we discuss in some details the origin of the convexity of $S_B(E, N)$, by means of large deviation theory considerations. Let us stress that the values of N to observe a large deviation principle are not necessarily “thermodynamic” ($N \rightarrow \infty$). In general, these value of N will depend on the specific system, corresponding to situations in which some common approximations (e.g. Laplace approximation for exponential integrals) can be safely applied. It is easy to understand that the convexity assumption implies the validity of the second principle of thermodynamics, as discussed in the next subsection.

1.4.1 Second law and energy flux between two systems in contact

Let us consider a system \mathcal{A} of $N_{\mathcal{A}}$ particles described by the variables $\{\mathbf{q}_{\mathcal{A}}, \mathbf{p}_{\mathcal{A}}\}$ and hamiltonian $H_{\mathcal{A}}(\mathbf{q}_{\mathcal{A}}, \mathbf{p}_{\mathcal{A}})$, a system \mathcal{B} of $N_{\mathcal{B}}$ particles described by the variables $\{\mathbf{q}_{\mathcal{B}}, \mathbf{p}_{\mathcal{B}}\}$ and hamiltonian $H_{\mathcal{B}}(\mathbf{q}_{\mathcal{B}}, \mathbf{p}_{\mathcal{B}})$ and a small coupling among the two, so that the global hamiltonian is

$$H = H_{\mathcal{A}}(\mathbf{q}_{\mathcal{A}}, \mathbf{p}_{\mathcal{A}}) + H_{\mathcal{B}}(\mathbf{q}_{\mathcal{B}}, \mathbf{p}_{\mathcal{B}}) + H_I(\mathbf{q}_{\mathcal{A}}, \mathbf{q}_{\mathcal{B}}). \quad (1.46)$$

If the two hamiltonians have the same functional dependencies on the canonical variables (i.e. they correspond to systems with same microscopic dynamics, with possibly different sizes $N_{\mathcal{A}}$ and $N_{\mathcal{B}}$), for large N , we can introduce the (Boltzmann) entropy per particle

$$S_B(E, N) = NS(e) \quad , \quad e = \frac{E}{N}, \quad (1.47)$$

with $S(e)$ a convex function, identical for systems \mathcal{A} and \mathcal{B} . Let us now suppose that systems \mathcal{A} and \mathcal{B} have, respectively, energy $E_{\mathcal{A}} = N_{\mathcal{A}}e_{\mathcal{A}}$ and $E_{\mathcal{B}} = N_{\mathcal{B}}e_{\mathcal{B}}$ and the corresponding inverse Boltzmann temperatures $\beta_B^{(\mathcal{A})}$ and $\beta_B^{(\mathcal{B})}$.

When the two systems are put in contact, a new system is realized with $N = N_{\mathcal{A}} + N_{\mathcal{B}}$ particles. Let us call $a = N_{\mathcal{A}}/N$ the fraction of particles from the system \mathcal{A} . We have that the final energy is $E_f = E_{\mathcal{A}} + E_{\mathcal{B}} = Ne_f$, where $e_f = ae_{\mathcal{A}} + (1 - a)e_{\mathcal{B}}$ and final entropy

$$S_B(E_f, N) = NS(e_f) \geq N_{\mathcal{A}}S(e_{\mathcal{A}}) + N_{\mathcal{B}}S(e_{\mathcal{B}}) = N[aS(e_{\mathcal{A}}) + (1 - a)S(e_{\mathcal{B}})]. \quad (1.48)$$

The previous inequality follows from the convexity assumption for $S(e)$ which implies

$$S(ae_{\mathcal{A}} + (1 - a)e_{\mathcal{B}}) \geq aS(e_{\mathcal{A}}) + (1 - a)S(e_{\mathcal{B}}). \quad (1.49)$$

⁹ It is interesting to notice that Kubo in [62] uses the adjective “normal” for systems satisfying $\Sigma(E, N) \sim e^{N\phi(E/N)+o(N)}$. It is easy to verify that for such systems one has $\beta_G = \beta_B + O(1/N)$. However our assumption is different: we ask that, in the large N limit, $\omega(E, N) \sim e^{N\psi(E/N)+o(N)}$. Since $\Sigma(E, N) = \int^E \omega(E')dE'$, a simple steepest descend computation shows that, if $d\psi(E'/N)/dE' > 0$ for $E' < E$, then $\psi(E/N) = \phi(E/N)$: this is equivalent to say that $T_B = T_G$ in the thermodynamic limit (i.e. up to $O(1/N)$) whenever $T_B > 0$ for an example). On the other hand if ψ has a maximum at E^* then $\Sigma(E, N)$ is roughly constant for $E > E^*$. In summary, for “normal” systems the temperatures must coincide, while with our assumption, one can have different temperatures in the region $E > E^*$. Note also that normal systems also satisfy our assumption, while the opposite is not true. Moreover, even if not all the systems satisfying our assumption could be named “normal”, all of them satisfy the equivalence of ensembles in the thermodynamic limit (as discussed below).

The final inverse temperature $\beta_B^{(f)}$ is intermediate between $\beta_B^{(A)}$ and $\beta_B^{(B)}$, e.g. if $e_B > e_A$ – that is $\beta_B^{(A)} > \beta_B^{(B)}$ – then

$$\beta_B^{(B)} < \beta_B^{(f)} < \beta_B^{(A)}. \quad (1.50)$$

The energy flux obviously goes from smaller β_B (hotter) to larger β_B (colder). The consequence of convexity is that $\beta_B(E)$ is always decreasing and a negative value does not lead to any ambiguity. Confusion may arise from the fact that $T_B < 0$ is, for the purpose of establishing the energy flux, hotter than $T_B > 0$. However if β_B is used, the confusion is totally removed [85].

We also briefly discuss a particularly interesting case with different hamiltonians. Suppose that for the system \mathcal{A} negative temperatures can be present, whereas system \mathcal{B} has only positive temperatures; it is quite easy to see that the coupling of the system \mathcal{A} at negative temperature with the system \mathcal{B} at positive temperature always produces a system with final positive temperature. Indeed, at the initial time the total entropy is

$$S_I = S^{\mathcal{A}}(E_{\mathcal{A}}) + S^{\mathcal{B}}(E_{\mathcal{B}}), \quad (1.51)$$

while, after the coupling, it will be

$$S_F = S^{\mathcal{A}}(E'_{\mathcal{A}}) + S^{\mathcal{B}}(E'_{\mathcal{B}}), \quad (1.52)$$

where $E'_{\mathcal{A}} + E'_{\mathcal{B}} = E_{\mathcal{A}} + E_{\mathcal{B}}$ and, within our assumptions, $E'_{\mathcal{A}}$ is determined by the equilibrium condition [51] that S_F takes the maximum possible value, i.e.

$$\beta_{\mathcal{A}} = \frac{\partial S^{\mathcal{A}}(E'_{\mathcal{A}})}{\partial E'_{\mathcal{A}}} = \beta_{\mathcal{B}} = \frac{\partial S^{\mathcal{B}}(E'_{\mathcal{B}})}{\partial E'_{\mathcal{B}}}. \quad (1.53)$$

Since $\beta_{\mathcal{B}}$ is positive for every value of $E'_{\mathcal{B}}$, the final common temperature must also be positive. The above conclusion can also be found, without a detailed reasoning, in some textbooks [63, 15].

1.4.2 Subsystems

Let us revisit in some detail the derivation of the canonical ensemble from the microcanonical one: consider a vector \mathbf{X} in \mathbb{R}^{2dN_1} (with $N_1 < N$), that is a subsystem of the full phase space (\mathbf{q}, \mathbf{p}) , and let us indicate with $\tilde{\mathbf{X}}$ in $\mathbb{R}^{2d(N-N_1)}$ the remaining variables. We have

$$H = H_1(\mathbf{X}) + H_2(\tilde{\mathbf{X}}) + H_I(\mathbf{X}, \tilde{\mathbf{X}}) \quad (1.54)$$

with an obvious meaning of symbols.

Let us consider the case $N \gg 1$ and $N_1 \ll N$. In the microcanonical ensemble with energy E , the probability density function (pdf) for the full phase space (\mathbf{q}, \mathbf{p}) is

$$P(\mathbf{q}, \mathbf{p}) = \frac{1}{\omega(E, N)} \delta(H(\mathbf{q}, \mathbf{p}) - E). \quad (1.55)$$

The pdf of \mathbf{X} can be obtained from the latter, by integrating over $\tilde{\mathbf{X}}$. If the hamiltonian $H_I(\mathbf{X}, \tilde{\mathbf{X}})$ is negligible (a consequence of our assumption for non long-range interaction) then we have

$$P(\mathbf{X}) \simeq \frac{\omega(E - H_1(\mathbf{X}), N - N_1)}{\omega(E, N)}. \quad (1.56)$$

1.4 Isolated small systems and the problem of the definition of temperature 27

It is now possible to exploit the definition of S_B and get

$$\omega(E, N) = e^{S_B(E, N)} \quad (1.57)$$

$$\omega(E - H_1(\mathbf{X}), N - N_1) = e^{S_B(E - H_1(\mathbf{X}), N - N_1)} \propto e^{S_B(E, N - N_1) - \beta_B(E)H_1(\mathbf{X})}, \quad (1.58)$$

which, together with (1.56) leads to

$$P(\mathbf{X}) \propto e^{-\beta_B H_1(\mathbf{X})}. \quad (1.59)$$

When H_1 is bounded (as in our assumptions), the previous simple derivation can be done irrespective of the sign of β_B . It is immediately clear from the above argument that T_B is the temperature ruling the statistics of fluctuations of physical observables in a subsystem. For instance, the pdf of the subsystem (i.e. the canonical ensemble) energy E_1 reads

$$P(E_1, N_1) \propto \omega(E_1, N_1) e^{-\beta_B E_1} \propto e^{[S_B(E_1, N_1) - \beta_B E_1]}. \quad (1.60)$$

Of course the above result holds in the (important) case where the two subsystems are weakly interacting and $H_1 \ll E$. Therefore, for $e_1 = E_1/N_1$, one has

$$P(e_1, N_1) \propto e^{N_1[S(e_1) - \beta_B e_1]}, \quad (1.61)$$

which is a large deviation law where the Cramer's function $C(e_1)$ is $C(e_1) = \beta_B e_1 - S(e_1) + \text{const}$. From general arguments of theory of probability, we know that - if a large deviation principle holds - $\frac{d^2 C(e_1)}{de_1^2} \geq 0$ so $\frac{d^2 S(e_1)}{de_1^2} \leq 0$. The validity of the large deviation principle can be easily shown for non-interacting systems with $N \gg 1$. For weakly interacting systems it is quite common and reasonable, and can be stated under rigorous hypothesis [98, 105]. The precise size of the system is in general not specified: nonetheless it is safe to assume that, in many systems of interest, a large deviation approximation can be carried out already for $N \sim (10 \div 100)$: in [Cerino et al., 2015a] a system with $N \sim 100$ was exhibited, for which all the above considerations applied.

1.4.3 The generalised Maxwell-Boltzmann distribution

The extreme case of the above considerations is when $N_1 = 1$, that is to say the fluctuations of a single degree of freedom (e.g. a momentum component of a single particle) are observed. This becomes interesting when the hamiltonian has the form

$$H = \sum_{n=1}^N g(p_n) + \sum_{n,k}^N V(q_n, q_k) \quad (1.62)$$

where the variables $\{p_n\}$ are limited and the same happens for the function $g(p)$.

Repeating the arguments in the previous subsection, one may compute the probability density for the distribution of a single momentum p , obtaining

$$P(p) \simeq \frac{\omega(E - g(p), N - 1)}{\omega(E, N)} \propto e^{-\beta_B g(p)}, \quad (1.63)$$

which, again, is valid for both positive and negative β_B . We mention that in the experiment in [9], the above recipe has been applied to measure both positive and negative system's temperatures.

From Eqs. (1.60) and (1.63) the true deep meaning of the (Boltzmann) temperature is quite transparent: it is a quantity which rules the pdf of energy of a subsystem (or the momentum of a single particle). Let us note that since T_B is associated to the large microcanonical system (in physical terms the reservoir) it is a non-fluctuating quantity [37] also for each sub-system and, in general, for non-isolated systems.

1.4.4 How to measure T_B and T_G

The definitions of β_B and β_G given in Eqs. (1.37) and (1.39) are based on the functional dependence of the phase space occupations $\omega(E)$ and $\Sigma(E)$ upon the energy. In a real or numerical experiment it may be cumbersome or even impossible to make use of those definitions to measure the two temperatures: for instance, an empirical estimate of $\omega(E)$ (and therefore of $\Sigma(E)$) will always be limited by the available statistics (number of independent measurements of E) and therefore cannot provide a clear answer, for both β_B and β_G , in the interesting regimes where $\omega(E) \sim 0$.

On the other hand it has been shown [87] that β_B can be obtained as a microcanonical average of a suitable observable. The recipe is the following

$$\beta_B = \langle R(\mathbf{X}) \rangle, \quad R(\mathbf{X}) = \nabla \cdot \frac{\nabla H}{|\nabla H|^2} \quad (1.64)$$

where ∇ stands for the vector of derivative operators along the degrees of freedom in the full phase space $\mathbf{X} \equiv (\mathbf{Q}, \mathbf{P})$. From (1.64) one has, assuming the ergodicity, that β_B can be computed with a molecular dynamics simulation, and, at least in principle, by a long-time series from an experiment. It is interesting to notice that such a kind of recipe does not exist for $S_B(E, N)$ or $S_G(E, N)$ [87]. It is clear that, in view of the considerations in Sections 1.4.2 and 1.4.3, one may always measure fluctuations of appropriate observables, such as subsystem's energy or single particle momentum, to get an estimate of T_B .

Coming to β_G , a way, even discussed in textbooks and considered sometimes rather important [50], to approach the problem of its measurement is via the equipartition theorem, which states

$$\left\langle x_i \frac{\partial H}{\partial x_j} \right\rangle = \delta_{ij} T_G. \quad (1.65)$$

However the usual derivation of Eq. (1.65) implies the possibility to neglect boundary terms in an integration by parts. Such a possibility is challenged in the class of systems with bounded energy and phase space that we are considering.

In particular it is easy to show that (1.65) does not hold under the simultaneous realization of the following conditions:

- bounded space of the canonical variables;
- bounded derivatives of the hamiltonian $\frac{\partial H}{\partial x_j}$;

1.4 Isolated small systems and the problem of the definition of temperature 29

- bounded energy from above and below: $E_m \leq E \leq E_M$;
- vanishing density of states at the boundaries, i.e. $\omega(E_M) = 0$.

Given such conditions, one has that, on one side,

$$T_G(E) = \frac{\Sigma(E)}{\omega(E)} \quad (1.66)$$

diverges when $E \rightarrow E_M$. On the other side, $\langle x_i \frac{\partial H}{\partial x_j} \rangle$ is limited, resulting in a contradiction.

A failure of the equipartition formula Eq. (1.65) is also possible in systems where there are no negative temperatures, i.e. $T_G \simeq T_B > 0$ for all E . Consider, for instance, the following hamiltonian

$$H = \sum_{n=1}^N \frac{p_n^2}{2} + \epsilon \sum_{n=1}^N (1 - \cos(\phi_n - \phi_{n-1})) \quad (1.67)$$

where $\phi_n \in [-\pi, \pi)$. For large E , i.e. $E \gg \epsilon N$, the contribution to $\Sigma(E)$ of the variables $\{\phi_n\}$ does not depend too much on the value of E , so that

$$\Sigma_\epsilon(E) \simeq \Sigma_0(E) \propto E^{N/2}, \quad (1.68)$$

and $T_G \simeq 2E/N$ and, for large N , $T_B = T_G + O(1/N)$.

On the other hand it is easy to see that

$$\left| \phi_n \frac{\partial H}{\partial \phi_n} \right| \leq 2\pi\epsilon, \quad (1.69)$$

and, therefore, the equipartition formula $\langle \phi_n \frac{\partial H}{\partial \phi_n} \rangle = T_G$ does not hold for large value of E and N .

1.4.5 Equivalence of ensembles and the equipartition formula

Let us briefly discuss the problem of the equivalence of ensembles. From a physical point of view it is possible to obtain the canonical ensemble from the microcanonical one only for large systems with short range interactions. In such a class of systems, if $N \gg 1$, the $S(e)$ is convex and it is easy to obtain the equivalence of the ensembles. Such a property is a fundamental requirement to obtain equilibrium thermodynamics, where there is no difference between thermostatted and isolated macroscopic systems. Assuming that $S(e)$ is convex and performing a steepest descent analysis, for large N , one obtains the canonical functions from the (Boltzmann) microcanonical ones, e.g.:

$$T_B(e)S(e) = e - f(T_B(e)), \quad (1.70)$$

where $f(T)$ is the free energy per particle in the canonical ensemble. In addition the energy fluctuations are negligible. In such a derivation, the relevant point is only the convexity of $S(e)$ and nothing about its first derivative is asked. Therefore, the equivalence of ensembles naturally holds under our hypothesis even for negative T_B . Since T_B and T_G can be different even for large N it is evident that T_G is not relevant for the ensemble equivalence: for this reason, in systems exhibiting negative temperatures, where S_B and S_G are no longer equivalent in the large N limit, thermodynamic can be recovered for $N \rightarrow \infty$ only through the Boltzmann formalism.

1.4.6 Can single-particle “thermodynamics” rule out the Boltzmann temperature?

In this Section we have given a series of arguments to support the thesis of the Boltzmann temperature T_B as a useful parameter to describe the statistical features of a system with many particles and short-range interactions, even when it takes negative values. Moreover, with this analysis, we confute the possibility of choosing an entropy over another by looking at single particle systems: it is clear from the above analysis that ruling out Boltzmann entropy on the basis of single or few particles systems analysis would exclude many actual thermodynamic behaviors of system with a larger number of components.

For short-range systems with large $N \gg 1$ we have two possible scenarios:

- In “standard system”, i.e. with a quadratic kinetic term in the hamiltonian, S_B and S_G converge for large N , therefore there is no need to choose one over the other.
- In “non-standard systems”, that could produce negative Boltzmann temperatures, we have shown that the temperature T_B is the proper quantity which describes the distribution of the energy fluctuations in the canonical ensemble. It also enters in an immediate generalization of the Maxwell-Boltzmann distribution to the case of “kinetic energy” which is not a quadratic function of momentum.

If the microcanonical entropy $S(e)$ is a convex function, independently of the sign of T_B , there is no ambiguity in determining the flux of energy: it always goes from the hotter system, i.e. with smaller β_B to the colder one (with larger β_B). It should be reminded that the convexity of $S(e)$ is a direct consequence of the presence of many degrees of freedom and can be violated only for very small systems or systems with long range interaction.

1.5 Conclusions

The above considerations should convince the reader that it is not possible to have a “small system thermodynamics” for isolated systems with very few particles $N \sim \mathcal{O}(1)$. In fact, in those systems, an analogy can only be sustained at a formal level with purely mathematical considerations. Let us summarize our conclusions:

- In few particle systems there is not distinction between “macro” and “micro” states: observables depending on the configuration of the whole system are not many-to-one functions. Most of the values of the collective variables correspond to regions of the space of parameters which have comparable volumes. A consequence of this, is that, due to the wildness of the fluctuations, it is not possible to distinguish between equilibrium (i.e. close to the average value) and non-equilibrium configurations.
- Ergodicity for these systems is a mere consequence of the periodicity of their trajectories. Therefore, even if it is possible to compute time-averages through phase-space averages, the cause of such an identification is completely

unrelated to the one responsible for the ergodicity of macroscopic systems, i.e. the presence of a large number of degrees of freedom.

- Few particles systems do not display irreversible behaviors. In chaotic system with a small number of degrees of freedom, the mixing property is an ensemble property and, thus, cannot be observed at the level of single trajectories. Irreversibility can emerge only as an average computed over many different realizations of the experiment: this is unrelated to the genuine thermodynamic irreversibility which can be observed in a system evolving from a non-equilibrium configuration.
- Very small systems present an ambiguity in the choice of the correct definition of the microcanonical entropy: this is due to the fact that many thermodynamic properties (e.g. the convexity of the entropies) do not hold when N is very small. Other examples of systems that do not present convex entropy are long-range systems, very well known examples where the tools of statistical mechanics do not allow to get thermodynamics from microscopic dynamics. In large standard systems, the differences between the two possible definitions of entropy (Boltzmann and Gibbs) vanish for large N ; in non-standard systems (with non monotonically increasing entropies) the choice of the Boltzmann entropy is justified by the many properties that it possesses for large N systems.

On the other hand, there exists two different classes of systems that preserve many thermodynamic properties even when composed by a small number of components.

1. Stochastic systems, that can be obtained, under some rather reasonable hypothesis, as a projection of the state of a large system onto a small set of macroscopic observables.
2. Many-particles small systems (MPSS), that, as showed above, possess all the fundamental properties of macroscopic systems and, in addition, have large, non-negligible fluctuations and non-trivial equilibrium states. This class of systems, which is the one usually considered in numerical experiments, constitutes the main interest of this Thesis.

Chapter 2

Irreversible and Stochastic Thermodynamics: a Brief Review

Thermodynamics, at its origins, received a crucial impulse from the study of heat engines [18]. It is interesting to realize that - after almost two centuries - engines still represent a relevant driving force towards new developments in this science. A challenging frontier in thermodynamics is the world of small and fast systems, where the assumptions of “quasi-reversible” transformations and the thermodynamic limit of statistical mechanics are not valid [99, 90]. Obviously, “fast” thermodynamic transformations, i.e. those such that the typical cycle time τ is shorter than the slowest relaxation time of the system, constitute a key problem in the industry and, for this reason, have been under the scrutiny for many decades. In the 70’s of the last century, several results were obtained in the so-called finite time thermodynamics [1], especially in the analysis of the efficiency at maximum power: this quantity, that represents the efficiency of a motor that operates in an external condition such that the output power is maximum, is an important figure when trying to optimize the performance of an engine. Of course, the maximum power occurs far from the adiabatic, i.e. infinitely slow, limit, where the maximum possible efficiency, the Carnot efficiency, is reached. For this reason a finite-time analysis is necessary: in this context an estimate of the maximum possible value of efficiency at maximum power has been obtained by Curzon-Ahlborn (CA) [28]. Such an estimate has been revised in the recent years, with the introduction of new and more general classes of engines with respect to the original model considered by Curzon and Ahlborn [99, 36, 35]. A more recent wave of studies concerns the thermodynamics of systems with a “small” number N of degrees of freedom [60], motivated by the tremendous increase of resolution in the observation and in the manipulation of the micro-nano world, involving mainly biophysical systems and artificial machines [44]. The distinguishing feature of small systems is the relevance of fluctuations, which are negligible only when the number of constituents is very large, as for macroscopic bodies. The study of fluctuations of thermodynamics functions such as energy or entropy goes back to Einstein, Onsager and Kubo, but has recently taken an acceleration with the establishing of new results in response theory [71] and in the

so-called stochastic thermodynamics [93, 94]. Such a turning point concerns the properties of fluctuations in system which are *far* from thermodynamic equilibrium, and therefore is intimately connected to the previously mentioned problem of fast transformations. In a nutshell, modern stochastic thermodynamics addresses the finiteness of both transformation's time τ and system's size N .

In this Chapter we try to summarize most of the main results obtained in the last decades regarding the finite-time thermodynamics of small systems. In particular we will revisit the microscopic interpretation of the peculiar and characteristic quantities of thermodynamics, namely work, heat, entropy etc... Furthermore we will give a consistent illustration of the previously mentioned stochastic thermodynamics, i.e. the attempt to extend thermodynamic quantities to stochastic processes. In this context a notion of work, heat and entropy may be defined and applied at the level of Langevin equations involving very few (in general one) degrees of freedom. For this reason, Brownian motors, i.e. single-particle stochastic engines, have been deeply tested and investigated and many interesting properties found. Eventually, after having established a general framework for the study of micro and mesoscopic engines, we will discuss the efficiency at maximum power, the importance and utility of this quantity and we will retrace one of the derivations leading to the upper bound to this quantity, the Curzon-Ahlborn efficiency. In the Appendices we will briefly explore the realm of the fluctuation theorem, by deriving some of the fundamental equations necessary to derive some properties of the observables we are interested in.

2.1 On the definition of microscopic work and heat

In recent years, i.e. since when the thermodynamic jargon has been used to describe micro and mesoscopic situations, the problem emerged on the correct microscopic definition of work and heat. In particular, as we will show in the rest of this section, the naïve definition $dW = pdV$ is not appropriate in a large variety of situations, e.g. when the external varied parameter is not the volume. In fact this definition comes from a basic application of the classical definition of work ($dW = \mathbf{F} \cdot d\mathbf{x}$) and pressure ($\mathbf{F} \cdot \hat{\mathbf{n}} = pS$, where S is the surface): in systems exceeding the usual applications of thermodynamics the definition must be reconsidered. We will now discuss the definition in the context of Hamiltonian systems.

Consider a time-dependent Hamiltonian system with hamiltonian $\mathcal{H}(\mathbf{X}, t)$. Let $\{\tilde{\mathbf{X}}(t)\}_{t=0}^T$ be a solution of Hamilton equations of motion; then, the time derivative of a generic function $A(t) = A(\tilde{\mathbf{X}}(t), t)$.

$$\frac{dA(t)}{dt} = \frac{\partial A(\mathbf{X}, t)}{\partial \mathbf{X}} \dot{\tilde{\mathbf{X}}} + \frac{\partial A(\mathbf{X}, t)}{\partial t} = \{A, \mathcal{H}\} + \frac{\partial A}{\partial t}. \quad (2.1)$$

It is natural to identify the internal energy of the system with the value of the hamiltonian $E(t) = \mathcal{H}(\tilde{\mathbf{X}}(t), t)$ and compute its time derivative via Eq. (2.1)

$$\frac{dE(t)}{dt} = \{\mathcal{H}, \mathcal{H}\} + \frac{\partial \mathcal{H}}{\partial t} = \left. \frac{\partial \mathcal{H}}{\partial t} \right|_{\tilde{\mathbf{X}}(t)}. \quad (2.2)$$

Therefore, when the external parameter varies, the total energy of the system changes and the above equation represents its time derivative. It is not difficult

to see that, from a thermodynamic point of view, this quantity corresponds to the time derivative of the work $\dot{W}(t)$. In fact, an Hamiltonian system is a closed system, since it does not exchange energy with a thermostat: all the energy is transferred in an Hamiltonian manner with the (hidden) device responsible for the time variation of the hamiltonian. In thermodynamics, the energy exchange of closed systems is exactly the work produced or consumed by the system itself. This definition contains, of course, a certain degree of arbitrariness: nevertheless, this energy is the only one that can be converted into useful movement (for instance by an engine). Summarizing,

$$\dot{W} = \frac{\partial \mathcal{H}}{\partial t}. \quad (2.3)$$

It is important to remark that this convention implies that whenever the energy of the system increases the work is positive, whereas, when the opposite occurs, the work is negative: this is in contrast with the usual thermodynamic convention, but with a far more transparent energetic interpretation.

Suppose that the explicit time dependence of the hamiltonian is given by an additional term $\mathcal{H} = \mathcal{H}_0 + h(t)$, and that $h(t) = \lambda(t)X_i$, with X_i one of the component of the phase space variable \mathbf{X} ; one has

$$\dot{W}(t) = \tilde{X}_i(t)\dot{\lambda}(t), \quad (2.4)$$

equivalent to $dW = Vdp$: this is different from the standard definition $dW = -pdV$. The explanation was reported in Ref. [56]: assume that, generalizing the above situation, the hamiltonian depends upon a certain number of external parameters $\lambda_k(t)$, with $k = 1, \dots, M$:

$$\mathcal{H} = \mathcal{H}_0(\mathbf{X}) + \sum_k \lambda_k(t)g_k(\mathbf{x}). \quad (2.5)$$

An alternative definition of internal energy, that can be applied in this case, is $E(t) = \mathcal{H}_0(t)$: its time derivative can be readily computed and reads

$$\frac{d\mathcal{H}_0}{dt} = \{\mathcal{H}_0, \mathcal{H}\} = - \sum_k \lambda_k(t)\{g_k, \mathcal{H}\} = - \sum_k \lambda_k(t)\dot{g}_k(\mathbf{x}(t)). \quad (2.6)$$

Therefore, with this new convention consisting in a different separation of the hamiltonian in which all the time-dependent terms are considered *external* to the system of interest, the definition of work reads

$$\dot{W}' = - \sum_k \lambda_k(t)\dot{g}_k(\mathbf{x}(t)). \quad (2.7)$$

In particular, when $k = 1$ and $g(\mathbf{x}) = X_i$, we recover the usual thermodynamic definition of work

$$\dot{W}(t) = -\lambda(t)\dot{X}_i, \quad (2.8)$$

that is equivalent to $dW = -pdV$. It is important to remark that this definition gives a nonzero work also when the external parameters are fixed, $\dot{\lambda} = 0$: this occurs because some energy may be exchanged between the internal energy \mathcal{H}_0 and the external terms appearing in the hamiltonian. This happens despite the fact that

the value of the hamiltonian $\mathcal{H}(\mathbf{X}(t))$ does not change over time. In the rest of this work we will use the first definition, Eq. (2.3).

It is not possible to define the heat exchange in the context of purely Hamiltonian systems, since these systems do not transfer energy with the external environment. Moreover, in general, there are several different ways to define the interaction of a system with an external thermostat at temperature T (stochastic, deterministic, etc...). For this reason, to keep this discussion on general terms, it is important to give a definition of heat that does not rely on the specificity of the model. In the next Section we will discuss this topic in the context of coarse-grained stochastic differential equations.

When a time-dependent Hamiltonian system is coupled to an external thermostat the variation of the total energy $E(t) = \mathcal{H}(\mathbf{X}(t), t)$ is due to different causes: indeed, a part of energy is funneled through the external parameter that varies in time, whereas another part is exchanged with the thermostat to which the system is attached. It is customary to denote with the term heat all the energy that is not exchanged in an ‘‘Hamiltonian’’ manner, i.e.

$$\dot{Q}(t) = \dot{E}(t) - \dot{W}(t), \quad (2.9)$$

where $\dot{W}(t)$ is the quantity defined in Eq. (2.3). This last equation is the microscopic equivalent of the first principle of thermodynamics, when the appropriate sign convention is chosen. From a practical point of view, there may be some ambiguity in the definition of \dot{Q} in the context of stochastic differential equations: in particular, as we will see later, it is very important to specify whether the derivative should be taken with the Itô or the Stratonovich convention. To avoid such ambiguities it can be useful to report the integrated version of Eq. (2.9):

$$Q(\mathcal{T}) = \Delta E - \int_0^{\mathcal{T}} \frac{\partial \mathcal{H}(t)}{\partial T} dt, \quad (2.10)$$

where \mathcal{T} is the total time of the measurement, and $\Delta E = E(\mathcal{T}) - E(0)$ is the total energy variation in such an interval. This last equation does not present any ambiguity, since the integration variable of the integral appearing on the r.h.s. is the time t (and not the phase space position \mathbf{X}), and therefore $E(t)$ is a well-defined function of time.

The context in which we are going to study these quantities are periodic engines, i.e. systems where the external parameters are periodic functions of time $\lambda_k(t) = \lambda_k(t + \mathcal{T})$: it can then be interesting to evaluate the average values of W and Q measured over a period (or cycle). From a practical point of view, if we analyze a long evolution of the system during a time $\tau = K\mathcal{T}$, with $K \gg 1$ we can define

$$\langle W \rangle = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{i=1}^K W^{(i)}, \quad (2.11)$$

with $W^{(i)}$ the work measured in the i -th cycle (the analogous definition holds for $\langle Q \rangle$),

$$W^{(i)} = \int_{i\mathcal{T}}^{(i+1)\mathcal{T}} \frac{\partial \mathcal{H}}{\partial t}(\mathbf{X}(t)) dt. \quad (2.12)$$

In order to express this quantity through phase space averages, it is possible to assume that, due to the time periodicity of the functions λ_k , there exists a periodic time dependent phase space distribution:

$$\rho^{inv}(\mathbf{X}, t) = \rho^{inv}(\mathbf{X}, t + \mathcal{T}), \quad (2.13)$$

where the evolution of the distribution function is determined by the Liouville equation for Hamiltonian system and the Fokker-Planck equation in stochastic systems. With this assumption, it is possible to write down the average work as

$$\langle W \rangle(\mathcal{T}) = \int_0^{\mathcal{T}} dt \int d\mathbf{X} \rho^{inv}(\mathbf{X}, t) \dot{W}(\mathbf{X}, t). \quad (2.14)$$

It is not possible to obtain a rigorous proof of the complete equivalence of Equations (2.11) and (2.14): nevertheless such an identification is reasonable thanks to arguments similar to the ones leading to the law of large numbers. Since the energy $E(t)$ only depends upon the state of the system $\mathbf{X}(t)$, it is immediate to see that $\langle \Delta E \rangle = 0$: as a consequence,

$$\langle Q \rangle(\mathcal{T}) = -\langle W \rangle(\mathcal{T}). \quad (2.15)$$

In view of the fact that we will define an efficiency for molecular engines it is important to split, when the temperature T of the thermostat varies in time, the heat Q into two contributions Q_{in} and Q_{out} , such that $Q = Q_{in} + Q_{out}$. The two quantities represent, respectively, the heat absorbed from the reservoir(s) and the heat released: this splitting is straightforward when the temperature T only assumes two values, $T_H > T_C$, i.e. $Q_{in} = Q_H$ and $Q_{out} = Q_C$. When the time dependence of the temperature is more complicated, the identification is not trivial and, moreover, may not be unique. A general recipe, based on the definition of entropy production, will be discussed in the following section, and an application to a practical example will be exhibited in the second part of the Thesis.

2.2 Stochastic Thermodynamics

In this Section we introduce the so-called *stochastic thermodynamics*, i.e. that discipline that studies how thermodynamic concepts can be applied in the context of stochastic differential equations. In particular, following Ref. [94], we will examine the simple case of a unidimensional brownian particle at environmental temperature T moving in an external time dependent potential $V(x, t)$. The stochastic differential equation describing this system in the underdamped regime reads

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= -\gamma v - \frac{\partial_x V(x, t)}{m} + \sqrt{\frac{2k_B T \gamma}{m}} \eta \end{aligned} \quad (2.16)$$

where η is white noise $\langle \eta(t)\eta(t') \rangle = \delta(t - t')$, γ a constant parameter and k_B the Boltzmann constant.

2.2.1 Work and Heat in the SDE formalism

A very natural choice for the energy of the particle $E(t)$ is

$$E(t) = \frac{1}{2}mv(t)^2 + V(x(t), t), \quad (2.17)$$

and, with considerations analogous to the ones discussed in the previous section, we can define the work performed on the system as

$$\dot{W}(t) = \left. \frac{\partial V(x, t)}{\partial t} \right|_{x=x(t)}. \quad (2.18)$$

Those definitions, despite their apparent coherence, conceal some inconsistencies due to the fact that Eq. (2.16) is a coarse grained equation.

It is very easy to show that, in the simple case when $\partial_t V = 0$, the (equilibrium) invariant distribution of the stochastic system is

$$\rho(x, v) \propto \exp \left\{ -\beta \left(\frac{m}{2}v^2 + V(x) \right) \right\}. \quad (2.19)$$

On the other hand, when considering x as a single component of a much larger Hamiltonian system $\mathbf{X} = (x, v, x_1, v_1, x_2, v_2, \dots, x_N, v_N)$ with hamiltonian $\mathcal{H}(\mathbf{X})$ we have, at equilibrium,

$$\rho^{eq}(x, v) = \int \frac{dx_1 dv_1 \dots dx_N dv_N}{Z} e^{-\beta \mathcal{H}(\mathbf{X})}, \quad (2.20)$$

where Z is the partition function and $\beta = 1/T$. Whenever the hamiltonian can be split into the sum of a kinetic term of the particle and another part involving all the remaining degrees of freedom of the system,

$$\mathcal{H}(\mathbf{X}) = \frac{mv^2}{2} + \mathcal{H}_{int}(x, x_1, v_1, \dots, x_N, v_N), \quad (2.21)$$

it is possible to recast Eq. (2.20) in the following form

$$\rho^{eq}(x, v) = \frac{Z(x)}{Z_0} e^{-\beta m \frac{v^2}{2}}, \quad (2.22)$$

where $Z(x) = \int \exp\{-\beta \mathcal{H}_{int}\} dx_1 dv_1 \dots dx_N dv_N$, and Z_0 the normalizing constant. Therefore, by comparing the last expression with Eq. (2.19), it is easy to see that

$$V(x) = -\frac{1}{\beta} \ln Z(x), \quad (2.23)$$

which is interpreted by many authors, like [94], as a conditional free energy function rather than a proper potential energy.

When the hamiltonian \mathcal{H} explicitly depends upon the time t , also V will have a time-dependency: suppose for simplicity that the time derivative of the hamiltonian $\partial_t \mathcal{H}(\mathbf{X}, t)$ only depend upon the variable x . Even in this case, the two definition of work, i.e. the hamiltonian one $\dot{W}_H(x) = \partial_t \mathcal{H}(\mathbf{X}, t)$ and the stochastic one $\dot{W}_S(x) = \partial_t V(x, t)$ may give different results: an example of this issue will be discussed in the second part of the Thesis.

A rule must be given in order to choose the correct definition of work: in general we prefer to use the microscopic definition whenever we have an explicit description of the model from which the stochastic equation is obtained. In all the other cases, i.e. when the model is directly defined through an SDE, the stochastic equivalent must be adopted.

When Eq. (2.18) is adopted, heat Q is the difference between energy and work (first principle of thermodynamics). An explicit formula for heat can be readily obtained:

$$\begin{aligned}\dot{Q} &= \frac{dE}{dt} - \dot{W} = \\ &= \partial_v E \dot{v} + \partial_x E \dot{x} = mv \left(-\gamma v - \frac{\partial_x V}{m} + \sqrt{\frac{2\gamma k_B T}{m}} \eta \right) + v \partial_x V = \\ &= m \left(-\gamma v + \sqrt{\frac{2\gamma k_B T}{m}} \eta \right) v, \quad (2.24)\end{aligned}$$

where, since we applied the usual calculus rules, the differential equation must be interpreted with the Stratonovich convention. Nevertheless, in order to compute its average value $\langle \dot{Q} \rangle$, it is useful to derive its Itô equivalent:

$$\dot{Q} = m \left(-\gamma v + \sqrt{\frac{2\gamma k_B T}{m}} \eta \right) v + \gamma k_B T. \quad (2.25)$$

Of course, the average of Eq. (2.25) and (2.24) must give the same result, the Itô expression being more explicit:

$$\langle \dot{Q} \rangle = -\gamma m \langle v^2 \rangle + k_B T \gamma = -2\gamma \left(\frac{m \langle v^2 \rangle}{2} - \frac{k_B T}{2} \right). \quad (2.26)$$

It is clear that the quantities \dot{Q} and \dot{W} , and the integrated heat and work are fluctuating quantities because they depend upon the single trajectories $\{x(t), v(t)\}_{t=0}^T$.

2.2.2 Work and heat in the Fokker-Planck formalism

The stochastic equation (2.16) gives a description of the system at level of single trajectories: naturally, an equivalent description can be given in terms of the probability distribution function $\rho(x, v, t)$, i.e. the probability density of finding the particle at time t with position x and velocity v . The time evolution of this quantity is given by a partial differential equation, the Fokker-Planck equation [43]:

$$\partial_t \rho(x, v, t) = -\nabla \cdot \mathbf{J}(x, v, t) \quad (2.27)$$

where the two components of the current $\mathbf{J} = (J_x, J_v)$ are, respectively

$$J_x = v \rho(x, v, t), \quad (2.28)$$

$$J_v = -\gamma v \rho(x, v, t) - \frac{\partial_x V}{m} \rho(x, v, t) - \frac{k_B T \gamma}{m} \partial_v \rho(x, v, t). \quad (2.29)$$

The average values of the thermodynamic quantities introduced above can be obtained by taking the time derivative of the average energy of the system, i.e.

$$\langle E(t) \rangle = \int dx dv \left(\frac{1}{2} m v^2 + V(x, t) \right) \rho(x, v, t), \quad (2.30)$$

that yields

$$\begin{aligned} \langle \dot{E} \rangle &= \int dx dv \partial_t V \rho(x, v, t) + \int dx dv E(x, v, t) \partial_t \rho(x, v, t) = \\ &= \langle \dot{W} \rangle + \langle \dot{Q} \rangle. \end{aligned} \quad (2.31)$$

The identification of the first term of the r.h.s. in the first line with the average mechanical power is immediate, since this quantity trivially coincides with Eq. (2.18). The other term of the sum is associated with the average heat rate, because, in agreement with the first principle of thermodynamics, the sum of the two terms must represent the total (internal) energy variation of the system. In order to show that this second definition, i.e.

$$\langle \dot{Q} \rangle = \int dx dv E(x, v, t) \partial_t \rho(x, v, t) \quad (2.32)$$

is consistent with Eq. (2.26), it is necessary to perform some algebra:

$$\begin{aligned} \langle \dot{Q} \rangle &= - \int dx dv E(x, v, t) \nabla \cdot \mathbf{J}(x, v, t) = \int dx dv \nabla E(x, v, t) \cdot \mathbf{J}(x, v, t) = \\ &= \int dx dv (m v J_v + \partial_x V J_x) = \int dx dv \left(-m \gamma v^2 p - v \partial_x V p - k_B T \gamma v \partial_v p + v \partial_x V p \right) = \\ &= \int dx dv (-\gamma v) (m v p + k_B T \partial_v p) = \int dx dv (-\gamma v) (m v + k_B T \partial_v \ln p) \rho(x, v, t) \end{aligned} \quad (2.33)$$

By performing one last integration by parts, one immediately gets Eq. (2.26) from the above expression. It is very important to remark that it is very useful to have two different formalisms describing the stochastic thermodynamics of one-particle systems: in fact on one hand Eqs. (2.18) and (2.25) gives a recipe on how to measure the work and heat exchanged on a single trajectory. Therefore, e.g. in numerical simulations where the state of the system is accessible at every time t , one can integrate the two expressions and obtain the (fluctuating) heat and work exchanged in a single realization of the experiment. On the other hand, the expressions involving the distribution $\rho(x, v, t)$ and its derivatives are more useful in an analytic context, and, as we will show in the following subsection, are necessary to obtain a connection between this quantities and the entropy production.

2.2.3 Connection with standard thermodynamics and entropy production

Standard thermodynamics, i.e. the one that we learn on textbooks, is intrinsically an equilibrium theory: nonequilibrium steady-states are in general not considered (or reduced to continuous systems in which local equilibrium is assumed) and

nonequilibrium transformations always connect two equilibrium states. One of the consequence of this is that the entropy $S(A)$ of a thermodynamic state A is a function that makes sense only if A is a set of thermodynamic variables identifying an equilibrium state. In particular, the entropy is a state function obtained through the formula

$$\Delta S = S(B) - S(A) = \int_A^B \frac{dQ}{T}, \quad (2.34)$$

where dQ is the heat exchanged with the exterior of the system, T the temperature of the system on any quasi-static transformation, i.e. a generic path consisting in a succession of equilibrium states that connects the state A to the state B . Neither dQ nor T represent the actual heat exchanged or temperature of the system. The second principle of thermodynamics, in the form of Clausius inequality reads

$$\oint \frac{dQ}{T} \leq 0, \quad (2.35)$$

i.e. $\Delta S \geq \int_B^A \frac{dQ}{T}$, and the equivalence only holds in the case of quasi-static, i.e. equilibrium transformations. Assuming that the system is attached to a thermostat at constant temperature T , when considering the system of interest and the thermostat attached to it, one simply gets

$$\Delta S_{tot} = \Delta S_{sys} + \Delta S_{therm} \geq 0 \quad (2.36)$$

or, equivalently

$$\Delta S_{sys} = -\Delta S_{therm} + \Sigma, \quad (2.37)$$

where Σ is a positive quantity which, with an evident meaning, is called total entropy production (it is the increase of entropy in an isolated system due only to the irreversibility of the dynamics), and the other term of the sum is

$$\Delta S_{therm} = -\frac{Q}{T}, \quad (2.38)$$

where Q is the heat *absorbed* by the system $Q = \int_A^B dQ$. This last quantity has been called in recent years *entropy production of the medium* [92], since it is the variation of the entropy of the thermostat.

We will now enter in some of the details of the reasoning allowing to connect stochastic thermodynamics and standard thermodynamics: even if, from an historical point of view it is not easy to determine the precise origin of this theory (this ideas were implicit in the pioneering works of Gallavotti-Cohen [42] and Lebowitz-Spohn [66] in the 1990s), it is customary to assume Refs. [92, 94] as its starting point. In order to obtain an entropy equivalent function in the context of stochastic thermodynamics it is necessary to introduce the Gibbs, or Shannon, entropy

$$S(t) = - \int dx dv \rho(x, v, t) \ln \rho(x, v, t) \quad (2.39)$$

This definition has a very long history in the context of statistical mechanics: it is immediate to see that it reproduces the correct entropy definition when the probability distribution is the equilibrium distribution both in the canonical and

microcanonical ensemble. On the other hand, it is also well known that, whenever the dynamics is conservative, the value of the Gibbs entropy is always constant, for every initial distribution $\rho(x, v, 0)$: for this reason this entropy is not appropriate to describe some irreversible processes like the free expansion of a gas in a container (see the discussion on irreversibility in Sec. 1.3.1 and in [Cerino et al., 2016a]).

At this stage it is possible to compute the derivative of the Gibbs entropy and express it as a sum of different terms:

$$\begin{aligned} \frac{dS(t)}{dt} &= - \int dx dv (\partial_t \rho(x, v, t) \ln \rho(x, v, t) + \partial_t \rho) = \\ &= - \int dx dv \frac{\mathbf{J}}{\rho} \cdot \nabla \rho = - \int dx dv \left(v \partial_x \rho - \gamma v \partial_v \rho - \frac{\partial_x V}{m} \partial_v \rho - \frac{k_B T \gamma}{m} \frac{(\partial_v \rho)^2}{\rho} \right) = \\ &= \int dx dv \left(\gamma v \partial_v \rho + \frac{k_B T \gamma}{m} \frac{(\partial_v \rho)^2}{\rho} \right), \end{aligned} \quad (2.40)$$

where, in the last line, we used the fact that many terms cancels with an integration by parts. Carrying on the calculations further one gets

$$\begin{aligned} \frac{dS(t)}{dt} &= \int dx dv \left(2\gamma v \partial_v \rho + \frac{k_B T \gamma}{m} \frac{(\partial_v \rho)^2}{\rho} \right) - \int dx dv \gamma v \partial_v \rho = \\ &= \int dx dv \frac{1}{\rho} \left(\sqrt{\frac{k_B T \gamma}{m}} \partial_v \rho + \sqrt{\frac{\gamma m}{k_B T}} v \rho \right)^2 - \frac{1}{k_B T} \int dx dv (\gamma m v^2 \rho + \gamma v k_B T \partial_v \rho) = \\ &= A(t) + \frac{\langle \dot{Q} \rangle}{k_B T} \end{aligned} \quad (2.41)$$

where $A(t)$ is an evident always positive quantity. The identification of the two contributions with their thermodynamic equivalents is immediate: $S(t)$ is the entropy production of the system, $A(t)$ is the total entropy production $\dot{A}(t) = \dot{\Sigma}$ and the last term is the entropy production of the medium \dot{S}_{therm} .

This calculation also allows to obtain a single trajectory, i.e. fluctuating, entropy variation of the system. The state-dependent entropy of the system is

$$s(x, t) = - \ln \rho(x, v, t), \quad (2.42)$$

since the average of this quantity gives the Gibbs entropy in Eq. (2.39). This quantity can be measured once the state of the system x and the probability distribution function $\rho(x, v, t)$, i.e. the solution of the Fokker-Planck equation, are known: for instance, in every trajectory that originates at time t_0 in (x_0, v_0) and finishes at time t_f in (x_f, v_f) one can measure the entropy difference:

$$\Delta s = - \ln \frac{\rho(x_f, v_f, t_f)}{\rho(x_0, v_0, t_0)}. \quad (2.43)$$

Averaging on all the possible trajectories one gets $\langle \Delta s \rangle = \Delta S$, where S is the Gibbs entropy, which do not depend upon the state of the system. Moreover, by computing the sum of the two terms,

$$\dot{\Sigma}(x, v, t) = \dot{s}(x, v, t) - \frac{\dot{Q}}{T}(x, v, t), \quad (2.44)$$

it is possible to recover a quantity, $\dot{\Sigma}$, which is the single-trajectory equivalent of the total entropy production: the value of this observable depend upon the single realization of the experiment, while $\langle \Sigma \rangle$ is positive on average. This observation constitutes a probabilistic interpretation of the second principle of the thermodynamics: from the point of view of the single realization, there may happen “violations” of the second principle, i.e. trajectories on which the entropy production is negative, but on average this occurrences must be compensated in order to give a positive average. Another confirm of the above observation comes from the so-called fluctuation relations: a fluctuation relation is an equation that, given an observable \mathcal{W} relates the probability of measuring $\mathcal{W} = X$ and $\mathcal{W} = -X$,

$$\frac{P(\mathcal{W} = X)}{P'(\mathcal{W} = -X)} = e^X, \quad (2.45)$$

where P and P' are the probabilities of a given trajectory and its time-reversed counterpart. A detailed proof of the fluctuation relations in some physically relevant systems and their implications are discussed in Appendix 2.A. For the moment we just want to emphasize that an immediate consequence of Eq. (2.45) is that the average of the exponential \mathcal{W} is equal to 1, $\langle \exp\{-\mathcal{W}\} \rangle = 1$, and that, by applying the Jensen equality

$$\langle \mathcal{W} \rangle \geq 0. \quad (2.46)$$

In the Appendix it also proven that, in many interesting situations, including the ones described in this section and in the previous one, the quantity \mathcal{W} is the total entropy production $\dot{\Sigma}$, the last equation being the equivalent of the second principle of thermodynamics. In particular it is possible to give an alternate, single trajectory definition of the entropy production of the medium:

$$\frac{Q(t)}{T} = \ln \frac{P(\{x(s)\}_{s=0}^t | x_0)}{P'(\{\mathcal{R}x(t-s)\}_{s=0}^t | x_t)} \quad (2.47)$$

where the two probabilities above represent the probability of observing a forward and a reversed trajectory. Such probabilities can be obtained, for instance, via the action functional in the Onsager-Machlup formalism, see Ref. [70, 52].

2.3 Heat engines and linear irreversible thermodynamics

In the second half of the nineteenth century, the attempt of designing and realizing a maximally efficiency engine was one of the main reasons pushing physicists to “invent” the branch of physics that is now known as thermodynamics. An engine consists in a physical system that performs cyclical transformations from which, by exchanging heat with a certain number of thermostats (or reservoirs) at different temperatures, some mechanical work is extracted.

In order to establish the notation and to state the problem with a certain degree of generality, we refer to Fig. 2.1. The engine state is identified by its internal energy $E(t)$: the energy of the system may be exchanged through two different channels, heat $Q(t)$ and work $W(t)$. As already stated, the work W is mechanical energy which

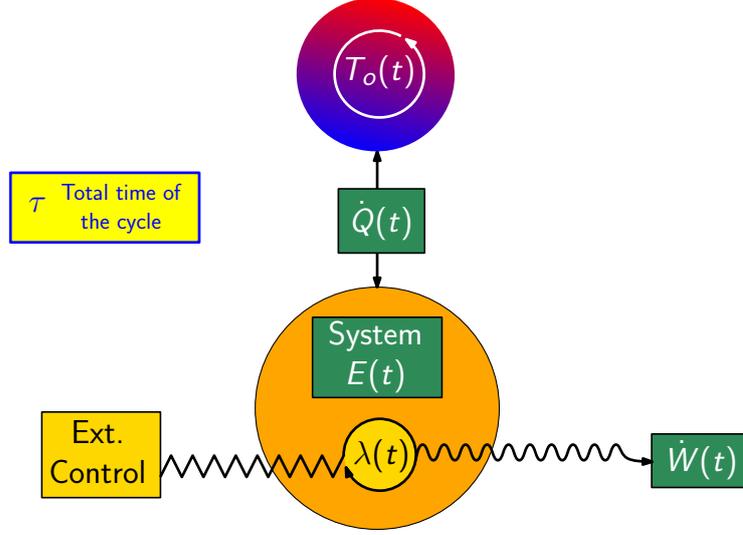


Figure 2.1. A scheme representing the working principle of an heat engine, i.e. a device characterized by an internal energy E that mechanically interacts with the exterior via the parameter $\lambda(t)$ and that exchanges heat with several thermostats at different temperatures $T_0(t)$.

is funneled through a time-varying parameter $\lambda(t)$, while the heat Q is the energy absorbed from or given to a certain number of thermostats, each one at a different temperature T . In order to include also the case where the system is attached to a single thermostat whose temperature continuously varies in time, we may assume, without losing generality, that there is a single thermostat whose temperature is a time-varying parameter $T_0(t)$. Naturally, in order to have a cyclical engine, one must assume that all the time-varying parameters appearing in the description must be periodic functions of time of period τ . When this is the case one must have that, on average, the state of the system does not change if the system is observed every τ : $\langle E \rangle(t) = \langle E \rangle(t + \tau)$. In addition, the first principle of thermodynamics holds, i.e.

$$W(\Delta t) + Q(\Delta t) = E(t + \Delta t) - E(t), \quad (2.48)$$

where W and Q are the total work and heat integrated over a generic time interval Δt . Whenever $\Delta t = \tau$, due to the periodicity of the transformation, this last equation reduces to

$$\langle W \rangle(\tau) + \langle Q \rangle(\tau) = 0. \quad (2.49)$$

We remark that the average symbol $\langle \cdot \rangle$ means that the desired quantity, defined over a single cycle, must be measured in a long experiment where the cycle is repeated a large number of times: in this way, if the system reaches a stationary periodic asymptotic state, this average must coincide with one computed with respect to the invariant periodic distribution of the system, Eq. (2.13).

We now want to address the most common example of an engine, i.e. the one in which the system is alternatively attached to two different thermostats, respectively at the temperatures T_H and T_C , with $T_H > T_C$: we will denote with Q_H and Q_C , respectively, the heat exchanged with the hot and the cold reservoir, such that

$\dot{Q}_H + \dot{Q}_C = \dot{Q}$. In the previous Section we showed that for a vast class of systems, i.e. all the ones for which a fluctuation theorem holds, the total entropy production Σ is a quantity that, on average, is non-negative. Furthermore, see the Appendix at the end of the Chapter, under some rather general conditions the entropy production of such an engine can be expressed as

$$\Sigma = \Delta s_{sys} - \frac{Q_H}{T_H} - \frac{Q_C}{T_C}, \quad (2.50)$$

where Δs_{sys} is the entropy variation of the system. By taking the average, and using the periodicity of the transformation and, hence, the fact that the variation of the system entropy must vanish over a cycle, one has

$$\langle \Sigma \rangle(\tau) = -\frac{\langle Q_H \rangle}{T_H} - \frac{\langle Q_C \rangle}{T_C} \geq 0, \quad (2.51)$$

where the last inequality holds because of the positivity of $\langle \Sigma \rangle$, and the equality only holds in the case of quasi-static transformations. By using Eq. (2.48) one gets

$$\langle \Sigma \rangle = -\frac{\langle Q_H \rangle}{T_H} - \left(\frac{-\langle Q_H \rangle - \langle W \rangle}{T_C} \right) = \frac{\langle W \rangle}{T_C} + \frac{\langle Q_H \rangle}{T_C} \left(1 - \frac{T_C}{T_H} \right), \quad (2.52)$$

i.e. an equation expressing the average entropy production by means of the average work and absorbed heat. If one defines the efficiency of the engine

$$\eta = -\frac{\langle W \rangle}{\langle Q_H \rangle}, \quad (2.53)$$

where the minus sign is considered in order to have a positive efficiency when work is produced, it is easy to show that

$$\eta \leq \eta_C = 1 - \frac{T_C}{T_H}, \quad (2.54)$$

where η_C is the Carnot efficiency; this inequality is a direct consequence of the positivity of the entropy production, the equality holding only in the case of a vanishing entropy.

Because of Eq. (2.54), unfortunately, the maximum possible efficiency is only reached for quasi-static transformations, i.e. for $\tau \rightarrow \infty$ which corresponds to a vanishing power $\langle W \rangle/\tau$ and hence to a useless heat engine: for this reason, in actual engines, like the ones operating in cars or power plants, in general, an efficiency lower than the maximum is accepted, if that is compensated by the maximum possible output power. Therefore an interesting figure of merit, i.e. a quantity describing the “goodness” of the heat engine, is the *efficiency at maximum power* $\tilde{\eta}$, i.e. the efficiency reached by the system in the configuration of the external parameters that maximizes the power $\langle W \rangle/\tau$. In many interesting situations, as we will show in the next paragraph, such a quantity has an upper bound, i.e. the *Curzon-Ahlborn efficiency*,

$$\tilde{\eta} \leq \eta_{CA} = 1 - \sqrt{\frac{T_C}{T_H}}, \quad (2.55)$$

which, for small value of $|T_H - T_C|$ is close to the half of the Carnot efficiency. Of course, it is not possible to obtain results regarding the efficiency at maximum power using the standard tools of thermodynamics: indeed, in order to maximize the power, it is necessary to turn our attention far from the quasi-static regime, and therefore to use a non-equilibrium formalism, which will be developed in the next paragraph.

2.3.1 Linear regime and Onsager reciprocity relations

Every irreversible process is characterized by a non-negative average entropy production (all the considered quantities are integrated in time on a cycle period τ) Σ which, in turn, can be expressed as the sum of n products of some thermodynamic (time-integrated) fluxes J_i with the associated thermodynamic forces f_i [29]:

$$\Sigma = \sum_{i=1}^n J_i f_i, \quad (2.56)$$

where, here and for the rest of this paragraph, we drop the average symbols on Σ , in order to momentarily shorten and simplify the notation. The entropy production and all the fluxes are expected to vanish at equilibrium (i.e. when there are no thermodynamic forces). Consequently, when the forces are small, the fluxes can be expressed as linear combinations of the forces,

$$J_i = \sum_{k=1}^n L_{ik} f_k, \quad (2.57)$$

where L_{ik} are the so-called *Onsager coefficients*. Thus, in the linear regime, the entropy production rate is a quadratic form of the thermodynamic forces,

$$\Sigma = \sum_{i,k=1}^n L_{ik} f_i f_k. \quad (2.58)$$

The matrix of Onsager coefficients (*Onsager matrix*), given the positivity of Σ , must be positive-semidefinite.

In the context of heat engines, there are two “forces” that drive the system far from equilibrium, namely the difference between the temperatures of the thermostats and the variation over time of the external parameter: in fact, whenever the temperature and the parameters are constant in time the system is at equilibrium. To set the notation, we denote with ϵ the relative excursion of the external parameter λ , $\epsilon = (\lambda_H - \lambda_L)/\lambda_L$, where λ_H and λ_L are, respectively the maximum and the minimum value assumed by the parameter. With the same notation we will be addressing the temperature variation: $\delta = (T_H - T_C)/T_H$.

In many physical problems the identification of fluxes and forces is unambiguous: that is not the case in the problem we are considering [81, 80]. Nevertheless it is quite natural, by looking at Eq. (2.52), to do the following identification:

$$J_1 = \frac{\langle W \rangle}{\epsilon T_C}(\tau), \quad (2.59)$$

$$J_2 = \frac{\langle Q_H \rangle}{T_C}(\tau). \quad (2.60)$$

Indeed, it is quite easy to see that the work $\langle W \rangle$ must be proportional to ϵ , since the work is always zero when the external parameter is not varied. Furthermore, it is reasonable to assume that both the fluxes are linear functions of the thermodynamic forces:

$$J_1 = L_{11}(\tau)\epsilon + L_{12}(\tau)\delta, \quad (2.61)$$

$$J_2 = L_{21}(\tau)\epsilon + L_{22}(\tau)\delta. \quad (2.62)$$

This decomposition is particularly suited for discussing the efficiency at maximum power, as we will see in the following section. It is very important to stress that the Onsager coefficients depend upon the total time of the cycle: this allows to study the properties of the system also by varying the velocity of the cycle. In particular, when computing the Onsager matrix, one must recover the fact that in the adiabatic limit ($\tau \rightarrow \infty$), the entropy production must vanish, i.e. the Onsager matrix must be degenerate. The other case in which the entropy is zero corresponds to the equilibrium situation, i.e. when ϵ and δ are both equal to zero (independently of the form of the matrix).

This considerations can be generalized to a system \mathbf{x} with hamiltonian $\mathcal{H}(\mathbf{x}(t), \lambda(t))$ and coupled to a thermostat at the time-dependent temperature $T(t)$. Both $\lambda(t)$ and $T(t)$ are periodic functions of time, with period τ . Since the thermostating dynamics satisfies, at every time t , the detailed balance condition with the equilibrium (Gibbs) distribution at temperature $T(t)$, the total average entropy production (see Appendix 2.A.3) of the system reads

$$\Sigma(\tau) = - \int_0^\tau \frac{\langle \dot{Q}(t) \rangle}{T(t)} dt \geq 0, \quad (2.63)$$

where \dot{Q} is the rate of heat absorption from the thermostat. To obtain a decomposition analogous to Eq. (2.56), it is useful to express the temperature, following [8], as

$$T(t) = \frac{T_0(1 - \delta^2)}{(1 + \delta) - 2\delta\gamma(t)}, \quad (2.64)$$

where $T_0 = (T_H + T_C)/2$ and

$$\delta = \frac{T_H - T_C}{T_H + T_C}. \quad (2.65)$$

The time-dependence of $T(t)$ is expressed through the function $0 \leq \gamma(t) \leq 1$, so that $\gamma(t) = 1 \iff T(t) = T_H = T_0(1 + \delta)$ and $\gamma(t) = 0 \iff T(t) = T_C = T_0(1 - \delta)$. We will also use the notation $\lambda_0 = (\lambda_H + \lambda_L)/2$ and

$$\epsilon = \frac{\lambda_H - \lambda_L}{\lambda_H + \lambda_L}, \quad (2.66)$$

to indicate, respectively, the intermediate force and the relative excursion. By plugging Eq. (2.64) into Eq. (2.63) and using $\langle W \rangle + \langle Q \rangle = 0$ one gets

$$\Sigma(\tau) = \frac{\langle W \rangle}{T_0(1 - \delta)} + \delta \left(\frac{2}{T_0(1 - \delta^2)} \int_0^\tau \langle \dot{Q} \rangle \gamma(t) dt \right). \quad (2.67)$$

By interpreting ϵ and δ as the two (adimensional) independent thermodynamic forces, or affinities, acting on the system, and recalling that $\langle W \rangle \propto \epsilon$ for small ϵ , one gets an expression analog to Eq. (2.56):

$$\Sigma(\tau) = \epsilon J_1(\tau) + \delta J_2(\tau), \quad (2.68)$$

where

$$J_1(\tau) = \frac{\langle W \rangle}{\epsilon T_0(1 - \delta)}, \quad (2.69)$$

$$J_2(\tau) = \frac{2}{T_0(1 - \delta^2)} \int_0^\tau \langle \dot{Q} \rangle \gamma(t). \quad (2.70)$$

It is easy to see that J_2 , in the limiting case where $T(t)$ is a square wave function between T_C and T_H (i.e. when $\gamma(t)$ assumes only the values $\gamma = 1$ or $\gamma = 0$) is proportional to the heat Q_H exchanged with the hot thermostat, i.e. we recover the usual known results. It is worth noticing that expressions different from Eq. (2.64) can lead to different (legit) definition of fluxes: nonetheless, this decomposition is particularly suited for an analysis of the efficiency at maximum power. In fact, with this last observation, we can simply define a generalized efficiency:

$$\eta = -\frac{\langle W \rangle}{\langle Q_{in} \rangle} \quad (2.71)$$

where the denominator is a quantity proportional to J_2

$$Q_{in} = \int_0^\tau dt \dot{Q} \gamma(t) = 2T_0(1 - \delta^2)J_2 \quad (2.72)$$

. The definition of the efficiency in terms of the thermodynamic fluxes reads:

$$\eta = -\frac{2\epsilon J_1}{(1 + \delta)J_2}. \quad (2.73)$$

We now discuss a generalization of the Onsager reciprocity relations for systems undergoing cyclical transformations, proposed in in Ref. [8]. For each protocol determined by $T(t)$ and $\lambda(t)$, it is possible to construct its “time-reversed” counterpart $\tilde{T}(t) = T(\tau - t)$ and $\tilde{\lambda}(t) = \lambda(\tau - t)$: if we indicate with $\tilde{\cdot}$ quantities measured in the time-reversed cycle, the following relation is a direct consequence of the reversibility of the microscopic dynamics:

$$L_{12}(\tau) = \tilde{L}_{21}(\tau). \quad (2.74)$$

A common situation is when the protocol are even or odd functions under time-reversal transformations. In this case one has

$$L_{12}(\tau) = \pm L_{21}(\tau), \quad (2.75)$$

where the plus sign (symmetric matrix) applies when the protocol is invariant under time reversal and the minus sign (antisymmetric matrix) when it changes sign under time-reversal.

2.3.2 Efficiency at maximum power

The efficiency at maximum power (EMP) is the efficiency of an engine operating at a certain value of its external parameters corresponding to the maximum possible output power. We now briefly show, following Ref. [99], that in the most simple maximization process, the EMP has an upper bound, the Curzon-Ahlborn efficiency [28]:

$$\eta_{CA} = 1 - \sqrt{\frac{T_H}{T_C}} \approx \frac{1}{2}\eta_C. \quad (2.76)$$

Let us consider the time-rescaled version of Eq. (2.56) and to use the same symbols, with an abuse of notation, we denote the quantities defined above and divided by τ : for this reason if Σ represents the entropy production rate, $\langle W \rangle$ the power etc. . . the form of the Onsager matrix does not change. The power, expressed in terms of the thermodynamic fluxes, is

$$\langle W \rangle = \epsilon T_0(1 - \delta)J_1 = T_0(1 - \delta)(L_{11}\epsilon^2 + L_{12}\epsilon\delta). \quad (2.77)$$

We now want to maximize this expression with respect of one of the thermodynamic forces: since, in general, it is customary to keep the temperature difference fixed, and since the Carnot efficiency only depends upon the temperature difference, we do the maximization procedure with respect to ϵ , obtaining

$$\epsilon^* = -\frac{L_{12}(\tau)\delta}{2L_{11}(\tau)}. \quad (2.78)$$

By plugging this last expression into the definition of efficiency Eq. (2.73) one gets

$$\tilde{\eta} = \frac{q(\tau)^2}{2 - q(\tau)^2} \frac{\delta}{2}, \quad (2.79)$$

where, depending on whether the matrix is symmetrical or antisymmetrical, the parameter q is

$$q(\tau) = \pm \frac{L_{12}(\tau)}{\sqrt{L_{11}(\tau)L_{22}(\tau)}}. \quad (2.80)$$

In the limit of “strong coupling”, i.e. when $|q| \rightarrow 1$, this result is exactly one half of the Carnot efficiency ($\delta/2$), i.e. the power expansion of the Curzon-Ahlborn efficiency at first order in δ .

This result, which appears very general, has been strongly debated: in particular in Ref. [90] a single-particle Brownian motion was exhibited in which the efficiency can overcome the Curzon-Ahlborn limit. Moreover the maximization process used in the above derivation is limited, above all, by the fact that the τ -dependence of the Onsager matrix is neglected: in the second part of the Thesis, by resorting to actual models, we will show that a maximization with respect both to the external parameter and to τ will lead to different results.

2.4 Conclusions

In this Chapter we revisited the main results obtained in the context of stochastic thermodynamics and finite-time thermodynamics in the last decades. These results

give a new deep insight into the thermodynamics of small system, in which the two main assumptions of thermodynamics, i.e. large systems and slow transformation, are violated.

A thermodynamic formalism can be properly defined also for system in the mesoscopic realm, stochastic single-particle systems described by a Langevin equation being the most relevant examples. Because of the relevance and the simplicity of these models, most of the literature on this topic focused on brownian motor moving in time-varying external potentials.

One of the aim of the Thesis is trying to extend and validate these results to much more complex systems, in particular thermostatted systems composed by few hundreds of particles in which the cyclic protocol is performed over a finite interval. By having a tighter control over the microscopic model, i.e. not resorting on coarse-grained model but, instead, on Hamiltonian models attached to some thermostats, one can understand in depth the mechanisms underlying the physics of these systems. Moreover, by using more realistic models which depend upon a large number of control variables. one can also explore the limit of the general theory and overcome some bounding values prescribed by the theory.

Since developing a general theory for the class of systems we are interested in is not possible, due to the huge variety of different models and to the difficulty of performing analytical calculations on systems composed of a large (but not infinite) number of degrees of freedom, we will resort to specific models. In the second part of the Thesis we will study and analyze a model of a gas enclosed in a container with a moving wall that is the microscopic equivalent of the usual engines encountered in standard thermodynamics. We will apply most of the considerations exposed in this section to the model and test the possibility of designing small engine.

2.A Fluctuation theorems and Jarzynski Equality

Consider a system evolving in time: every possible trajectory $\Omega = \{\mathbf{x}(s)\}_{s=t_0}^{\tau}$ occurs with a probability $\mathcal{P}_{\lambda}[\Omega]$. The evolution of the system may be either stochastic or deterministic (in the latter case the functional \mathcal{P} will trivially coincide with the probability distribution function of the unique initial condition generating a trajectory). The dynamics will in general depend upon an external time-varying protocol $\lambda(t)$: different protocols generate different dynamics. We define the *action functional*:

$$\mathcal{W}[\Omega] = \ln \frac{\mathcal{P}_{\lambda}[\Omega]}{\mathcal{P}_{\lambda'}[\mathcal{R}\Omega]} \quad (2.81)$$

where \mathcal{R} can be any involutive transformation, such that $\mathcal{R}(\mathcal{R}[\Omega]) = \Omega$ and λ and λ' are two different protocols. Note that, the action functional measured in a system ruled by a dynamics in which the protocol is λ' simply reads

$$\mathcal{W}'[\Omega] = \ln \frac{\mathcal{P}_{\lambda'}[\Omega]}{\mathcal{P}_{\lambda}[\mathcal{R}\Omega]} = -\mathcal{W}[\mathcal{R}\Omega]. \quad (2.82)$$

Though not necessary, it is quite natural, also in view of the context in which such quantities are usually measured, to take \mathcal{R} as the time reversal transformation $\mathcal{R}\{\mathbf{q}(s), \mathbf{p}(s)\}_{s=t_0}^{\tau} = \{\mathbf{q}(\tau - s), -\mathbf{p}(\tau - s)\}_{s=t_0}^{\tau}$ and $\lambda'(t) = \lambda(\tau - t)$. The probability

of measuring a given value of the action (the summation symbol is purely symbolic, and should be replaced with an integration when the space of trajectory is continuous) is

$$\begin{aligned}
P(\mathcal{W} = X) &= \\
&= \sum_{\Omega: \mathcal{W}[\Omega]=X} \mathcal{P}_\lambda(\Omega) = \sum_{\Omega: \mathcal{W}[\Omega]=X} \frac{\mathcal{P}_\lambda[\Omega]}{\mathcal{P}_{\lambda'}[\mathcal{R}\Omega]} \mathcal{P}_{\lambda'}[\mathcal{R}\Omega] = \\
&= e^X \sum_{\Omega: \mathcal{W}[\Omega]=X} \mathcal{P}_{\lambda'}[\mathcal{R}\Omega] = e^X \sum_{\Omega: \mathcal{W}'[\Omega]=-X} \mathcal{P}_{\lambda'}[\Omega] = e^X P'[\mathcal{W}' = -X]. \quad (2.83)
\end{aligned}$$

This last equality can be recasted into the form

$$\frac{P(\mathcal{W} = X)}{P'(\mathcal{W}' = -X)} = e^X, \quad (2.84)$$

which constitutes the so-called *fluctuation relation* (FR) or fluctuation theorem (FT). Note that this equation is very general but not very meaningful: for any λ and λ' , i.e. when one compares two completely unrelated processes, the FT establishes a relation between the histogram of \mathcal{W} (measured in the system of interest) with the one of \mathcal{W}' (measured in a different system). Moreover the physical meaning of \mathcal{W} is not clear. Note that, also in great generality, the fluctuation theorem can be expressed in the following form

$$\langle e^{-\mathcal{W}} \rangle = 1, \quad (2.85)$$

which is called Integral Fluctuation Theorem (IFT). We will see later that the Jarzynski Equality is a special case of IFT (in the literature it is customary to call Jarzynski equality every equation analogous to Eq. (2.85)). From Jensen inequality ($\langle e^X \rangle \geq e^{\langle X \rangle}$), we get

$$\langle \mathcal{W} \rangle \geq 0. \quad (2.86)$$

By identifying \mathcal{W} in Eq. (2.86) with entropy, we get a microscopic, probabilistic interpretation of the second principle of thermodynamics.

In [66] it was proved that for a generic Markov process a fluctuation theorem holds. In particular the following quantity

$$\mathcal{W}_{LS}[\Omega = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_t\} | \mathbf{x}_0] = \ln \frac{p(\mathbf{x}_0 \rightarrow \mathbf{x}_1) \dots p(\mathbf{x}_{t-1} \rightarrow \mathbf{x}_t)}{p(\mathbf{x}_1 \rightarrow \mathbf{x}_0) \dots p(\mathbf{x}_t \rightarrow \mathbf{x}_{t-1})}, \quad (2.87)$$

which we will denote by *Lebowitz-Spohn Action Functional* (LSAF), satisfies a large deviation principle

$$\lim_{t \rightarrow \infty} -\frac{1}{t} \ln P \left(\frac{\mathcal{W}_{LS}[t]}{t} = x \right) = C(x), \quad (2.88)$$

and the Cramér function $C(x)$ has the following symmetry

$$C(x) = C(-x) - x. \quad (2.89)$$

It is easy to show that Eq. (2.89) is indeed a fluctuation theorem, in the sense of Eq. (2.81):

$$\frac{P(\mathcal{W}_{LS} = X)}{P(\mathcal{W}_{LS} = -X)} \underset{t \gg 1}{\sim} \exp \left\{ -tC \left(\frac{X}{t} \right) + tC \left(-\frac{X}{t} \right) \right\} = e^X. \quad (2.90)$$

This is an asymptotic result that only holds for large values of t : therefore is less general than the one explained before. Nonetheless it is worth noticing that the two functionals \mathcal{W}_{LS} and \mathcal{W} differ of a boundary term

$$\mathcal{W} - \mathcal{W}_{LS} = \ln(\rho(\mathbf{x}_0) - \ln \rho'(\mathbf{x}_t)). \quad (2.91)$$

Since in the vast majority of problems involving Non-Equilibrium Steady States (NESS) the initial (stationary) distribution function is not known, the \mathcal{W}_{LS} is often used. The literature discussing the relevance of the boundary term is huge (see, i.e., [83]).

2.A.1 Isolated Hamiltonian systems

Let us now consider a system described by a time dependent hamiltonian $H(\mathbf{x}, \lambda(t))$, where λ goes from $\lambda(t=0) = \lambda_0$ to $\lambda(t=\tau) = \lambda_1$. Let us also assume that, at time $t=0$, the initial condition of the system is distributed according to the equilibrium (canonical) distribution function at $\lambda = \lambda_0$,

$$\rho_0(\mathbf{x}) = \frac{1}{Z_0} \exp\{-\beta H(\mathbf{x}, \lambda_0)\}, \quad (2.92)$$

where β is the inverse temperature of the system and Z_0 its partition function. In order to build the action functional we compare this system with an analogous one obtained by reversing the protocol: $\lambda'(t) = \lambda(1-t)$ (i.e. $\lambda'(t=0) = \lambda_1$ and $\lambda'(t=1) = \lambda_0$). Moreover we assume that such a “reversed system”, at $t=0$ is in equilibrium at the inverse temperature β with $\lambda = \lambda_1$, i.e.

$$\rho'_0(\mathbf{x}) = \frac{1}{Z_1} \exp\{-\beta H(\mathbf{x}, \lambda_1)\}. \quad (2.93)$$

Let us also assume that the dynamics of the two systems is deterministic and ruled by the hamiltonian H (i.e. that the systems are decoupled from the thermostats for the whole duration of the process). The action functional (Eq. (2.81)) reads

$$\begin{aligned} \mathcal{W}[\Omega = \{\mathbf{x}_0, \dots, \mathbf{x}_f\}] &= \ln \left(\frac{\rho(\mathbf{x}_0)}{\rho'(\mathbf{x}_f)} \right) = \\ &= \ln \left(\frac{Z_1}{Z_0} \right) - \beta(H(\mathbf{x}_0, \lambda_0) - H(\mathbf{x}_f, \lambda_1)) = \beta(\Delta H - \Delta F). \end{aligned} \quad (2.94)$$

where $\Delta F = F_1 - F_0 = \ln(Z_1/Z_0)/\beta$ is the free energy difference. Since the system is not attached to any thermostat, the hamiltonian (energy) variation is simply the work performed on the system: $W = \Delta H$.

By using Eq. (2.84) and the fact that, for analogous reasons, $\mathcal{W}' = \beta(W' + \Delta F)$, where W' is the work measured in the reversed process and ΔF the same quantity as before, one gets

$$\frac{P(W = X)}{P(W' = -X)} = \frac{P(\beta(W - \Delta F) = \beta(X - \Delta F))}{P(\beta(W' + \Delta F) = -\beta(X - \Delta F))} = e^{\beta(X - \Delta F)}, \quad (2.95)$$

that is the so-called *Crooks relation*:

$$\frac{P(W)}{P(-W')} = e^{\beta(W - \Delta F)}. \quad (2.96)$$

The corresponding integral relation is the well-known Jarzynski equality:

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}. \quad (2.97)$$

2.A.2 Thermostatted Hamiltonian systems

For thermostatted systems the derivation is more complex, for this reason we will prove it in the simple case of systems with discrete time and states: apart from avoiding some technical difficulties, this proof contains the essence of the one in the continuous case. At every time $t = 0, 1, \dots, \tau$, the state of a system is denoted by i_t , its energy represented by the hamiltonian $H(i, \lambda_t)$, and λ_t is a (discrete) time-dependent protocol. Since the stochastic dynamics is intended to reproduce the equilibrium for fixed λ , it must satisfy the detailed balance relation:

$$p_\lambda(i \rightarrow j) \exp\{-\beta H(i, \lambda)\} = p_\lambda(\mathcal{R}j \rightarrow \mathcal{R}i) \exp\{-\beta H(\mathcal{R}j, \lambda)\} \quad (2.98)$$

Given a trajectory $\Omega = (i_0, \dots, i_\tau)$ and assuming that the initial distributions for the forward and reversed processes are the equilibrium ones, the action functional can be easily computed (we assume $H(i) = H(\mathcal{R}i)$):

$$\begin{aligned} \mathcal{W}[\Omega] &= \ln \frac{p_{\lambda_1}(i_0) p_{\lambda_1}(i_0 \rightarrow i_1) \dots p_{\lambda_\tau}(i_{\tau-1} \rightarrow i_\tau)}{p_{\lambda_\tau}(\mathcal{R}i_\tau) p_{\lambda_\tau}(\mathcal{R}i_\tau \rightarrow \mathcal{R}i_{\tau-1}) \dots p_{\lambda_1}(\mathcal{R}i_1 \rightarrow \mathcal{R}i_0)} \\ &= \ln \left(\frac{Z_1}{Z_0} \right) - \beta(H(i_0, \lambda_1) - H(i_\tau, \lambda_\tau)) - \beta \sum_{t=1}^{\tau} [H(i_t, \lambda_t) - H(i_{t-1}, \lambda_t)] \end{aligned} \quad (2.99)$$

At every time t the total energy difference can be decomposed into two contributions, heat and work, due, respectively, to the change of the state and change of the external parameter:

$$H(i_t, \lambda_t) - H(i_{t-1}, \lambda_{t-1}) = \underbrace{H(i_t, \lambda_t) - H(i_{t-1}, \lambda_t)}_{\text{Heat } Q} + \underbrace{H(i_{t-1}, \lambda_t) - H(i_{t-1}, \lambda_{t-1})}_{\text{Work } W}. \quad (2.100)$$

Going back to Eq. (2.99)

$$\mathcal{W}[\Omega] = \beta \left(-\Delta F + \Delta H - \sum_{t=1}^{\tau} Q_t \right) = \beta(W - \Delta F), \quad (2.101)$$

which, again, gives the Crooks and Jarzynski equations.

2.A.3 Thermostatted Hamiltonian systems with a time-dependent temperature

For the sake of consistency, in this Appendix we prove Eq. (2.63) for the simple case of discrete phase space and time. This formula, which holds also if time and space are both continuous, has appeared many times in the literature (see e.g. Ref. [48] for a nice pedagogical derivation).

Let us consider a discrete Markov process with time-dependent transition probabilities $p(i \rightarrow j, t)$ satisfying the detailed balance (DB) condition

$$e^{-\frac{\mathcal{H}(i, \lambda_t)}{T_t}} p(i \rightarrow j, t) = e^{-\frac{\mathcal{H}(j, \lambda_t)}{T_t}} p(j \rightarrow i, t), \quad (2.102)$$

for every value of i, j and t , where λ_t is an external time-dependent protocol and T_t the time-dependent temperature. The entropy production of the medium for a given trajectory $\{i_t\}_{t=0}^\tau$ reads [66]

$$\Sigma_m(\tau) = \log \frac{P[\{i_s\}_{s=0}^\tau | i_0]}{\tilde{P}[\{i_{\tau-s}\}_{s=0}^\tau | i_\tau]} \quad (2.103)$$

where \tilde{P} denotes the probability of the trajectory in a process with the time reversed protocol $\tilde{\lambda}_t = \lambda_{\tau-t-1}$. By expliciting Eq. (2.103) and using DB one gets

$$\begin{aligned} \Sigma_m(\tau) &= \log \frac{p(i_0 \rightarrow i_1, \lambda_0) \dots p(i_{\tau-1} \rightarrow i_\tau, \lambda_{\tau-1})}{p(i_\tau \rightarrow i_{\tau-1}, \lambda_{\tau-1}) \dots p(i_1 \rightarrow i_0, \lambda_0)} = \\ &= - \sum_{t=1}^{\tau} \frac{\mathcal{H}(i_t, \lambda_{t-1}) - \mathcal{H}(i_{t-1}, \lambda_{t-1})}{T_{t-1}}. \end{aligned} \quad (2.104)$$

At every step the energy difference of the system can be decomposed according to

$$\Delta E = \mathcal{H}(i_t, \lambda_t) - \mathcal{H}(i_{t-1}, \lambda_{t-1}) = W_t + Q_t, \quad (2.105)$$

where the work W_t is the contribution due only to the change of λ , $W_t = \mathcal{H}(i_t, \lambda_t) - \mathcal{H}(i_t, \lambda_{t-1})$, and the heat Q_t is due to the change of state at fixed λ , $Q_t = \mathcal{H}(i_t, \lambda_{t-1}) - \mathcal{H}(i_{t-1}, \lambda_{t-1})$. Consequently

$$\Sigma_m(\tau) = - \sum_{t=0}^{\tau-1} \frac{Q_t}{T_t}. \quad (2.106)$$

The non-negative total entropy production $\Sigma_{tot}(\tau)$ [92] is the sum of the medium entropy production Σ_m and the system entropy production

$$\Sigma_s(\tau) = - \log \frac{\rho(i_0, t=0)}{\rho(i_\tau, t=\tau)}, \quad (2.107)$$

where $\rho(i_t, t)$ is the probability distribution function (pdf) of the system at time t . If λ_t and T_t are periodic function of time with period τ , there exists a periodic “stationary” pdf ρ_∞ such that $\rho_\infty(i, t) = \rho_\infty(i, t + \tau)$, for every i and t . This also means that if the initial pdf $\rho(i, t=0) = \rho_\infty(i, t=0)$, on average, the system entropy production vanishes, $\langle \Sigma_s \rangle = 0$ and finally

$$\langle \Sigma(\tau) \rangle = \langle \Sigma_m \rangle = - \sum_{t=0}^{\tau} \frac{\langle Q \rangle}{T_t}, \quad (2.108)$$

which is the discrete time equivalent of Eq. (2.63).

Part II

A case of study: the gas-piston engine prototype

Chapter 3

Equilibrium description and coarse-grained equations of motion

3.1 Introduction

In this second part of the Thesis we will analyze a specific example in which most of the feature of statistical mechanics and thermodynamics of small systems can be observed, studied and analyzed. In this Chapter we will introduce the system of interest, and analyze its statistical equilibrium characteristics, trying to derive a coarse grained equation of motion describing the evolution of a set of relevant macroscopic observables of the system.

A class of systems that is important for nanosciences is that of partitioning objects containing an extra degree of freedom (a wall) which separates the system into subsystems. A paradigmatic example is given by the adiabatic piston [27, 49, 22]: a system of N particles of mass m (e.g. an ideal gas) in a container of length L and cross-section A , separated in two regions by a movable wall (the piston) of mass M . The walls of the container are supposed to be perfect insulators preventing any mass or heat exchanges with the exterior. Gas particles undergo purely elastic collisions with the piston and the walls, and the piston is constrained to move along one axis. If at initial time the temperatures T_L , T_R and pressures P_L , P_R in the left and right parts do not coincide, the system shows a rather rich phenomenology (depending on M/m , N/L etc) in the approach to the mechanical and thermodynamic equilibrium.

A physical version of the adiabatic piston is a big Brownian particle sliding along a microtubule filled with particles [30]. The authors of Ref. [30] showed how the presence of the wall is able to induce, even in the equilibrium state, rather non trivial (and slow) dynamical behaviors.

Our work is devoted to the statistical mechanics of a system where particles are confined in a tube with a non fixed wall, on which an external force acts, see Fig. 3.1. The pressure on the piston due to the interaction with the gas particles on one side is balanced by the external force, so that the piston reaches a stationary state. This system is a clear generalization of the classical gas-piston system, i.e. a schematization of a mechanical engine in which the movement of the piston, due to

the expansion-compression cycle of the gas, is converted into the rotational movement of a wheel, e.g. in cars, steam trains, power plants etc. . . .

Our interest in small systems led to focus on the case where N is not too large, i.e. approximately $10^2 - 10^3$. This microscopic setup will help understanding the microscopic mechanics underlying the conversion of thermal energy into mechanical energy. In addition, as we will show in the following, we have a model in which the derivation of a coarse-grained stochastic differential equation (Langevin equation) can be carried on without many difficulties. This equation permits to obtain some analytic predictions: the comparison between the analytic results and computer simulation will show that some common assumptions and simplifications given for granted in the derivation of the coarse-grained equation are actually not satisfied in systems with a small number of degrees of freedom.

Nevertheless, despite being only an approximated description of the physics of the system, we will show that, from a qualitative point of view, the stochastic differential equation can be actually used in order to predict its behavior, especially in the non-equilibrium regime.

In this Chapter we will mainly focus on the equilibrium problem: i.e. we will determine the state of the system in the canonical and microcanonical ensemble. One of the most striking equilibrium features of this system is that the fluctuations of the position of the piston are different whether the system is isolated (microcanonical) or attached to a thermal reservoir (canonical). As already discussed in Chapter 1, this is not a violation of the equivalence of ensemble, since such a theorem only involves average values of macroscopic observables (i.e. it does not give information regarding their variances and standard deviations). This consideration, whose explanation is quite simple and immediate, leads to a very interesting result, i.e. the fact that the dynamics of the piston should be different in the canonical and microcanonical ensemble. This is quite surprising, since in the usual derivation of the coarse-grained equation of motion of the piston, no difference emerges when the computations are carried on in the canonical or microcanonical ensemble. An important consequence of such a difference, which holds both in the case of non-interacting and interacting particles, is that the correlation function (of the velocity) $C(t)$ must be different in the two ensembles.

Numerical simulations show a non trivial behavior of $C(t)$ with a negative minimum around a characteristic time $\tau(N)$ increasing linearly with N . A comparison between the numerical results and the Langevin equation shows how even for large N the presence of the wall has non trivial consequences which can have a role for an effective modeling of the system.

3.2 The model

We consider a two-dimensional system composed by a gas of N point-like particles with mass m , positions $\mathbf{x}_i = \{x_i, y_i\}$ and momentum \mathbf{p}_i , with $i = 1, \dots, N$, contained in a rectangular box with one moving adiabatic wall of length L (hereafter referred to as the “piston”). The position of the piston is denoted by Y and its momentum and mass are P and M , respectively (see Fig 3.1 for visual explanation). An external force $\mathbf{F} = -F \cdot \hat{y}$, directed along the horizontal axis \hat{y} , acts on the piston, which is also

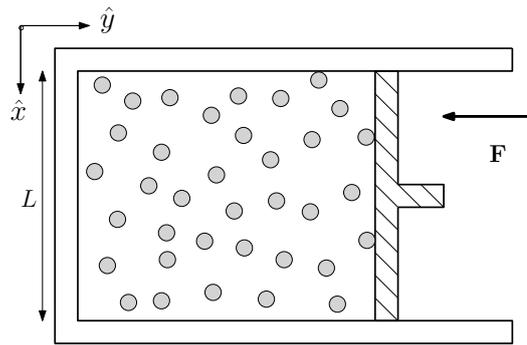


Figure 3.1. Sketch of the piston model: a gas of particles is confined by a moving wall which is subjected to a constant external force.

subject to the collisions with the particles. In the tubular geometry that we consider, in which the size of the system is increased anisotropically only along one direction when adding particles, the piston plays the role of a “partitioning” object with respect to the particle gas, namely its position determines the volume available for the gas. This system has been studied in [37] as an effective thermometer model. In the following the particle-particle and particle-piston interactions are described in an Hamiltonian (conservative) context and the piston can slide without dissipation along the y axis. The case of dissipative interactions, inducing nonequilibrium behaviors, of similar systems have been studied for instance in [11, 53, 25, 41, 89, 88]. The two-dimensionality of the system is necessary because, for the sake of completeness, we also want to study the case with an interaction among the particles: nevertheless, in the following Chapter, an idealized and simplified version of the system will be studied, in which the particles do not directly interact, i.e. they constitute a perfect gas. In this situation there is no difference, apart from inessential constants that depend upon the dimensionality of the system, between the multi-dimensional and the uni-dimensional case. For the sake of consistency, in this section we will only discuss the two-dimensional case, postponing the illustration of the characteristic of the uni-dimensional case in the following Chapter, see Fig. 4.1.

3.3 Statistical Equilibrium Description

We start by considering the case of a non-interacting gas, so that the hamiltonian of the system reads

$$\mathcal{H} = \sum_{i=1}^N \frac{|\mathbf{p}_i|^2}{2m} + \frac{P^2}{2M} + FY, \quad (3.1)$$

with geometrical constraints

$$\begin{cases} Y > 0; \\ 0 < x_i < L; \\ 0 < y_i < Y. \end{cases} \quad (3.2)$$

We are interested in the study of the fluctuations at varying the number of gas particles, and, in particular, in the comparison between the microcanonical and canonical ensembles.

3.3.1 Microcanonical

In the microcanonical ensemble at energy E , the invariant measure is non-zero only on the hypersurface of constant energy S_E . If \mathcal{M} is a subset of S_E and $d\sigma$ is the infinitesimal surface element

$$\mathbb{P}(x \in \mathcal{M} \subseteq S_E) = \int_{\mathcal{M}} \frac{d\sigma}{\omega(E)} \frac{1}{|\nabla \mathcal{H}|}, \quad (3.3)$$

where $\omega(E) = \partial \Sigma(E) / \partial E$. In order to derive the expression of the temperature of the system as a function of the energy, we must compute $\Sigma(E)$. This quantity is given by

$$\begin{aligned} \Sigma(E) &= \int_{\mathcal{H} < E} d^N x d^N y dY d^N \mathbf{p} dP \\ &= L^N \int_{\sum_i |\mathbf{p}_i|^2 / 2m + P^2 / 2M + FY < E} d^N y dY d^N \mathbf{p} dP. \end{aligned} \quad (3.4)$$

Recalling that the volume of a D -dimensional sphere of radius R is $V(R) = \int_{\sum_i x_i^2 < R^2} d^D x = \frac{\pi^{\frac{D}{2}}}{\Gamma(\frac{D}{2}+1)} R^D$, where $\Gamma(x)$ is the Euler Gamma, from Eq. (3.4) we obtain

$$\begin{aligned} \Sigma(E) &= (2m)^N \sqrt{2M} L^N \frac{\pi^{N+\frac{1}{2}}}{\Gamma(N+\frac{3}{2})} \\ &\times \int_0^Y d^N y \int_0^{E/F} dY (E - FY)^{N+\frac{1}{2}} \\ &= (2m)^N \frac{\sqrt{2M}}{F} \left(\frac{L}{F}\right)^N \frac{\pi^{N+\frac{1}{2}}}{\Gamma(N+\frac{3}{2})} E^{2N+\frac{3}{2}} \\ &\times \int_0^1 dx x^N (1-x)^{N+\frac{1}{2}} \end{aligned} \quad (3.5)$$

and, eventually,

$$\Sigma(E) = (2m)^N \frac{\sqrt{2M} L^N}{F^{N+1}} \pi^{N+\frac{1}{2}} \frac{\Gamma(N+1)}{\Gamma(2N+\frac{5}{2})} E^{2N+\frac{3}{2}}, \quad (3.6)$$

and

$$\omega(E) = (2m)^N \frac{\sqrt{2M} L^N}{F^{N+1}} \pi^{N+\frac{1}{2}} \frac{\Gamma(N+1)}{\Gamma(2N+\frac{3}{2})} E^{2N+\frac{1}{2}} \quad (3.7)$$

Now we can compute the temperature of the system using the relation $S = k_B \ln \Sigma(E)$, namely

$$k_B T = k_B \left(\frac{\partial S}{\partial E} \right)^{-1} = \frac{\Sigma(E)}{\omega(E)} = \frac{E}{2N + \frac{3}{2}}. \quad (3.8)$$

Using alternative definitions of S , e.g. $S = k_B \ln \omega(E)$ or $S = k_B \ln \Gamma_{\Delta E}(E)$, where $\Gamma_{\Delta E}(E) = \Sigma(E + \Delta E) - \Sigma(E) \simeq \omega(E) \Delta E$ where ΔE is the tolerance on E , for $N \gg 1$ one has negligible differences [51].

We are interested in the probability density function of the position of the piston Y . Observing that for a generic phase space function $A(\mathbf{X})$ in the microcanonical ensemble one has [59]:

$$\rho_A(a) = \frac{1}{\omega(E)} \frac{\partial}{\partial E} \mathcal{I}(E, a), \quad (3.9)$$

where

$$\mathcal{I}(E, a) = \int_{\mathcal{H} < E} \delta(A(\mathbf{x}) - a) d\mathbf{x}, \quad (3.10)$$

putting $A(\mathbf{X}) = Y$ one readily obtains

$$\begin{aligned} I(E, Y = \tilde{Y}) &= \int_{H < E} dY d^N x d^N y d^N \mathbf{p}_i dp \delta(Y - \tilde{Y}) \\ &= (2m)^N \sqrt{2ML}^N \frac{\pi^{N+\frac{1}{2}}}{\Gamma(N + \frac{3}{2})} \tilde{Y}^N (E - F\tilde{Y})^{N+\frac{1}{2}}, \end{aligned} \quad (3.11)$$

for $0 < \tilde{Y} < E/F$; therefore

$$\begin{aligned} \rho_E(Y) &= \frac{1}{\omega(E)} \frac{\partial I}{\partial E} = \frac{\Gamma(2N + \frac{3}{2})}{\Gamma(N + \frac{1}{2})\Gamma(N + 1)} \\ &\times \frac{F}{E} \left(\frac{FY}{E}\right)^N \left(1 - \frac{FY}{E}\right)^{N-\frac{1}{2}}. \end{aligned} \quad (3.12)$$

From the above result, we obtain

$$\langle Y \rangle = \frac{(N + 1)k_B T}{F} \quad (3.13)$$

and

$$\sigma_Y^2 = \frac{(N + \frac{1}{2})(N + 1)}{2N + \frac{5}{2}} \left(\frac{k_B T}{F}\right)^2, \quad (3.14)$$

where, in the two last equations, we used Eq. (3.8) to express $\langle Y \rangle$ and σ_Y^2 as functions of T instead of E .

3.3.2 Canonical

In the canonical ensemble at constant temperature T with $\beta = 1/(k_B T)$, the partition function of the system is given by

$$\begin{aligned} Z &= \int d^N x d^N y dY d^N \mathbf{p} dP e^{-\beta \mathcal{H}} = \\ &= \left(\frac{2}{\pi}\right)^{N+\frac{1}{2}} N! m^N \sqrt{M} \beta^{-(2N+\frac{3}{2})} F^{-(N+1)}. \end{aligned} \quad (3.15)$$

We can easily compute the mean energy of the system

$$E = \langle \mathcal{H} \rangle = -\frac{\partial \ln Z}{\partial \beta} = \left(2N + \frac{3}{2}\right) k_B T. \quad (3.16)$$

Now we want to find the probability distribution function of the position of the piston Y : starting from

$$\rho_\beta(Y, \{y_i\}) = \frac{e^{-\beta F Y}}{\int dY d^N y e^{-\beta F Y}} \prod_i \Theta(Y - y_i), \quad (3.17)$$

and integrating over all the y_i , one obtains

$$\rho_\beta(Y) = \frac{Y^N e^{-\beta F Y}}{\int dY Y^N e^{-\beta F Y}}. \quad (3.18)$$

The mean value of this distribution is

$$\langle Y \rangle = \frac{k_B T (N + 1)}{F} \quad (3.19)$$

whereas its variance is

$$\sigma_Y^2 = \frac{(N + 1)(k_B T)^2}{F^2}. \quad (3.20)$$

3.3.3 Comparison between microcanonical and canonical ensemble

In order to compare the results for static quantities $\langle Y \rangle$ and σ_Y^2 in the two ensembles, for each temperature T in the canonical ensemble we consider the corresponding energy in the microcanonical, such that $T = E/(2Nk_B)$, in the limit of large number of particles $N \gg 1$. While the average position is always the same, from Eqs. (3.20) and (3.14) one observes that fluctuations differ by a factor 1/2, also in the large N limit. In the Table 3.1 we summarize these findings.

The equivalence of ensembles in the thermodynamic limit is expected only for average values, and not for fluctuations [68, 77]. Indeed, the observed “discrepancy” is explained by noting that the variance in the canonical ensemble can be expressed as the sum of two contributions, namely a term which corresponds to the variance of the piston in the microcanonical ensemble at fixed energy plus a term corresponding to energy fluctuations at fixed temperature:

$$\sigma_Y^2(T) = \alpha \sigma_E^2(T) + \sigma_Y^2(E) \Big|_{E=\langle E \rangle_\beta}, \quad (3.21)$$

where $\alpha = 1/(4F^2) + O(1/N)$ and $\sigma_E^2 = \langle \mathcal{H}^2 \rangle - E^2$. Therefore, for $N \gg 1$, since $\sigma_E^2(T) \approx 2N(k_B T)^2$, one has $\sigma_Y^2(T) = 2\sigma_Y^2(E) \Big|_{E=2Nk_B T}$.

Let us open a short digression on terminology. With the term “canonical ensemble” we mean the system with hamiltonian in Eq. (3.1) and Eq. (3.2) (in the following we will include also the interactions among the particles) interacting with a thermal bath at temperature T . Noting that the pressure is nothing but F/L , one can then say that we are dealing with an ensemble at fixed temperature and fixed pressure for the system without the terms FX and $P^2/(2M)$ in the hamiltonian [77]. In a similar way our microcanonical ensemble correspond to an ensemble with fixed enthalpy for the system without the terms FX and $P^2/(2M)$ in the hamiltonian. We prefer the terms canonical and microcanonical because they put the dynamical variables describing the wall on the same level of those for the particles. Let us note that the mass of the piston is important for the dynamical properties.

	Canonical	Microcanonical
Temperature:	T	$\frac{E}{2Nk_B}$
$\langle Y \rangle$	$\frac{Nk_B T}{F}$	$\frac{E}{2F} = \frac{Nk_B T}{F}$
σ_Y^2	$\frac{N(k_B T)^2}{F^2}$	$\frac{E^2}{8NF^2} = \frac{N(k_B T)^2}{2F^2}$

Table 3.1. Comparison of average position and variance in the microcanonical and canonical ensembles.

The above results on the fluctuations immediately imply two important consequences on the dynamical correlations in the two ensembles. First, notice that the finite value of the variance σ_Y^2 in both cases for finite N implies that the diffusion coefficient D of the piston is zero, implying that the piston remains confined. Second, the difference in the static fluctuations have repercussions on the shape of the velocity-velocity fluctuations in the canonical and microcanonical ensemble. Let us note that

$$\sigma_Y^2 = \langle (Y - \langle Y \rangle)^2 \rangle = \int_0^\infty \int_0^\infty \langle V(t')V(t'') \rangle dt' dt'', \quad (3.22)$$

where $V(t)$ is the velocity of the piston. Since σ_Y^2 are different in the canonical and microcanonical ensembles also the correlation $\langle V(t)V(0) \rangle$ must be different. These issues will be addressed in the next section, in the case of interacting gas.

Exactly the same considerations about the difference of fluctuations in the canonical and microcanonical ensembles hold in the case that a different thermodynamic limit is considered, in which the size of the piston is increased isotropically. In this case, in order to have that for each value of N the shape of the gas compartment is isotropic, namely $\langle Y \rangle = L$, and that the density $\rho = N/L^2$ and the pressure $p = F/L$ are constant, we need the scaling $F \sim \sqrt{N}$ for the force acting on the piston. If we insert such scaling for F in the equations Eq. (3.14,3.20), we find that increasing isotropically the size of the compartment, at variance with the tubular geometry, the mean square displacement σ_Y^2 of the partitioning wall becomes asymptotically constant for increasing N in the two ensembles. On the contrary the factor 2 by which canonical and microcanonical fluctuations differ remains the same. The comparison between the two different thermodynamic limits tell us on one hand that the result on the difference in canonical and microcanonical fluctuations is robust and on the other hand allows us to point out the peculiarities of the tubular geometry.

3.4 The interacting case

In order to understand whether the previous results are peculiar to the non-interacting case, and to study a more realistic case, we perform molecular dynamics simulations of the system with an interacting particle gas. We consider a repulsive interaction potential $V(\mathbf{r})$ for soft disks, with cut-off r_c

$$V(\mathbf{r}) = \begin{cases} V_0 \left[\left(\frac{r_0}{r} \right)^{12} - \left(\frac{r_0}{r_c} \right)^{12} + 12 \left(\frac{r_0}{r_c} \right)^{12} \left(\frac{r}{r_c} - 1 \right) \right] & \text{for } r < r_c \\ 0 & \text{for } r > r_c, \end{cases} \quad (3.23)$$

where $r = |\mathbf{r}|$ is the distance between particles, V_0 is the potential intensity and r_0 is the average interaction range. The same potential also describes the interaction of particles with walls. In the simulations of the canonical ensemble the coupling with the *reservoir* at temperature T is implemented in the following way. We consider that the side of the box opposite to the piston acts as a thermostat, so that when a particle enters the interaction region with the wall, namely its distance from the wall is smaller than r_0 , the velocity is changed along the y axis according to the Maxwellian distribution $p(v_y) \propto v_y \exp(-v_y^2/2mk_B T)$, for $v_y > 0$ [97]. The study of the system upon varying N is performed by retaining a tubular geometry, namely keeping the length L and the force F constant and letting the equilibrium position $\langle Y \rangle$ increase accordingly, so that the gas density remains fixed. The results here described do not depend too much on the specific interaction. Indeed, we also studied the case of a stronger interaction potential $V(r) \sim r^{-64}$, which at low density reproduces the behavior of hard-disk statistics [102], finding analogous results.

We start the numerical study of this interacting case by checking the validity of the relation (3.8). In Figure 3.2 we plot the temperature T as a function of the energy E in the microcanonical and canonical ensembles. The temperature is computed as $k_B T = M \langle V^2 \rangle$ whereas energy is $E = \langle \mathcal{H} \rangle$. As expected, the theoretical relation (3.8) derived in the non-interacting system is valid at high temperatures, where interactions become negligible. In Figure 3.3 we report the average values of the piston position and its variance in the two ensembles. Notice that also in this case the analytical predictions (3.13) and (3.19) hold in the high energy (or temperature) regions.

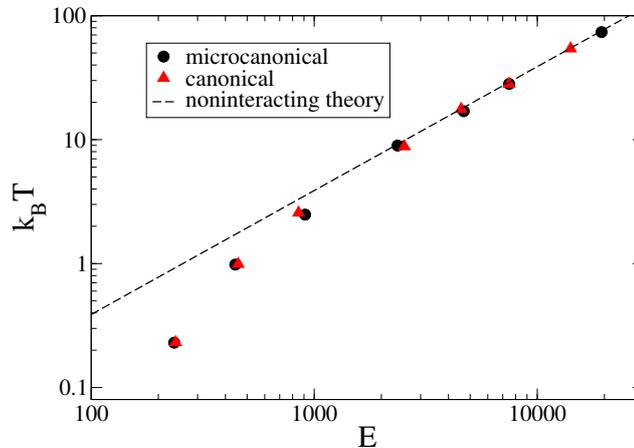


Figure 3.2. The temperature $k_B T$ (in the microcanonical ensemble is $M \langle V^2 \rangle$) is plotted as a function of energy E (in the canonical ensemble $E = \langle \mathcal{H} \rangle$) for $N = 128$. The dashed line represents the theoretical result for non-interacting particles $k_B T = E/(2N + 3/2)$, which is expected to hold for high temperatures. Other parameters in the simulations are $L = 10$, $F = 10$, $m = 1$ and $M = 128$.

It is interesting the fact that also in the interacting case the factor $1/2$ between the σ_Y^2 in the canonical and microcanonical is still present (see Fig. 3.4).

Interesting behaviors are also found for the dynamical properties of this system. Indeed, differences in the fluctuations between microcanonical and canonical are evident from the study of correlation functions. In particular, in Figure 3.5 we

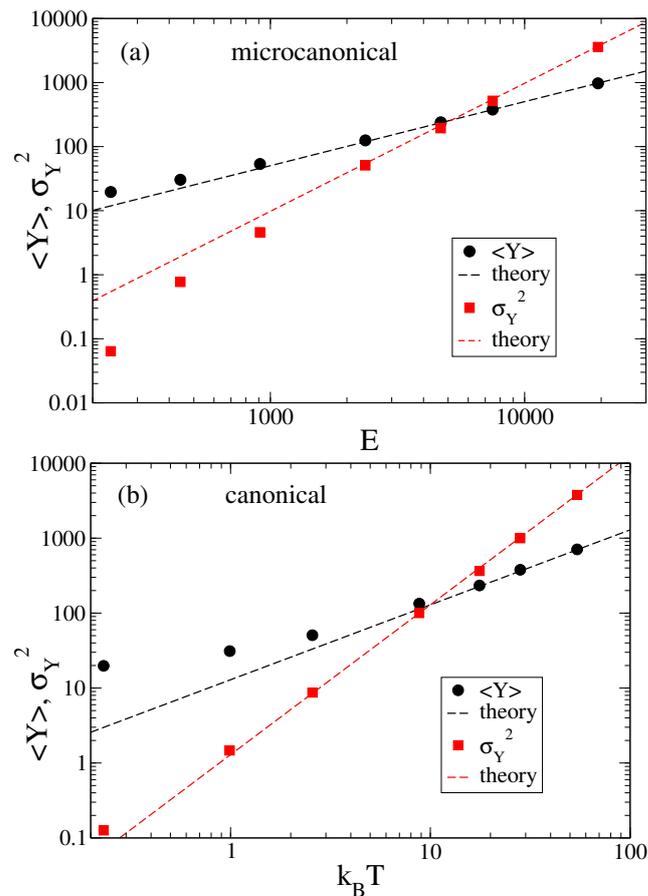


Figure 3.3. Panel (a): The average position $\langle Y \rangle$ and variance σ_Y^2 are plotted as a function of energy E in the microcanonical ensemble with $N = 128$. Dashed lines represent the theoretical results for the non-interacting gas: $\langle Y \rangle = (N + 1)E/(2N + 3/2)$ and $\sigma_Y^2 = (N + 1)(N + 1/2)/[(2N + 5/2)(2N + 3/2)^2](E/F)^2$. Panel (b): Same quantities as a function of $k_B T$ in the canonical ensemble. Theoretical results for the non-interacting gas are: $\langle Y \rangle = (N + 1)k_B T/F$ and $\sigma_Y^2 = (N + 1)(k_B T)^2/F^2$. Other parameters in the simulations are $L = 10$, $F = 10$, $m = 1$ and $M = 128$.

compare the behavior of the normalized velocity autocorrelation function of the piston, $C(t) = \langle V(t)V(0) \rangle / \langle V(0)V(0) \rangle$, for different values of N . First, one clearly observes that, as expected from the static results, fluctuations are larger in the canonical ensemble, namely the system is less correlated than in the microcanonical. Moreover, let us notice the nontrivial shape of $C(t)$. For small N one has a damped oscillatory relaxation, while, increasing N , a peculiar behavior emerges: after a first stage of relaxation, governed by a simple exponential decay, at later times a negative bump occurs, signaling the presence of another timescale in the system. This negative contribution to the correlation is necessary for the vanishing of the diffusion constant: $\int_0^\infty C(t)dt$ must be zero.

From the above results for $C(t)$, a two-time scenario emerges. We have the time τ_0 , characterizing the first exponential decay, empirically defined as the time necessary to cross the zero axis for the first time. In addition, we have the time $\tau(N)$

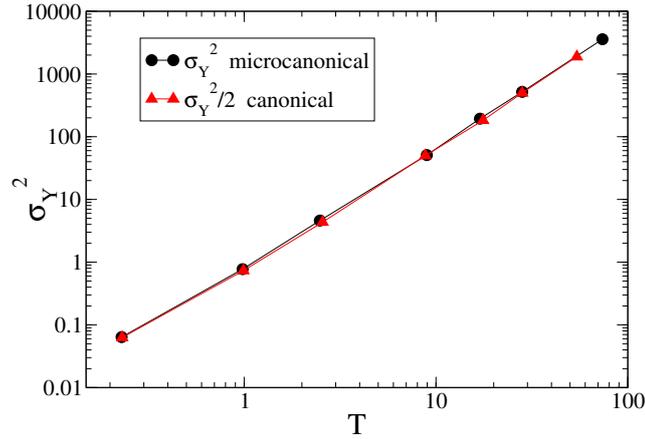


Figure 3.4. The variance σ_Y^2 is plotted as a function of $k_B T$ in the canonical and microcanonical ensembles (in the latter case we consider simulations at constant energy and the temperature is obtained from $k_B T = M \langle V^2 \rangle_E$) for $N = 128$. Other parameters in the simulations are $L = 10$, $F = 10$, $m = 1$ and $M = 128$.

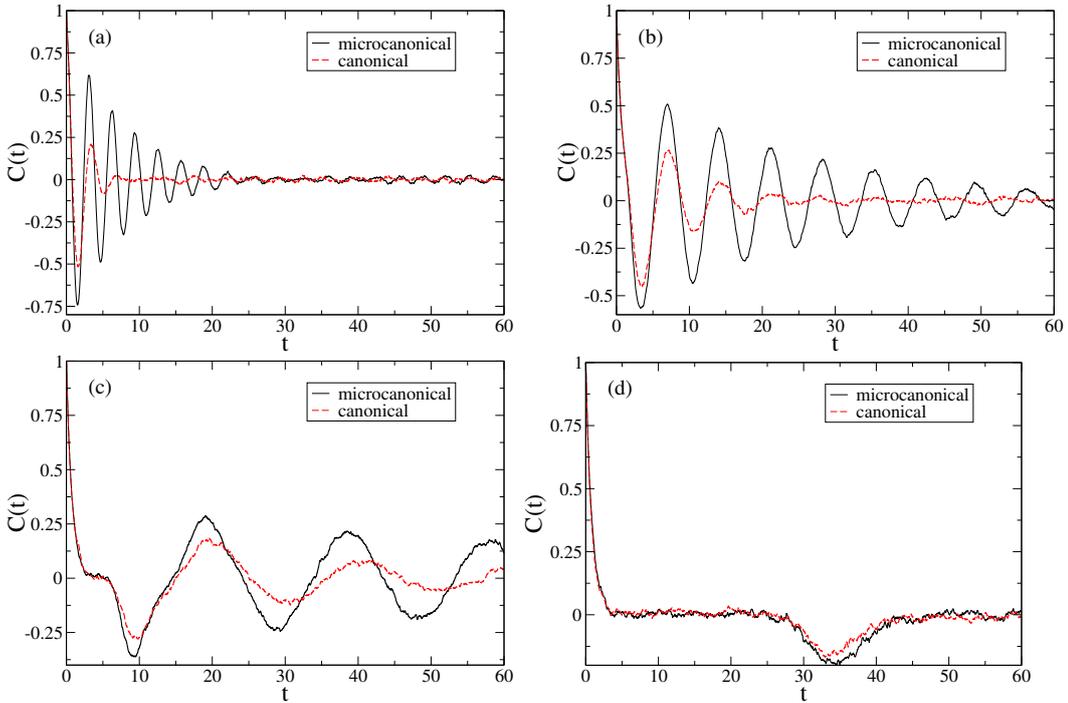


Figure 3.5. Velocity autocorrelation functions of the piston in the microcanonical and canonical ensembles, for $N = 16$, panel (a), $N = 64$, panel (b), $N = 256$, panel (c) and $N = 1024$, panel (d). Other parameters are $L = 30$, $F = 150$, $T = 10$, $m = 1$ and $M = 50$.

where the negative bump occurs. The first decay of the velocity correlation function $C(t)$ saturates upon increasing the number of particles and so the time τ_0 tends to a constant value, independent of N (see panel (a) of Figure 3.6 where τ_0 is plotted as a function of N in semilog scale, both for the microcanonical and the canonical

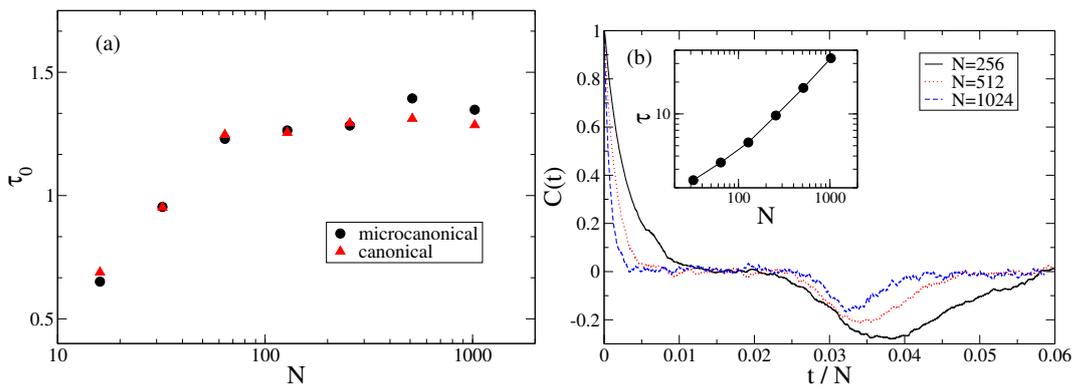


Figure 3.6. Panel (a): first relaxation time τ_0 of the piston velocity correlation for different values of N , in the microcanonical (black dot) and canonical (red squares) ensemble with parameters $M = 50$, $F = 150$, $L = 30$ and $k_B T = 10$. Notice that the first relaxation saturates for large N , and the time τ_0 reaches a constant value, both in the canonical and microcanonical ensembles. Panel (b): velocity correlation functions as a function of time rescaled by N in the canonical ensemble with same parameters. In the inset the time τ shows a linear dependence on N , for large N .

ensembles). On the other hand, we find that the second timescale τ depends linearly on N , as it is shown in panel (b) of Figure 3.6, where $C(t)$ is plotted as a function of t/N . In the inset we also plot $\tau(N)$ as a function of N in log-log scale for the canonical ensemble, showing the linear increasing with N (analogous results are observed for the microcanonical ensemble).

As discussed in the next section, such a peculiar behavior, induced by the presence of the partitioning piston, cannot be easily described by a standard Langevin-like approach.

3.5 Langevin equation

In the limit of N and M very large, the relaxation times of the piston and of the gas particles are well separated, and one may consider the gas particles weakly perturbed by the presence of the piston.

Within this strong assumption, the gas distribution is fixed and independent of the motion of the piston, and the dynamics can be described by a master equation for the probability density function $P(V, Y, t)$ from the velocity V of the piston at position Y at time t . In particular, for the first moment of this distribution, it is possible to write down the following equation (for the details refer to the Appendix 3.A):

$$\frac{d\langle V \rangle}{dt} = \langle F_{coll}(Y, V) \rangle \quad (3.24)$$

Then, the fluctuations around the equilibrium position ($Y \simeq Y_{eq}$ and $V \simeq 0$) are described by expanding up to the first order the right hand side of Eq. (3.24), obtaining

$$\frac{dV(t)}{dt} = -k_N y - \gamma V + \sqrt{2\gamma T} \eta(t), \quad (3.25)$$

where the displacement $y \equiv Y - Y_{eq}$ has been introduced. The parameters k_N and γ can be calculated by means of kinetic theory, and their explicit expressions are written in Eq. (3.37) of Appendix 3.A. One must notice that in Eq. (3.25) a noise term $\eta(t)$ has been added, whose expression cannot be directly derived from the Eq. (3.24) for the mean velocity. Actually, the correlation of the noise term can be determined by exploiting equipartition theorem valid for equilibrium dynamics. By requiring Maxwellian statistics for the stationary $P(V)$, it is well known that $\xi(t)$ must be white noise with variance

$$\langle \eta(t)\eta(t') \rangle = \delta(t - t'). \quad (3.26)$$

From the linearity of Eq. (3.25) it is possible to calculate the autocorrelation of velocity, obtaining:

$$\langle V(t)V(0) \rangle = \frac{T}{M} e^{-\frac{\gamma t}{2}} \left[\cosh\left(\frac{\Delta}{2}t\right) - \frac{\gamma \sinh\left(\frac{\Delta}{2}t\right)}{\Delta} \right] \quad (3.27)$$

where we introduced the parameter $\Delta = \sqrt{\gamma^2 - 4k_N}$, which rules the passage between underdamped and overdamped regime. More specifically, if $\frac{Nm}{m+M} > \frac{\pi}{2}$, the system is overdamped, else the system is underdamped.

Making a comparison between Eq. (3.27) and the numerical experiments presented in Fig. 3.5, it appears evident that the Langevin equation is able to capture, for N large, only the small time relaxation $\tau_0 \simeq \gamma^{-1}$, while is unable to detect the oscillation of $\langle V(t)V(0) \rangle$, that appears for times $\tau(N) \sim N$. We report in Fig. 3.7 the explicit comparison between the Langevin approximation (black curve) and the piston velocity correlation (red curve) in the non-interacting case. The same mismatch between analytical prediction and numerical results is observed also for interacting particles. The oscillations presented by $\langle V(t)V(0) \rangle$ are related with the interplay mechanism between the moving wall and a collective mode of the gas particles, that make the assumption of Markovianity to fail. We note how this phenomenon is quite general and it is present also in the case of non-interacting gas particles. In order to verify this point, one can analyze a natural collective variable of the gas, i.e. the center of mass velocity $v_{cm}(t) \equiv \frac{1}{N} \sum v_i(t)$. In the simpler case of a non-interacting gas confined in a fixed volume, the autocorrelation $\langle v_{cm}(t)v_{cm}(0) \rangle$ would be trivially equal to the one of a single particle in the gas. On the contrary this is not true anymore with the presence of the piston, since the different particles of the gas strongly correlates each other via the mutual interaction piston/border. The time scale of this process is very close to $\tau(N)$, as it can be observed in Fig. 3.7. Such a time scale is completely hidden if one consider only the single particle autocorrelation $\langle v_i(t)v_i(0) \rangle$.

3.6 Conclusions

In the present Chapter we have shown, with analytical calculations in the ideal gas case and with simulations for interacting particles, that the fluctuations in the canonical and microcanonical ensembles [26] show relevant differences when a partitioning object, like a moving wall, is introduced. The relevant points that we

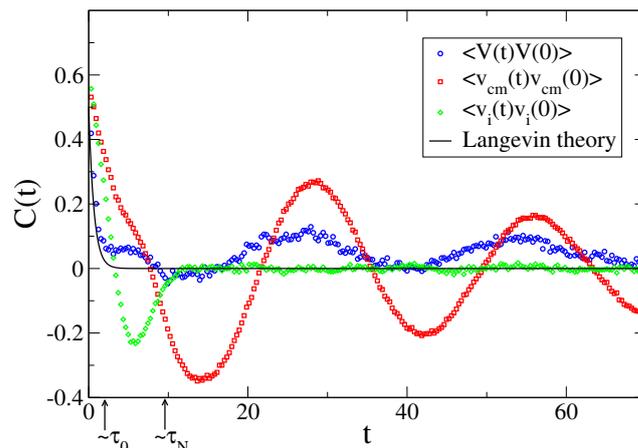


Figure 3.7. Autocorrelation of different observables in the case of a piston with non-interacting particles (canonical ensemble) measured in numerical simulations and Langevin approximation for the piston velocity correlation (black line). It is possible to observe how the oscillation in the autocorrelation of the piston velocity $\langle V(t)V(0) \rangle$ (blue circles) are in phase with the one of the center of mass of the gas particles $\langle v_{cm}(t)v_{cm}(0) \rangle$ (red squares). With green diamonds is represented the autocorrelation of a single particle velocity $\langle v_i(t)v_i(0) \rangle$. All the correlations are normalized to one for $t = 0$. Value of the parameters: $F = 150$, $T = 10$, $M = 50$, $N = 500$.

have highlighted are the following. First, we have shown that the interaction with the partitioning object induces nontrivial correlations among the particles even in the ideal gas approximation, see Fig. 3.7 in Sec. 3.5, irrespectively of the ensemble, canonical or microcanonical, where the dynamics is studied. Then, we have shown that the Langevin approach to the dynamics of the piston captures only partially the physics of the system. The Langevin equation, correctly predicts only the fast time scale, namely $\tau_0 \sim \gamma^{-1}$, but fails completely to catch the slower one, which grows linearly with the number of particles in the partitioned system, $\tau(N) \sim N$. This second time scale is produced by non-trivial correlations among the velocity of the gas particles and the one of the piston which are present, quite remarkably, also in the case of non-interacting particles, as shown in Fig. 3.7.

We recall that the macroscopic growth of $\tau(N)$ is related to the particular tubular geometry of the problem, where the size of the gas compartment is increased only in one direction. Notwithstanding the different behavior of the largest timescale, the factor 2 of difference between canonical and microcanonical fluctuations of the partitioning object σ_Y^2 , is independent from how the thermodynamic limit is taken.

We can therefore conclude that partitioning geometries with a single *macroscopic* degree of freedom which is *effectively* coupled to the motion of all the *microscopic* constituents of the system represent an eligible framework to study the dynamical properties of small systems.

3.A Derivation of the Langevin Equation

In this Appendix, following elementary kinetic theory, we detail the derivation of the Langevin equation for the motion of the piston. The basic idea is to compute the average force exerted by the gas particles which collide with the piston, by calculating the average momentum exchanged in the collisions. The following approach dates back to Smoluchowski [104] and it has been used to write a Langevin equation for colloidal particles [32]. For the variable $y = Y - Y_{eq}$ we will derive a stochastic equation

$$M \frac{d^2 y}{dt^2} = F_{av}(y, \dot{y}) + C\eta, \quad (3.28)$$

where $F_{av}(y, \dot{y})$ is the average force acting on the piston in the position $Y_{eq} + y$ and velocity \dot{y} , η is a white noise and the constant C can be fixed *a posteriori* from the condition $M \langle \dot{y}^2 \rangle = k_B T$.

Consider the gas at equilibrium, and focus on the collision of the piston, characterized by its mass M and precollisional velocity V , and a particle of the gas, which are characterized by m and \mathbf{v} , respectively. The collision rule is

$$V' = V + \frac{2m}{m+M}(v_y - V) \quad v'_y = v_y - \frac{2M}{m+M}(v_y - V) \quad (3.29)$$

where the primed quantity are postcollisional velocities, and v_y is the y -component of \mathbf{v} . The rate of such collisions can be obtained by considering the equivalent problem of a piston, at rest, hit by a flux of particles moving at relative velocity $V\hat{\mathbf{y}} - \mathbf{v}$. The rate is then determined by counting the number of point-like particles hitting the unit surface in the infinitesimal time interval dt . This number corresponds to the particles contained in a rectangle of infinitesimal base length δx and height $(v_y - V)\Theta(v_y - V)dt$. The step function $\Theta(s)$ selects the condition for having a collision. Setting $v = v_y$, the mean force exerted by the particles of the gas on the piston is

$$\begin{aligned} F_{coll}(Y, V) &= \left\langle M \frac{\Delta V}{dt} \right\rangle \\ &= M \int_{-\infty}^{\infty} dv \int_0^L dx \rho(x, Y - r'_0) \\ &\times \phi(v)(V' - V)(v - V)\Theta(v - V) \\ &= \frac{2mM}{m+M} \\ &\times \int_{-\infty}^{\infty} dv \int_0^L dx \rho(x, Y - r'_0) \phi(v)\Theta(v - V)(v - V)^2 \end{aligned} \quad (3.30)$$

where $\phi(v)$ is the equilibrium distribution of velocities of the gas, i.e. $\phi(v) = \sqrt{\frac{m}{2\pi k_B T}} e^{-\frac{mv^2}{2k_B T}}$ and $\rho(x, Y)$ is the spatial density of particles in the proximity of the piston. At equilibrium, this density is uniform on all the available volume and, therefore, depends on the position of the piston Y . Carrying on the integration on the spatial coordinates, we obtain

$$F_{coll}(Y, V) = \frac{2mM}{m+M} \lambda \int_V^{\infty} dv (v - V)^2 \phi(v) \quad (3.31)$$

where $\lambda = \frac{N}{Y}$. We note that the equilibrium properties of the gas used in the derivation of this equation don't depend on the choice of the ensemble. Of course, $F_{av}(y, \dot{y})$ is nothing but $F_{coll} - F$.

In order to decouple the motion of the piston from the one of the gas molecules it's necessary to assume that $M \gg m$ and that, moreover, V is always small if compared to the thermal velocity of the particles $v_m = \sqrt{\frac{2k_B T}{m}}$: the expansion of the integral in Eq. (3.31) in powers of $\sqrt{\frac{m}{M}}$, will give the viscous drag force appearing in the Langevin equation of motion. Defining $g = \sqrt{\frac{m}{2k_B T}}(v - V)$ and expanding perturbatively $\phi(v)$ as a function of g

$$\begin{aligned} e^{-\frac{m}{2k_B T}v^2} &= e^{-\left(g + \sqrt{\frac{m}{2k_B T}}V\right)^2} \simeq e^{-g^2 - \sqrt{\frac{2m}{k_B T}}gV} \\ &\simeq e^{-g^2} \left(1 - \sqrt{\frac{2m}{k_B T}}gV\right) \end{aligned} \quad (3.32)$$

we can compute the integral, performing the change of variables $v \rightarrow g$

$$\frac{2k_B T}{m\sqrt{\pi}} \int_0^\infty g^2 e^{-g^2} \left(1 - \sqrt{\frac{2m}{k_B T}}gV\right) dg = \frac{k_B T}{2m} - \sqrt{\frac{2k_B T}{\pi m}}V$$

namely

$$F_{coll} = \frac{N}{Y} \left[\frac{M}{m+M} k_B T - 2 \frac{M}{m+M} \sqrt{\frac{2mk_B T}{\pi}} V \right] \quad (3.33)$$

Expanding the previous expression at the first order in y and V around the equilibrium position of the piston Y_{eq} , defined by the condition $F = F_{coll}$ and $V = 0$, we obtain a linear Langevin equation. The equilibrium conditions are

$$\frac{M}{m+M} k_B T \frac{N}{Y_{eq}} = F \quad \text{and} \quad V_{eq} = 0 \quad (3.34)$$

and therefore

$$Y_{eq} = \frac{NMk_B T}{F(m+M)}. \quad (3.35)$$

The Langevin equation has the shape

$$\frac{d^2 y}{dt^2} = -k_N y - \gamma v + \sqrt{2\gamma T} \eta, \quad (3.36)$$

where

$$\gamma = \frac{2F}{M} \sqrt{\frac{2m}{\pi k_B T}} \quad \text{and} \quad k_N = \frac{F^2(m+M)}{M^2 N k_B T}, \quad (3.37)$$

and the coefficient in front of the fluctuating term is imposed in order to restore detailed balance (for a detailed discussion see Sec. 4.A). It is easy to compute the correlation function

$$\langle V(t)V(0) \rangle = \frac{k_B T}{M} e^{-\frac{\gamma}{2}t} \left[\cosh\left(\frac{\Delta}{2}t\right) - \frac{\gamma}{\Delta} \sinh\left(\frac{\Delta}{2}t\right) \right], \quad (3.38)$$

where $\Delta = \sqrt{\gamma^2 - 4k_N}$. Let us note that for any finite N (i.e. $k_N \neq 0$) one has $\int_0^\infty \langle V(t)V(0) \rangle dt = 0$.

A similar, less formal, derivation, when considering a uni-dimensional version of the system and by adding the internal temperature as a relevant slow degree of freedom is reported in Sec. 4.A.

Chapter 4

The piston engine

4.1 Introduction

In this Chapter we study a model of small heat engine, which is obtained by driving the system described in the previous Chapter far from equilibrium. This is done by introducing an explicit time dependency of the external parameters appearing in the piston-gas model: we let the temperature of the external thermostat and the external force which confines the system to vary in a cyclical manner, the protocol being repeated many times such that, in each cycle, a fraction of the heat absorbed from the thermostat may be converted into useful mechanical power, i.e. work.

In the literature about small systems thermodynamics a prominent role is covered by models, often inspired by minimal experiments at the microscale, with very few degrees of freedom, where typically one has $N = 1$: the overdamped dynamics of the position of a colloid in a non-conservative (e.g. time-dependent) potential is a seminal prototype [92]. Only a few studies have discussed the non-trivial effects of inertia [84, 21] where the relevant degrees of freedom are at least two (also with different parities under time-reversal). It is even more rare to find models with $N \gg 1$, still remaining in the domain of small N : for instance with an order of magnitude $N \sim 10^2$ fluctuations can still be relevant and possibly non-trivial, while the complexity of the dynamics is hugely raised. Such numbers are also closer to real biophysical applications with macromolecules, nanocapillaries, etc. [13]. On the front of the statistical mechanics of molecular models, an exception is certainly represented by the study in [54, 55], where a model similar to our proposal is analyzed, with the only exception that the piston's inertia is not considered because its (time-dependent) position is externally controlled during the cycle. In our model, on the contrary, we only fix the (time dependent) force acting on the piston, so that the piston's velocity is determined by the effect, mediated by inertia, of such a force and of the collisions with the gas' molecules. Because of a larger freedom in the piston's dynamics, this "machine" displays a much richer diagram of phases. In particular, the choice of τ determines different working regimes: engine, refrigerator and heat pump. In addition, also in view of an analysis of the linear regime of the engine similar to the one discussed in [55], in this model it is possible to disentangle the smallness of the external perturbation (represented by the excursion of forces and temperatures) with the slowness of the transformations (represented by the total time of the cycle).

In the present Chapter we study the performance of such an engine, by determining the dependence of the working regime on the total duration of the cycle. We also introduce an improved coarse grained description of the system in which an additional slow variable is used, i.e. the average kinetic energy of the gas. Unfortunately, despite increasing the level of accuracy of the model, this equation does not allow to obtain analytical expression for the performance of the engine: for this reason in order to compute the average work and heats we must resort to the simple, two-variables, coarse-grained equation of Chapter 3. A comparison with the result of numerical simulations will show that, despite the discrepancies already highlighted above, at least at a qualitative level, the analytic results reproduce the physics of the system.

We also introduce a formalization of the model in terms of fluxes and thermodynamics forces which allows to distinguish between a linear and a non-linear regime. In the linear regime, we get a matrix of Onsager coefficients which non-trivially depend on τ , a fact usually ignored in the recent literature. Along those lines, we can widen the study of the power optimization, considering different procedures of maximization (e.g. by varying different parameters, including the cycle duration τ) and comparing the results with the CA estimate for the efficiency.

Eventually, in the last Section of the Chapter, we report some analytical results regarding the fluctuations of the thermodynamic fluxes, obtained in the context of Large Deviation Theory.

4.2 The Mechanical Model (MM)

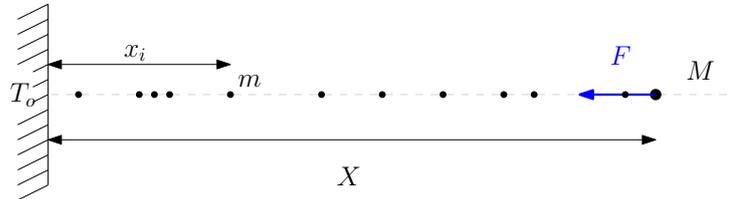


Figure 4.1. Sketch of the piston model. A gas of particles is confined by a fixed wall (the thermal bath) and a moving wall (the piston) that is subject to a constant external force.

In order to reduce the complexity of the system introduced in the previous Chapter, we will not consider an interaction among the particles: for this reason we can reduce the dimensionality of the model. In fact, in this new model, independently of the real dimensionality of the box, only the motion in the \hat{x} direction is relevant, as the particles interact only with the piston (see Fig. 4.1). For this reason we will consider a one-dimensional version of the previous model which allows to simplify some computations and numerical simulations. The full (one-dimensional) hamiltonian of the system reads

$$\mathcal{H} = \sum_{i=1}^N \frac{p_i^2}{2m} + \frac{P^2}{2M} + FX, \quad (4.1)$$

with the additional constraints $0 < x_i < X$, $i = 1, \dots, N$, and $X > 0$. The interaction potential between the piston and the particles is replaced with elastic

collisions, i.e. we assume that whenever the piston and one particle collide (have the same position) the velocities change with the law

$$\begin{aligned} V' &= V + \frac{2m}{m+M}(v_i - V), \\ v'_i &= v_i + \frac{2M}{m+M}(V - v_i), \end{aligned} \quad (4.2)$$

where v'_i and V' are post-collisional velocities. The wall at $x = 0$ acts as a thermostat at the temperature T_o (the subscript “o” is added in this Chapter in order to distinguish the temperature of the thermostat from that of the system): a collision of a particle with the wall is equivalent to give a new velocity v' to the particle with probability density

$$\rho(v') = \frac{m}{T_o} v' e^{-\frac{mv'^2}{2T_o}} \Theta(v), \quad (4.3)$$

where $\Theta(v)$ is the Heaviside Theta. Let us note that the presence of the piston introduces an interaction among the gas particles: two or more particles colliding successively with the piston exchange information because the velocity of the piston is reminiscent of previous collisions. Furthermore, for this reason, the dynamics of the system depends on the number of particles N (the system is not a collection of N non-interacting subsystems), and systems with different sizes will have different behaviors. We study the system in a range of Nm/M close to ~ 1 , meaning that there is a non-trivial interplay between the gas and the piston. Hereafter we use arbitrary units in numerical simulations, and we put $k_B = 1$ for the Boltzmann factor. At fixed F and T_o , the study of the system in the canonical ensemble by means of standard statistical mechanics, see Chapter 3 and Ref. [37] reveals that $\langle X \rangle^{eq} = (N+1)T_o/F$ and $\sigma_X^2 = (N+1)T_o^2/F^2$. In addition, if we define the estimate of the instantaneous temperature of the gas,

$$T = \frac{1}{N} \sum_{i=1}^N m v_i^2, \quad (4.4)$$

the ensemble average of this quantity reads $\langle T \rangle^{eq} = T_o$ and its variance is $\sigma_T^2 = 2T_o^2/N$.

4.2.1 The engine protocol

When the parameters F and T_o vary in time, mechanical work can be extracted from the system. In particular if we identify the (thermodynamical) internal energy of the system with the value of the hamiltonian, $E(t) = \mathcal{H}(\mathbf{x}(t), t)$, and, in addition, we define the input power as

$$\dot{W} = \left. \frac{\partial \mathcal{H}}{\partial t} \right|_{\mathbf{x}(t)}, \quad (4.5)$$

conservation of energy simply reads $\dot{E}(t) = \dot{Q}(t) + \dot{W}(t)$, where Q is the energy absorbed from the thermal wall. For an hamiltonian \mathcal{H} as in Eq. (4.1) one gets $\dot{W} = X\dot{F}$. Let us remark that this formula is different from the one obtained in standard thermodynamics, $\dot{W} = F\dot{X}$: this is due to the fact that we included the energy of the piston in the internal energy of the system [57]. For our choice of

thermostat discussed before, the formula for the energy adsorption (heat flux) can be formally written as $\dot{Q}(t) = m \sum_i \delta(t - t_i) [(v'_i(t))^2 - (v_i(t))^2]$ where t_i are the times of collisions between the gas particles and the thermostating wall at $x = 0$, whereas v_i and v'_i are the velocities before and after a collision respectively.

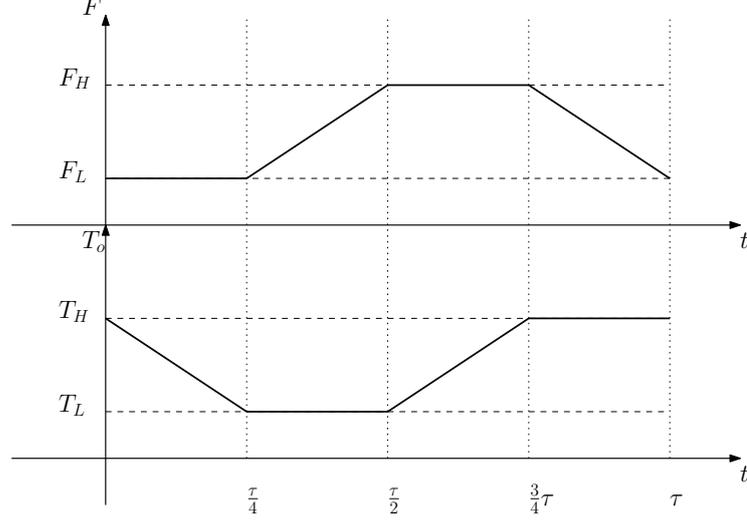


Figure 4.2. Graph of F and T_0 as a function of time over a cycle period τ .

Here we adopt the following cyclical protocol to obtain a heat engine: the parameters vary periodically in time over a cycle of length τ (see Fig. 4.2 and the inset of Fig. 4.3 for a visual explanation). If we set $t_0 = k\tau$, with k integer, the cycle has the following form:

- I) At times $t \in [t_0, t_0 + \tau/4]$: isobaric compression ($F(t) = F_L$ and $T_o(t) = T_H - 4(T_H - T_L)(t - t_0)/\tau$),
- II) At times $t \in [t_0 + \tau/4, t_0 + \tau/2]$: isothermal compression ($F(t) = F_L + 4(F_H - F_L)[t - (t_0 + \tau/4)]/\tau$ and $T_o(t) = T_L$);
- III) At times $t \in [t_0 + \tau/2, t_0 + 3\tau/4]$: isobaric expansion ($F(t) = F_H$ and $T_o(t) = T_L + 4(T_H - T_L)[t - (t_0 + \tau/2)]/\tau$);
- IV) At times $t \in [t_0 + 3\tau/4, t_0 + \tau]$: isothermal expansion ($F(t) = F_H - 4(F_H - F_L)[t - (t_0 + 3\tau/4)]/\tau$ and $T_o(t) = T_H$).

The main reason for such a protocol is to have a system which is always thermostatted, i.e. always near a canonical ensemble (at least for slow transformations): on the contrary, adiabatic transformations require a micro-canonical analysis which may become less transparent. In view of a Langevin-like analysis (see below) this protocol (called second type Ericsson cycle) – which is thermostatted for the whole duration of the cycle – is simpler than the more classical Carnot cycle. A similar model has been studied in [54] with a crucial difference: there the velocity of the piston is fixed at any time and cannot fluctuate (this choice corresponds to the infinite mass M limit).

The cycle of length τ is repeated a large number of times over a long trajectory. In the following, unless differently specified, we use the symbols W and Q to mean a time-integral over a cycle, i.e. $W = \int_t^{t+\tau} \dot{W}(s) ds$ and $Q = \int_t^{t+\tau} \dot{Q}(s) ds$. Due to the stochastic nature of collisions and of the random choice of initial conditions, W and Q are random variables. Conversely, the symbols $\langle W \rangle$ and $\langle Q \rangle$ indicate the average work and heat per cycle computed over a long (single) run composed of a large number of cycles. Due to the periodic nature of the protocol the system will reach, after a transient, a periodic asymptotic state with a probability distribution in the full phase space depending on time only through $t' = t \bmod \tau$. Therefore, thanks to the ergodic hypotheses, the average denoted by $\langle \cdot \rangle$, is equivalent to an average over the aforementioned periodic distribution.

The study of the thermodynamics of the engine in the quasi-static limit, i.e. assuming that the system is always at equilibrium: $X(t) = \langle X \rangle_{F(t), T_o(t)}^{eq}$ and $E(t) = \langle \mathcal{H} \rangle_{F(t), T_o(t)}^{eq}$, leads to the formula in Table 4.1.

Segment	$\langle W \rangle$	$\langle Q \rangle$
I)	0	$\frac{3}{2}(N+1)(T_L - T_H)$
II)	$(N+1)T_L \ln\left(\frac{F_H}{F_L}\right)$	$-\langle W \rangle$
III)	0	$\frac{3}{2}(N+1)(T_H - T_L)$
IV)	$-(N+1)T_H \ln\left(\frac{F_H}{F_L}\right)$	$-\langle W \rangle$

Table 4.1. Table with the adiabatic values of Q and W in each segment of the Ericsson cycle. The average $\langle \cdot \rangle$ is intended over many realization of the cycle.

During segments I) and III) no work is done on the system and the heats exchanged have same magnitude but opposite signs. Therefore, in the quasi-static limit $\tau \rightarrow \infty$, there is no net heat exchange with the intermediate reservoirs at temperature $T_C < T^* < T_H$. Therefore, one may assume that the two isobaric transformations do not contribute to the net exchange of heat and work: this is true for $\tau \rightarrow \infty$ and seems reasonable, for reasons of symmetry, at large τ , while (small) discrepancies at finite τ are observed in the simulations. For this reason we may identify the input heat $Q_{in} = Q_2$ with the energy absorbed from the hot reservoir at T_H in segment IV), the dissipated heat $Q_{diss} = Q_1$ with the energy released into the cold thermostat T_C in sector II) and, furthermore, assume $Q = Q_{in} + Q_{diss}$.

The more refined definition (Eq. (2.72)) of input heat for the case of thermostats with continuous varying temperatures introduced in Sec. 2.3.1 will be discussed in the following, in the context of an Onsager analysis of the system. The two definitions are substantially the same from a numerical point of view, with the only difference that the simpler definition is easier to be measured in numerical simulations. If $\langle Q_2 \rangle > 0$ and $\langle W \rangle < 0$, efficiency can be defined as

$$\eta = -\frac{\langle W \rangle}{\langle Q_2 \rangle}, \quad (4.6)$$

where W is the total work on a cycle, and $\langle \cdot \rangle$ denotes the average over many realizations of the cycle. Let us remark that this quantity is different from the average over many cycle of the fluctuating efficiency $\hat{\eta} = W/Q_2$.

4.3 The “phase diagram” of the engine

In order to perform molecular dynamics simulation of the system with time-dependent parameters F and T_o (where event-driven simulations are not possible), we introduce an interaction potential between the piston and the particles to mimic the effect of elastic collisions. We choose the repulsive soft sphere potential with cut-off radius r_0 similar to the one introduced in Chapter 3:

$$V(r) = \left[\left(\frac{r_0}{r} \right)^{12} + 12 \frac{r}{r_0} - 13 \right] \Theta(r_0 - r), \quad (4.7)$$

where $\Theta(r)$ is the Heaviside Theta. The parameter r_0 is to be chosen as small as possible, compatibly with integration time-step Δt , in order to simulate a contact interaction. In our case $r_0 = 0.2$ and $\Delta t = 0.0005$. The values of the other parameters, if not explicitly mentioned, are $m = 1$, $M = 100$, $N = 500$, $T_L = 11$, $T_H = 13$, $F_L = 180$, $F_H = 220$. The integration scheme adopted is based on the standard Verlet algorithm.

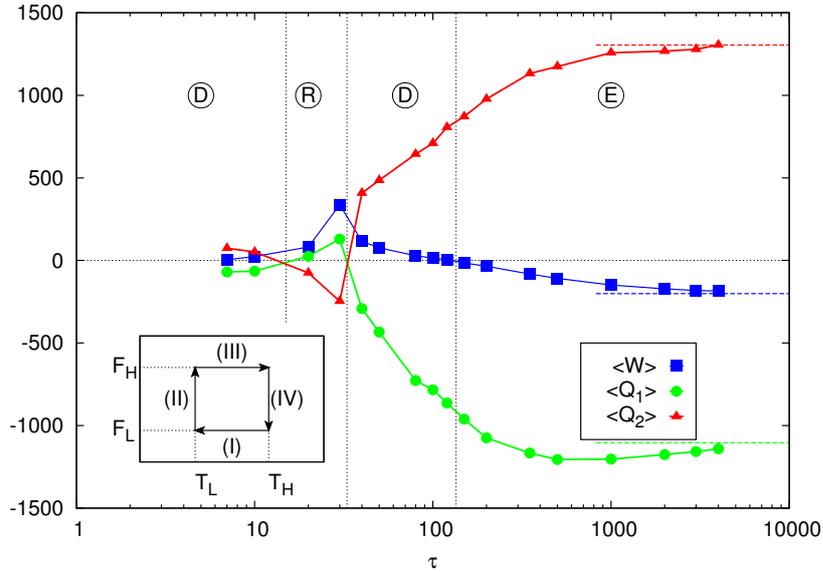


Figure 4.3. Average values per cycle of work W and heats Q_1 , Q_2 as a function of the cycle typical time τ . Dashed horizontal lines represent the adiabatic value of such quantities. Inset: schematic of the cycle protocol in the space of parameters F, T_o .

4.3.1 Average work and heats as functions of τ

In Fig. 4.3 we report the average values (over 500 cycles) of W , Q_1 and Q_2 as a function of τ . In the adiabatic limit, $\tau \gg 10^3$, we recover, for $\langle W \rangle$, $\langle Q_1 \rangle$ and $\langle Q_2 \rangle$, the values computed assuming quasi-static transformations in thermodynamics. At finite values of the cycle’s duration $\tau < 10^3$ quite a complex scenario emerges. The absolute value of $\langle W \rangle$ decreases upon reducing τ , until it vanishes at a stall time $\tau^* \sim 150$. For shorter cycles, the engine consumes work instead of producing it (the

regime at $\tau > \tau^*$ is marked, on Fig. 4.3, as “E”=engine). At smaller τ , the analysis of the heats reveals the existence of three regimes, marked on the Figure as “D”, “R” and again “D”. In the “R” regime the system acts as a refrigerator, i.e. consumes work to push heat from T_L to T_H . In the “D” phases, the heat flow is the standard one (from T_H to T_L), even if work is consumed: however, the rate of heat transfer $\langle Q_2 \rangle / \tau$ is higher than in the “E” phase, and therefore the machine acts as a more efficient heat sink, similar to dissipating fans. At a time $\tau_{res} < \tau^*$ we notice the presence of a maximum in $\langle W \rangle$: it is of the order of magnitude of the adiabatic limit, but with opposite sign. At smaller $\tau \rightarrow 0$ the consumed work goes to 0. Let us note that the relevant timescales emerged from this analysis are in fair agreement with the characteristic relaxation times computed in a simple Langevin model of this system, see below.

4.3.2 Fluctuations

In small systems, fluctuations are hardly negligible [105]. In Fig. 4.4 (A and B, red curves), we display the behavior of fluctuations of work W integrated in a cycle for two different regimes, at $\tau = 50 < \tau^*$ and $\tau = 500 > \tau^*$. Deviation from a Gaussian behavior are small, indicating that N , even if finite, is large enough to expect the validity of the central limit theorem. Interestingly the measure of the standard deviation (stdev) σ_W rescaled by the average value $\delta W = \sigma_W / |\langle W \rangle|$ (black curve in Fig. 4.4D) shows that $\delta W \ll 1$ close to τ_{res} and $\delta W \gg 1$ at the stall time τ^* . The relative stdev for the heat, $\delta Q_2 = \sigma_{Q_2} / |\langle Q_2 \rangle|$ behaves much more regularly. It is also interesting to analyze the fluctuation of the “fluctuating efficiency”, i.e. $\hat{\eta} = -W/Q_2$ measured in a single cycle, see Fig. 4.4C (restricted to positive values), which displays a long tail for values larger than the average [100]. At the end of this Chapter, Sec. 4.8 by means of large deviation theory, we will try to reproduce analytically the behavior of the fluctuations of the work as a function of τ , reported in Fig. 4.4D.

4.4 Coarse-grained description

In order to make contact with stochastic thermodynamics [92], which is a useful framework for small systems, we need a coarse-grained description with few relevant (slowly-changing) variables. The contribution of the fast degrees of freedom is in the noise. Since we showed in the previous Chapter that a two-variables description (X, V) of the system do not correctly reproduce the dynamics of the system, we try do derive a more refined model by increasing the number of slow degrees of freedom. Reasonable candidates are: the position of the piston X , its velocity V and the estimate of the instantaneous “temperature” of the gas $T(t) = m/N \sum_{i=1}^N v_i^2(t)$. Unfortunately, the disadvantage of increasing the accuracy of the coarse-grained description of the system, is that the three-variables (3V) model will not allow to perform analytic calculations.

The time evolution of these observables can be determined by computing the average rate of collision occurring between the particles of the gas and the walls of the container. Here, at any t , we assume the gas to be homogeneously distributed

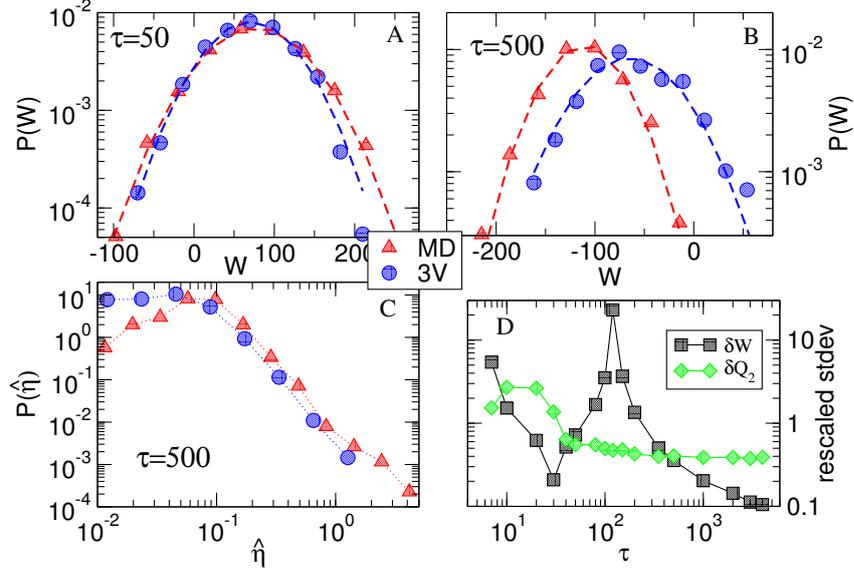


Figure 4.4. Study of fluctuations. A) and B): Pdf of the work in a cycle for two different values of τ (“E” and “D” regimes), from the MD and from the reduced “3V” model Eq. (4.11). C) Pdf of the fluctuating efficiency in a cycle at $\tau = 500$. D) rescaled stdev (see text) of $P(W)$ and of $P(Q_2)$ as a function of τ . The statistics in this Figure is obtained from 2000 cycles.

in the interval $(0, X(t))$ and each particle to have a velocity v , given by a Maxwell-Boltzmann distribution $\rho_T(v)$ at the temperature $T(t)$. In addition we use the fact that the collisions between the gas particles and the piston are elastic and that a particle that collides with the thermal wall gets a new velocity v' distributed according to a Maxwellian distribution $\rho_o(v')$. Taking into account the contributions of the external force and the collision, we have that the average derivative of the velocity of the piston $\langle \dot{V} \rangle = \lim_{\Delta t \rightarrow 0} \langle \Delta V \rangle / \Delta t$ is

$$\langle \dot{V} \rangle = -\frac{F}{M} + \frac{N}{X} \int dv \frac{2m(v-V)^2}{(M+m)} \Theta(v-V) \rho_T(v). \quad (4.8)$$

On the other hand, $\langle \dot{T}(t) \rangle$ is the sum of two terms coming from the collisions with the piston

$$\langle \dot{T} \rangle \Big|_{coll} = \frac{m}{X} \int dv (v'^2 - v^2) |v - V| \Theta(v - V) \rho_T(v), \quad (4.9)$$

where v' is the velocity after an elastic collision, and the interaction of the gas with the thermostat

$$\langle \dot{T} \rangle \Big|_{ther} = \frac{m}{X} \int dv dv' (v'^2 - v^2) |v| \Theta(-v) \rho_T(v) \rho_o(v'). \quad (4.10)$$

A detailed version of this derivation is reported in Sec. 4.A. In order to reduce the dynamics to a linear Langevin equation we assume the fluctuations of X, V and T to be small (such assumptions are reasonable if $N \gg 1$ and $M \gg m$) and expand Eqs. (4.8), (4.9) and (4.10) up to the first order around the equilibrium values

$X_{eq} = NT_o/F$, $V_{eq} = 0$ and $T_{eq} = T_o$. The linearity of the equation is, on one hand, inspired by the gaussianity of pdfs, and, on the other, it is a useful assumption that allows simple computations. The stochastic part is obtained by adding the gaussian noise terms with amplitudes determined by imposing that the variances of the variables coincide with those computed within the canonical ensemble and reported in Chapter 3. This yields

$$\begin{aligned}\dot{X} &= V, \\ \dot{V} &= -k(X - X_{eq}) - \gamma V + \mu(T - T_{eq}) + \sqrt{\frac{2\gamma T_o}{M}}\xi_1, \\ \dot{T} &= -\frac{2MT_o}{N}\mu V - \alpha(T - T_{eq}) + \sqrt{\frac{4\alpha T_o^2}{N}}\xi_2,\end{aligned}\quad (4.11)$$

where ξ_1 and ξ_2 are independent white noises $\langle \xi_i \rangle = 0$, $\langle \xi_i(t)\xi_j(t') \rangle = \delta_{ij}\delta(t - t')$, $k(t) = F(t)^2/[MNT_o(t)]$, $\gamma(t) = 2F(t)\sqrt{2m/[M^2\pi T_o(t)]}$, $\mu(t) = F(t)/[MT_o(t)]$ and $\alpha(t) = F(t)\sqrt{2/[mN^2\pi T_o(t)]}$. A numerical study of the “3 variables” (3V) model in Eq. (4.11) reveals a fair agreement with our main observations. In Fig. 4.5 the average values per cycle of work and heats are compared with those obtained in the original MD: we define \dot{W} and \dot{Q} as in MD, with $E(t) = NT(t)/2 + MV(t)^2/2 + F(t)X(t)$. The maximum and the inversion of the average work are fully reproduced, but with significant shifts of the values of τ where they occur. Indeed, a more detailed analysis (not reported here) has identified the relevance of two additional variables: taking into account the position and the velocity of the center of mass of the gas, it is possible to achieve a better agreement with the MD. Unfortunately, the parameters for such a “5 variables” model can only be obtained by fitting the MD data. Notwithstanding its degree of approximation, the 3V model gives a fair account of fluctuations, see Fig. 4.4 A and B, which have a very similar Gaussian shape and width. The overall shape of the efficiency fluctuations’ pdf (Fig. 4.4C) is also reproduced. The eigenvalues of the dynamic’s matrix in Equations (4.11) give also access to typical timescales. For instance, for $F = 200$, $T = 12$, $M = 100$, $m = 1$ and $N = 500$, the eigenvalues read $\lambda_1 \approx -0.02$ and $\lambda_{2,3} \approx -0.50 \pm i 0.10$ leading to three characteristic timescales to compare with the total duration of the cycle: $\tau_1 = 4/|\lambda_1| \approx 170$, $\tau_2 = 4/|\Re(\lambda_{2,3})| \approx 8$ and $\tau_3 = 4/|\Im(\lambda_{2,3})| \approx 40$. Remarkably, the order of magnitude of the relevant timescales in the MD system is correctly reproduced by the eigenvalues of the equilibrium dynamic’s matrix (see Fig. 4.3 and Fig. 4.5). A detailed study of the 3V model is out of our present scope, but certainly deserves future investigation.

4.5 Analytic results in the two-variables (2V) model

In spite of its apparent simplicity, it is not easy to derive analytical results for the 3V model in a cycle of the external parameters. Here we show that the qualitative dependence of $\langle W \rangle$ on the total time of the cycle τ can be obtained in a simplified version of Eq. (4.11), where we set the temperature $T(t)$ to be equal to the temperature of the thermostat $T_o(t)$ at every time t : i.e. we go back to the two-variable model that was introduced in Chapter 3, with the only difference that

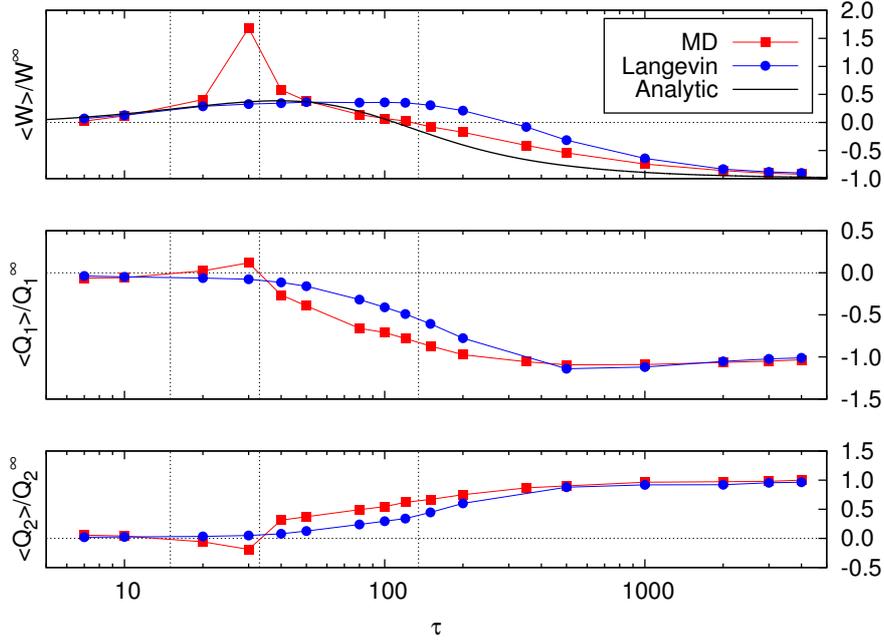


Figure 4.5. Average work and heats per cycle. Comparison between the MD, the 3-variables model, Eq. (4.11), and the analytical solution in Eq. (4.17) rescaled by their asymptotic values.

this model is uni-dimensional. Another necessary simplification with respect to the original model regards the engine protocol: the non-differentiable Ericsson cycle, following elementary kinetic theory replaced by a simpler protocol ($\omega = 2\pi/\tau$)

$$\begin{aligned} f(t) &= f_0(1 + \epsilon \cos(\omega t)), \\ T(t) &= T_0(1 + \delta \sin(\omega t)), \end{aligned} \quad (4.12)$$

where $f(t) = F(t)/M$, $\epsilon, \delta \ll 1$: we set $f_0 = 2$, $T_0 = 12$, $\epsilon = 0.1$ and $\delta = 0.08$. In the adiabatic limit, this simplified cycle (an approximation of the Ericsson protocol, see Fig. 4.2) produces a work not very different from the one of the Ericsson cycle. Passing to average values we obtain the equation (see Appendix 4.B):

$$\langle \ddot{X} \rangle + k(t)\langle X \rangle + \gamma(t)\langle \dot{X} \rangle = f(t). \quad (4.13)$$

The homogeneous solution associated to Eq. (4.13) goes to zero in the long time limit: therefore, since we are interested in the asymptotic stationary solution, we will focus only on the non-homogeneous solution. This will be done by expanding all the terms in Eq. (4.13) in powers of ϵ and δ . In particular, since $\langle X \rangle(t) = \langle X \rangle_0(t) + \langle X \rangle_1(t) + \mathcal{O}(\epsilon^2, \delta^2, \epsilon\delta)$, with $\langle X \rangle_1(t) = \mathcal{O}(\epsilon, \delta)$, by solving Eq. (4.13) for $\epsilon, \delta \rightarrow 0$ one gets $\langle X \rangle_0(t) = \frac{NT_0}{F_0}$, and

$$\langle \ddot{X} \rangle_1 + \omega_0^2 \langle X \rangle_1 + \nu \langle \dot{X} \rangle_1 = -\epsilon f_0 \cos \omega t + f_0 \delta \sin \omega t, \quad (4.14)$$

where $\omega_0^2 = F_0^2/(MNT_0)$ and $\nu = 2f_0\sqrt{2m}/(\pi T_0)$. The asymptotic solution is

$$\langle X \rangle_1(t) = A(\omega)[\delta \sin(\omega t + \phi(\omega)) - \epsilon \cos(\omega t + \phi(\omega))], \quad (4.15)$$

$$\begin{aligned}
A(\omega) &= \frac{f_0}{\sqrt{(\omega_0^2 - \omega^2)^2 + \nu^2 \omega^2}}, \\
\phi(\omega) &= \arctan\left(\frac{\nu\omega}{\omega_0^2 - \omega^2}\right).
\end{aligned} \tag{4.16}$$

The work performed over a cycle of the parameters of total time τ can be now expressed in a simple way

$$\begin{aligned}
W(\tau) &= -Mf_0 \frac{2\pi}{\tau} \epsilon \int_0^\tau dt (\langle X \rangle_0 + \epsilon \langle X \rangle_1(t)) \sin\left(2\pi \frac{t}{\tau}\right) = \\
&= Mf_0 \pi \epsilon A\left(\frac{2\pi}{\tau}\right) \left[\delta \cos \phi\left(\frac{2\pi}{\tau}\right) - \epsilon \sin \phi\left(\frac{2\pi}{\tau}\right) \right].
\end{aligned} \tag{4.17}$$

In Fig. 4.5 (black curve) it is seen that this result, when normalized to its adiabatic value, compares quite well, in spite of the many approximations introduced to obtain Eq. (4.17), with the average work performed by the MD system and the 3V model, recovering the change of sign at value not far from τ^* and a maximum at smaller values.

4.6 Linear regime

Many general results [99, 79, 100] regarding the performance of finite-time heat engines are obtained within the framework of irreversible thermodynamics [29]. For this reason and also for having a different insight into the physics of our model, in this Section we will fit our model into the Onsager formulation of out-of-equilibrium processes. Following the discussion of Sec. 2.3.1, we write down the the total average entropy production of the system

$$\Sigma(\tau) = - \int_0^\tau \frac{\langle \dot{Q}(t) \rangle}{T(t)} dt \geq 0, \tag{4.18}$$

where \dot{Q} is the rate of heat absorption from the thermostat. To obtain a decomposition analogous of the entropy production as the sum of the product of two fluxes and two thermodynamic driving forces, it is useful to express the temperature, following [8], as

$$T(t) = \frac{T_0(1 - \delta^2)}{(1 + \delta) - 2\delta\gamma(t)}, \tag{4.19}$$

where $T_0 = (T_H + T_C)/2$,

$$\delta = \frac{T_H - T_C}{T_H + T_C}. \tag{4.20}$$

and $\gamma(t)$ is a time-dependent function which is implicitly defined in Eq. (4.19). We will also use the notation $F_0 = (F_H + F_L)/2$ and

$$\epsilon = \frac{F_H - F_L}{F_H + F_L}, \tag{4.21}$$

to indicate, respectively, the intermediate force and the relative excursion. By plugging Eq. (4.19) into Eq. (4.18) and using $\langle W \rangle + \langle Q \rangle = 0$ one gets

$$\Sigma(\tau) = \epsilon J_1(\tau) + \delta J_2(\tau), \tag{4.22}$$

where

$$J_1(\tau) = \frac{\langle W \rangle}{\epsilon T_0(1 - \delta)}, \quad (4.23)$$

$$J_2(\tau) = \frac{2}{T_0(1 - \delta^2)} \int_0^\tau \langle \dot{Q} \rangle \gamma(t). \quad (4.24)$$

Let us recall that for small values of ϵ and δ , i.e. in the linear regime, the fluxes are linear function of the forces

$$J_i(\tau) = L_{i1}(\tau)\epsilon + L_{i2}(\tau)\delta, \quad i = 1, 2, \quad (4.25)$$

where the Onsager coefficients L_{ij} non-trivially depend on the total time τ of the transformation.

We also recall that a reciprocity relation exists for the coefficient of the Onsager matrix: for each protocol determined by $T(t)$ and $F(t)$, it is possible to construct its “time-reversed” counterpart $\tilde{T}(t) = T(\tau - t)$ and $\tilde{F}(t) = F(\tau - t)$: if we indicate with $\tilde{\cdot}$ quantities measured in the time-reversed cycle, the following relation is a direct consequence of the reversibility of the microscopic dynamics:

$$L_{12}(\tau) = \tilde{L}_{21}(\tau). \quad (4.26)$$

For the Ericsson protocol described in Sec. 4.2.1 the time reversal transformation can be obtained by taking the same form of the protocol for T and F with an (inessential) global shift of phase $t_0 = \tau/2$ and inverting the sign of the force difference $\epsilon \rightarrow -\epsilon$. For this reason $\tilde{L}_{21} = -L_{21}$, i.e.

$$L_{12}(\tau) = -L_{21}(\tau). \quad (4.27)$$

In Fig. 4.6 the results of a measurement of J_1 and J_2 for different values of ϵ and δ in molecular dynamics simulations of the MM are reported. By fixing, respectively, $\delta = 0.05$ or $\epsilon = 0.05$, J_2 and J_1 are plotted as functions of ϵ and δ . A linear dependence is obtained for small values of the thermodynamic forces: moreover, the data are compatible with the hypothesis of two linear relations with opposite coefficients describing the functional dependence of J_1 on δ and of J_2 on ϵ (straight lines in Fig. 4.6). The measurement also confirms the fact that the Onsager coefficients have a non trivial dependence on the total time of the transformation τ (inset of Fig. 4.6). In Fig. 4.7 we report a measurement of J_1 as a function of ϵ for different values of δ : since the curves are parallel straight lines, the Onsager coefficient L_{11} does not depend on the value of δ (analogous results, not reported here, can be obtained for all the Onsager coefficients).

In Fig. 4.8 we study the limits of the linear behavior of the MM: by taking $\delta = \epsilon$, we report the average work divided by ϵ^2 as a function of τ for different values of ϵ . In the linear regime the different curves, when rescaled, must superimpose: this is the case, of course, for small values of ϵ . At larger values of ϵ the appearing discrepancies are not uniform in τ . In particular we remark that around the maximum of $\langle W \rangle$ the separation is much more prominent (also signaled by the appearance of a second local maximum for $\epsilon \geq 0.25$). Note that the non-linearity appears also in the large

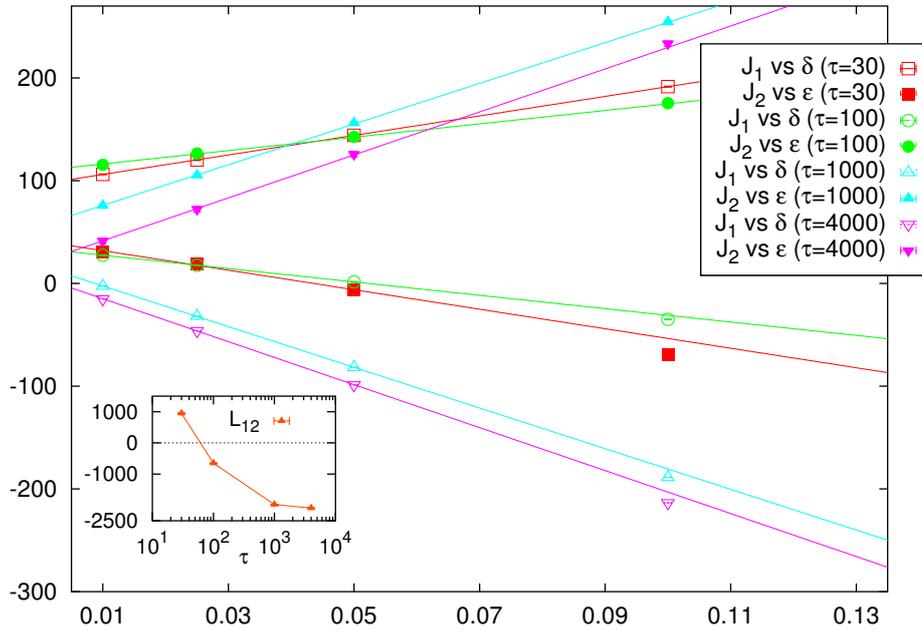


Figure 4.6. Thermodynamic currents J_1 (empty symbols) and J_2 (solid symbols) as functions, respectively, of thermodynamic forces δ and ϵ for different values of τ . Straight lines are obtained by fitting the two data sets at the same τ with two linear functions with slope L_{12} and $-L_{12}$ respectively: for this reason, lines with the same color have opposite slope. Inset: Off-diagonal Onsager coefficient L_{12} obtained with the above procedure as a function of τ . All the other parameters are the same of Fig. 4.3.

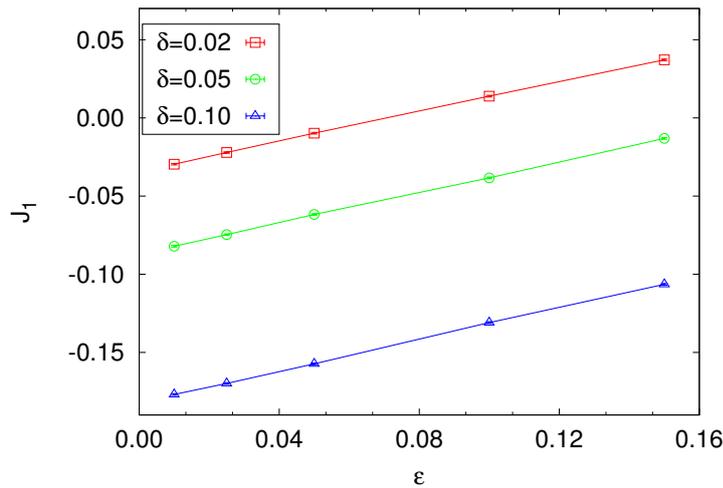


Figure 4.7. The thermodynamic current J_1 as a function of ϵ for different values of δ at $\tau = 500$. Other parameters are the same of Fig.4.3

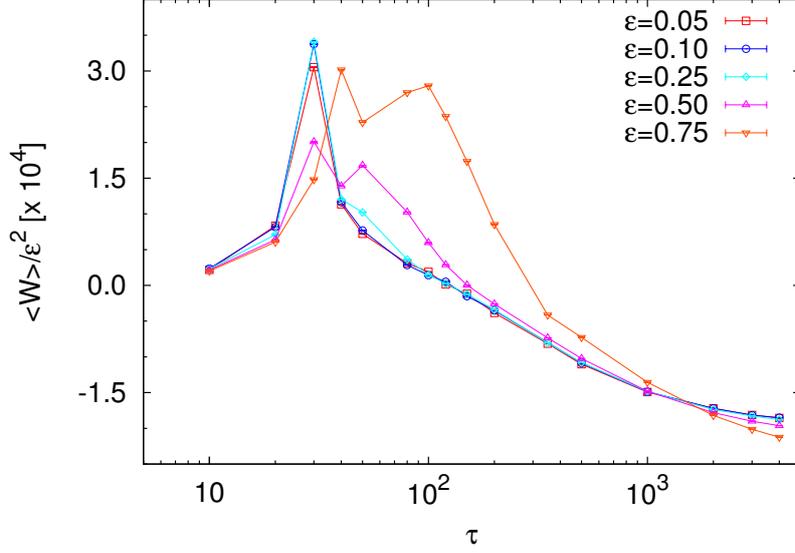


Figure 4.8. The average work per cycle, divided by ϵ^2 is reported as a function of τ for different values of ϵ . The other parameters are $\delta = \epsilon$, $F_0 = 200$, $T_0 = 12$, $N = 500$, $M = 100$, $m = 1$.

τ limit, since higher order terms of the expansion of the quantity $\log\left(\frac{1+\epsilon}{1-\epsilon}\right)$ in the adiabatic formula of work (see Table 4.1) become relevant. Unfortunately an analytic description of the non-linear regime is not yet available and the interesting features of such a regime will hopefully be the subject of future investigations.

4.6.1 Analytic expression of the Onsager coefficients in the 2V model

In the simplified 2V model, Eq. (4.13), it is possible to obtain an explicit expression for the above mentioned Onsager coefficients L_{ij} . By plugging Eq. (4.17) into definition (4.24) one immediately gets the linear expansion for J_1 , i.e.

$$J_1(\tau) = \frac{F_0}{T_0} \pi A \left(\frac{2\pi}{\tau} \right) \left(\epsilon \sin \phi \left(\frac{2\pi}{\tau} \right) - \delta \cos \phi \left(\frac{2\pi}{\tau} \right) \right). \quad (4.28)$$

In order to get the corresponding expansion for J_2 , we start by plugging the protocol $T(t) = T_0(1 + \delta \sin(2\pi t/\tau))$, Eq. (4.12), into Eq. (4.19) and get

$$\gamma(t) = \frac{1}{2} \left(1 + \sin \left(\frac{2\pi}{\tau} t \right) \right) + \mathcal{O}(\delta). \quad (4.29)$$

To obtain an explicit expression for $\langle \dot{Q} \rangle$ it is necessary to substitute the asymptotic solution for $\langle X \rangle$, Eq. (4.15) into the expression for energy $\langle E(t) \rangle = NT_o(t)/2 + M\langle \dot{X} \rangle^2/2 + F\langle X \rangle$, and then use the definition of heat $\langle \dot{Q} \rangle = \langle \dot{E} \rangle - \dot{F}\langle X \rangle = \mathcal{O}(\epsilon, \delta)$. Retaining only first order terms in ϵ and δ of Eq. (2.70) gives the following expression for J_2 :

$$J_2 = \frac{\pi F_0}{T_0} A \left(\frac{2\pi}{\tau} \right) \left(\epsilon \cos \phi \left(\frac{2\pi}{\tau} \right) + \delta \sin \phi \left(\frac{2\pi}{\tau} \right) \right). \quad (4.30)$$

In summary, the Onsager matrix takes the form

$$L = \frac{F_0}{T_0} \pi A \left(\frac{2\pi}{\tau} \right) \begin{pmatrix} \sin \phi \left(\frac{2\pi}{\tau} \right) & -\cos \phi \left(\frac{2\pi}{\tau} \right) \\ \cos \phi \left(\frac{2\pi}{\tau} \right) & \sin \phi \left(\frac{2\pi}{\tau} \right) \end{pmatrix} \quad (4.31)$$

The anti-reciprocal relation $L_{12} = -L_{21}$ in the Onsager matrix is due to the fact that the protocol used in the 2V model (Eq. (4.12)) behaves, under time-reversal, in the same way as the Ericsson protocol. We can also get a very simple expression for the total entropy production,

$$\Sigma(\tau) = \frac{\pi F_0}{T_0} A \left(\frac{2\pi}{\tau} \right) \sin \phi \left(\frac{2\pi}{\tau} \right) (\epsilon^2 + \delta^2), \quad (4.32)$$

which, as expected, is always positive because $A(\omega) \geq 0$ and $0 \leq \phi(\omega) \leq \pi$.

4.7 Efficiency at maximum power: linear regime and beyond

The aim of this Section is to study the efficiency at maximum power of our engine comparing three different levels of approximation: numerical simulations of the full Molecular Model, numerical solutions of its coarse-grained 2V version, and analytical solutions of the 2V model for small values of ϵ and δ (i.e. when fluxes are linear in the forces).

First, we need to find a suitable definition of efficiency η for our case. Usually the efficiency of a heat engine working at contact with two thermostats (at temperatures $T_C < T_H$) is simply the ratio of the output work divided by the energy absorbed from the hotter thermostat Q_{in} . For transformations that involve thermostats at temperatures ranging continuously in the interval $T_C < T(t) < T_H$, the input heat must be redefined [8] as

$$Q_{in} = \int_0^\tau \dot{Q}(t) \gamma(t) dt, \quad (4.33)$$

where τ is the total time of the transformation and $\gamma(t)$ is the function appearing at the denominator of the right-hand side of Eq. (2.64). This definition comes from the observation, already reported in the Chapter 2, that Eq. (4.33) gives the correct result in the simple case of two thermostats at T_C and T_H : moreover for the Ericsson protocol this expression reduces, in the quasi-static limit, to the heat extracted from the hot reservoir T_H .

From this definition, we get the expression for the efficiency of the engine,

$$\eta = \frac{-\langle W \rangle}{\langle Q_{in} \rangle} = -\frac{2\epsilon J_1}{(1 + \delta) J_2}, \quad (4.34)$$

where we used Eqs. (2.69) and (2.70) to recast the expression in terms of the thermodynamic currents J_1 and J_2 . When the total average entropy production Σ , Eq. (2.68), vanishes, it is straightforward to prove that the efficiency assumes the Carnot value $\eta_C = 1 - T_C/T_H = 2\delta/(1 + \delta)$. In the 2V model, the entropy Eq. (4.32), can vanish only for $\phi = 0$, i.e. in the adiabatic limit $\tau \rightarrow \infty$ (or in the trivial,

non-interesting case $\tau = 0$). It is quite reasonable to assume that, also in the general case, a vanishing entropy can be only obtained by varying very slowly the external parameters. As a consequence, the power corresponding to a maximally efficient engine must be zero. For this reason, in order to characterize the performance of the engine, we will study the efficiency at maximum power (EMP) $\tilde{\eta}$, i.e. the efficiency corresponding to a choice of the external parameters that maximizes the output power. In the last decades, a series of important results were obtained regarding the EMP: perhaps the most notable is that, under some rather general assumptions [28, 99], a universal bound for the EMP is given by the so-called Curzon-Ahlborn efficiency η_{CA} ,

$$\tilde{\eta} \leq \eta_{CA} = 1 - \sqrt{\frac{T_C}{T_H}} \approx \frac{1}{2}\eta_C + \mathcal{O}(\eta_C^2). \quad (4.35)$$

We now investigate the validity of such a bound, and compare our results with other recent works.

In our models the output power:

$$\mathcal{P} = -\frac{\langle W \rangle}{\tau} \quad (4.36)$$

depends on the details of the model (N, M, m) as well as on the choice of the external protocol $(T_0, F_0, \epsilon, \delta, \tau)$. Since the engine working state consists in a perturbation of the equilibrium state determined by T_0 and F_0 , it appears quite natural, in order to maximize \mathcal{P} , to fix N, m, M, T_0 and F_0 . Moreover, since η_C and η_{CA} only depend on the value of δ , we also fix the temperature difference and therefore maximize the two-variables function $\mathcal{P}(\epsilon, \tau)$.

4.7.1 Linear regime

In Fig. 4.9 the linear approximation for the output power in the 2V model (obtained by plugging Eq. (4.17) into Eq. (4.36)) is plotted as a function of ϵ and τ . In view of a comparison between the results in the 2V model and in the molecular model, whose protocols are slightly different, we rescaled the work and the power so that, for $\epsilon = 0.1$ the asymptotic value for work in the limit $\tau \rightarrow \infty$ is fixed to $\lim_{\tau \rightarrow \infty} W_{rescaled}(\epsilon = 0.1, \tau) = 100$. The plot shows that (in this linear approximation) the power does not reach a global maximum at a unique value of (τ, ϵ) . Indeed, there exists a curve $\tau_{mp}(\epsilon)$ consisting of τ -maxima points, i.e. where $\partial_\tau \mathcal{P} = 0$. The maximum power curve $\mathcal{P}[\epsilon, \tau_{mp}(\epsilon)]$ saturates to a constant value for increasing ϵ . In addition we also note that $\tau_{mp}(\epsilon)$ is an increasing function of ϵ , eventually saturating at the value $\lim_{\epsilon \rightarrow 1} \tau_{mp} \approx 1500$. In the plot we have also shown the curves at constant η : it is interesting to notice that the $\tau_{mp}(\epsilon)$ approaches the curve where $\eta = \eta_{CA}$ at increasing ϵ .

The EMP curves, $\tilde{\eta}(\epsilon, \tau_{mp}(\epsilon))$ for different values of δ are shown in Fig. 4.10: we observe that - consistently with the previous observation - the CA efficiency is only reached for large values of ϵ where, in principle, the linear approximation is no more reliable. However, by decreasing δ a faster convergence toward the CA efficiency is observed: this suggests the possibility to observe $\tilde{\eta} = \eta_{CA}$ even in the linear regime.

Let us remark again that in our system it is possible to separate the time τ of the transformation from the small force limit (small ϵ and δ): this means that we are able

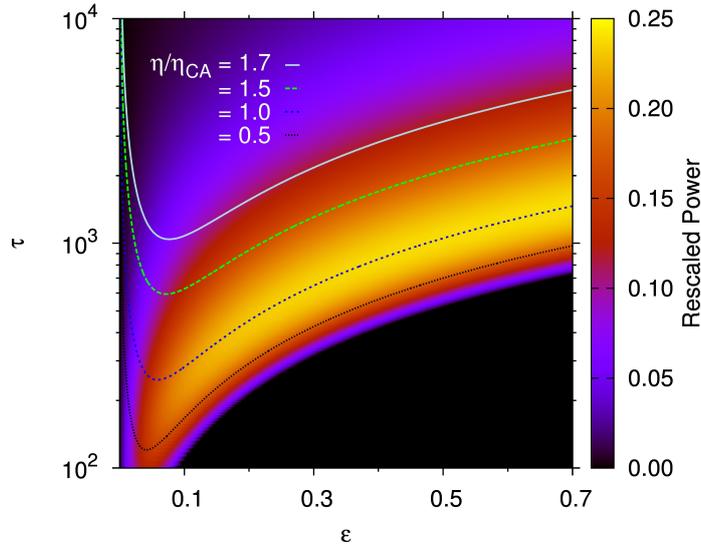


Figure 4.9. *2V model, linear approximation:* contour plot of rescaled $\mathcal{P}(\epsilon, \tau)$ as a function of τ and ϵ . The power is rescaled by the factor $w_{ad}^{2V} = \epsilon \delta N T_0 \pi / 100$ evaluated in $\epsilon = 0.1$, $\delta = 1/12$, $N = 500$, $T_0 = 12$, $w_{ad}^{2V} \approx 1.57$. Continuous line represents the set of points (ϵ, τ) where η/η_{CA} is constant.

to consider a linear approximation (and construct the corresponding τ -dependent Onsager matrix) which is valid, in the small $\epsilon - \delta$ limit, at every value of τ . On the contrary, in many recent papers (see e.g. [55]) one of the small thermodynamic forces must be the inverse of the time of the transformation τ .

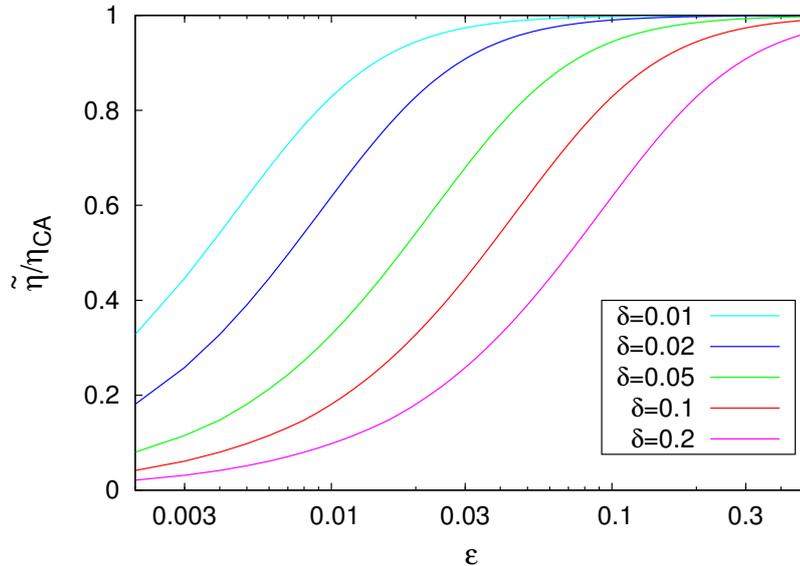


Figure 4.10. *2V model, linear approximation:* ratio of efficiency at maximum (with respect to τ) power $\tilde{\eta}$ and Curzon-Ahlborn efficiency η_{CA} , as a function of ϵ for different values of δ . Other parameters are $N = 500$, $M = 100$, $m = 1$, $T_0 = 12$, $F_0 = 200$.

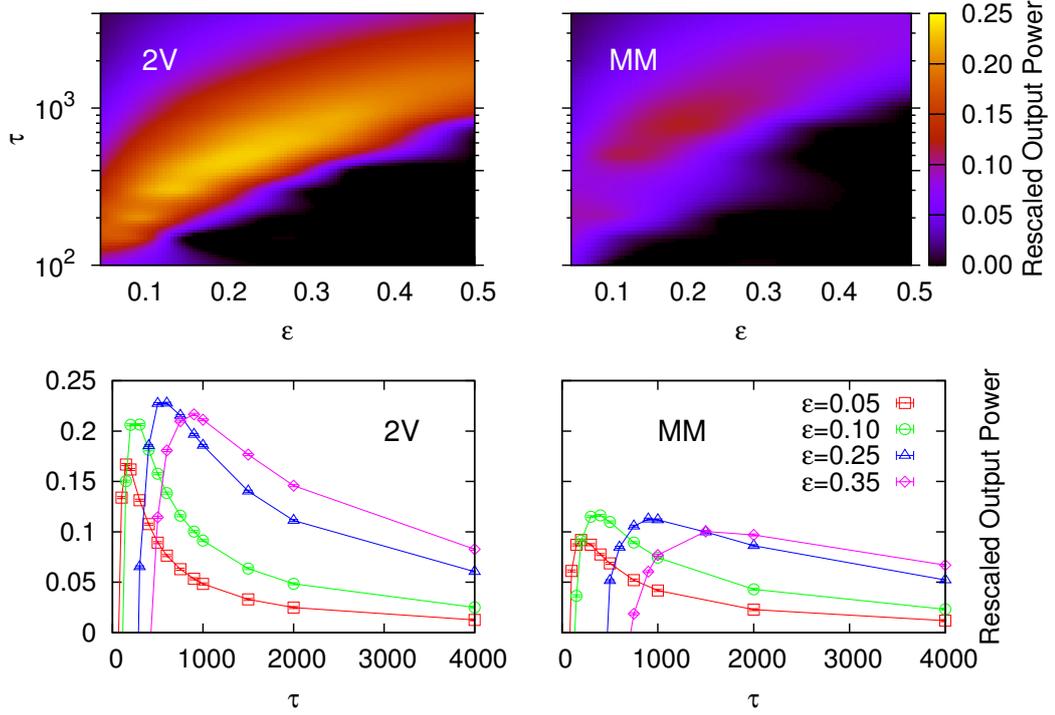


Figure 4.11. *Left: 2V model solved numerically, right: simulations of the MM.* Top panels: colour plot of the rescaled output power. Bottom panels: the rescaled output power as a function of τ for different values of ϵ . Parameters are the same in 2V and MM simulations: $\delta = 1/12$, $N = 500$, $M = 100$, $m = 1$, $F_0 = 200$, $T_0 = 12$. The power in MM simulation is divided by the quantity $w_{ad}^{MM} = 0.02 N T_0 \delta \ln((1 + \epsilon)/(1 - \epsilon))$ in $\epsilon = 0.1$, $w_{ad}^{MM} \approx 2.007$. In the 2V model the rescaling factor is $w_{ad}^{2V} = \epsilon \delta N T_0 \pi / 100$ which, in $\epsilon = 0.1$ gives $w_{ad}^{2V} \approx 1.57$

4.7.2 Non-linear regime

The absence of an absolute maximum for the power, however, appears only to be a consequence of the linear approximation used to solve the 2V equations. Indeed, by performing numerical integration of the full 2V model (Eq. (4.13)) and simulations of the MM, we observe a rather different situation, which is reported in Fig. 4.11: the two top panels represent the color map of the functions $\mathcal{P}(\epsilon, \tau)$ for the two models, the two bottom panels show some sections $\mathcal{P}(\epsilon^*, \tau)$ vs τ , for some values of ϵ^* . By analyzing these last plots, we observe that the maximum power increases when going from $\epsilon = 0.1$ to $\epsilon = 0.25$ and then decreases again in $\epsilon = 0.35$. This suggests that is indeed possible, at least numerically, to find a specific value for ϵ and τ corresponding to the global maximum power. The only significant difference between the 2V model and the MM is that the output power is smaller, in general, than the one obtained in the 2V model.

In Fig. 4.12 we focus on the MM and report the same output power as a function of the efficiency (which, interestingly, is a bijective function of τ). We observe that, at every value of ϵ , the maximum power is attained at a value $\eta < \eta_{CA}$: moreover the global maximum power corresponds to an efficiency that is approximately the 70%

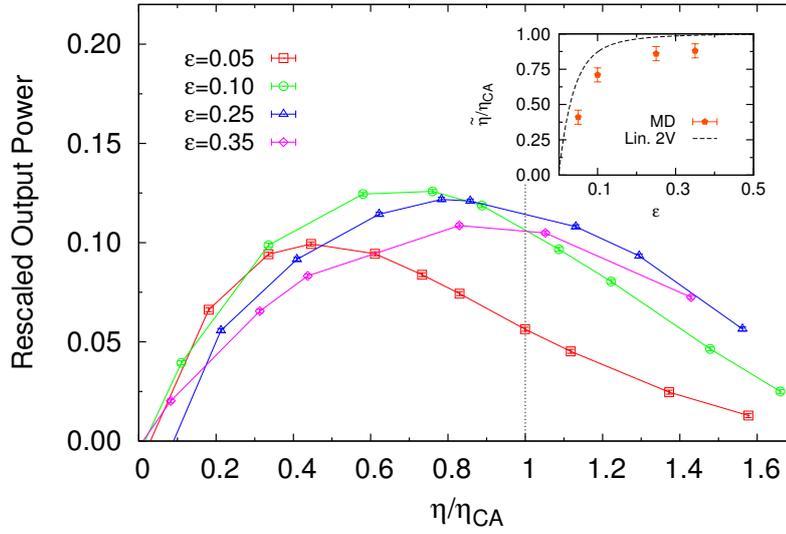


Figure 4.12. *Simulations of the MM:* the rescaled output power is reported as a function of the rescaled efficiency η/η_{CA} for different values of ϵ . The rescaling factor for power is $w_{ad}^{MM} \approx 2.007$ and $\eta_{CA} = 1 - \sqrt{(1 + \delta)/(1 - \delta)}$. Inset: comparison between numerically solved 2V model and simulations of the MM for the relative EMP $\tilde{\eta}/\eta_{CA}$ is reported as a function of ϵ . Other parameters are the same as in Fig. 4.11.

of the Curzon-Ahlborn efficiency (i.e. 35% of the Carnot efficiency). The Curzon-Ahlborn efficiency seems to be approached for larger values of ϵ . A comparison between the EMP measured in the numerical simulations of the MM and the corresponding result in the 2V model (inset of Fig. 4.12), shows that the simplified model overestimates the actual value of $\tilde{\eta}$.

We wish to spend a few words about the observed lower value of the EMP with respect to the CA efficiency. We see that it is a consequence of the widening of the space of parameters. In Ref. [99] (and its generalization to non-symmetric Onsager matrices in Ref. [8]), it is proved that, for fixed Onsager coefficients (i.e. for fixed τ) η_{CA} is reached whenever the value

$$q = \frac{L_{12}L_{21}}{L_{11}L_{22} - L_{12}L_{21}} \quad (4.37)$$

is close to $q = 1$ (“tight coupling” hypothesis). When the Onsager matrix is τ -dependent, the variable q is a function of τ , $q = q(\tau)$. Suppose the existence of a value of $\tau = \tau_0$ such that the tight-coupling hypothesis is verified $q(\tau_0) \simeq 1$: then, by denoting with ϵ_0 the value of ϵ that maximizes the power at $\tau = \tau_0$ (with respect to ϵ), we will obtain $\eta(\tau_0, \epsilon_0) = \eta_{CA}$. On the other hand, the global maximum power in the (τ, ϵ) -plane may occur in a point (τ_1, ϵ_1) for which the tight coupling condition is violated $q(\tau_1) < 1$, corresponding to an efficiency $\eta(\tau_1, \epsilon_1) < \eta_{CA}$. To summarize, extending the space of parameters, e.g. by allowing τ to vary, may permit to find a larger maximum power, but this does not guarantee that the corresponding efficiency would be closer to the CA efficiency.

4.8 An analytic study of fluctuations by means of Large Deviation Theory

In this last Section of the Chapter we will give an analytic description of the fluctuations of the work W measured in a cycle of the external parameters. We will resort to an even simpler model with respect to the 2V model introduced in the previous sections. For the sake of consistency, we recall the (time-dependent) stochastic differential equation describing the model:

$$\begin{aligned}\dot{X} &= v \\ \dot{V} &= -k(t)X - \gamma(t)V + \frac{F(t)}{M} + \sqrt{\frac{2\gamma(t)T(t)}{M}}\eta,\end{aligned}\quad (4.38)$$

where η represents white noise $\langle \eta(t)\eta(t') \rangle = \delta(t-t')$. The coefficients k and γ are functions of two external control parameters $F(t)$ (the “force”) and $T(t)$ (the “temperature”), that vary in time according to the periodical protocol described in Sec. 4.5.

The quantity we are interested in is the work absorbed by the system during a cycle,

$$W[\{X(s)\}_{s=0}^{\tau}, \tau] = \int_0^{\tau} ds \dot{V}(s)X(s), \quad (4.39)$$

We recall that the asymptotic expansion at first order in ϵ and δ of the solution of Eq. (4.38) reads

$$\langle X \rangle(t) = \frac{NT_0}{F_0} + A(\omega)[\epsilon \cos(\omega t + \phi(\omega)) - \delta \sin(\omega t + \phi(\omega))], \quad (4.40)$$

where, if we denote with $\omega_0^2 = k(t)|_{\epsilon, \delta=0}$, $\nu = \gamma(t)|_{\epsilon, \delta=0}$ and $f_0 = F_0/M$, we have

$$\begin{aligned}A(\omega) &= \frac{-f_0}{\sqrt{(\omega_0^2 - \omega^2)^2 + \nu^2\omega^2}}, \\ \phi(\omega) &= \arctan\left(\frac{\nu\omega}{\omega_0^2 - \omega^2}\right).\end{aligned}\quad (4.41)$$

Thus, the average work reads

$$W(\tau) = -Mf_0\pi\epsilon A\left(\frac{2\pi}{\tau}\right) \left[\epsilon \sin \phi\left(\frac{2\pi}{\tau}\right) - \delta \cos \phi\left(\frac{2\pi}{\tau}\right) \right]. \quad (4.42)$$

As shown in Figure 4.13, this analytic curve compares very well with the values of $\langle W \rangle$ obtained by numerical integration ($\epsilon = 0.1$ and $\delta = 0.08$).

In order to go behind the average value and to get an estimate of the fluctuations of W we will resort to a even simpler model. Heuristically we can assume that if, for every t , $\gamma(t) \ll 1$ and $k(t) \ll 1$, but $k(t) \sim \gamma(t)$, then the term \dot{v} in the second equation in Eq. (4.38) can be neglected, and obtain

$$\dot{X} + \alpha(t)X = \frac{F(t)}{\gamma(t)M} + \sqrt{\frac{2T(t)}{M\gamma(t)}}\eta \quad (4.43)$$

with $\alpha(t) = k(t)/\gamma(t)$. In order to investigate the validity of such approximation let us compare the (first order asymptotic expansion) of the long-time solution of Eq. (4.43) and Eq. (4.38). It is quite easy to show that, at first order in ϵ and δ , the solution of Eq. (4.43) has the same form of Eq. (4.40), with

$$\begin{aligned} A(\omega) &= \frac{-f_0}{\sqrt{\omega_0^4 + \nu^2 \omega^2}}, \\ \phi(\omega) &= \arctan\left(\frac{\nu\omega}{\omega_0^2}\right), \end{aligned} \quad (4.44)$$

i.e. it can be obtained from the previous expression by taking the limit $\omega \ll \omega_0$, or, equivalently, $\tau \gg 1/\omega_0$ (see Fig. 4.13, where we compare the average work). In the overdamped limit, the state of the system can be now represented by the only variable X .

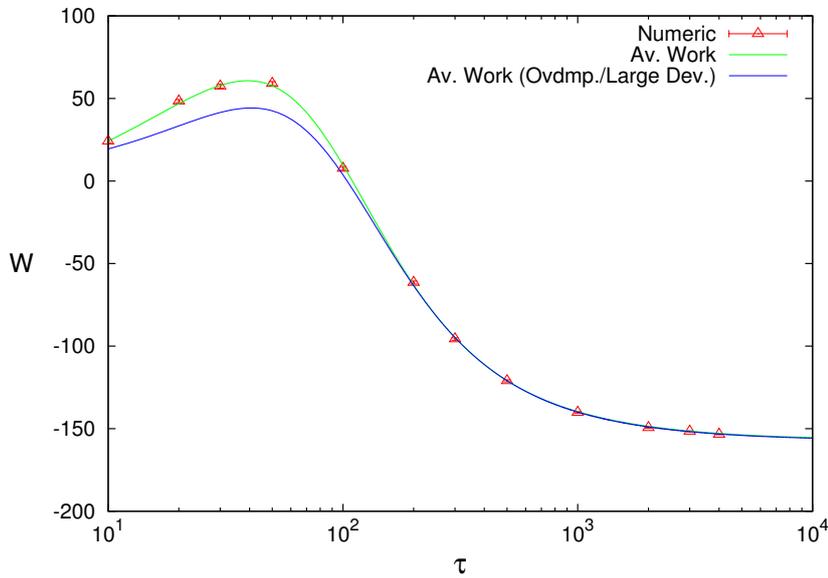


Figure 4.13. Comparison of the average work $\langle W \rangle$ over a cycle obtained from numerical integration of the 2V equation (triangles), the analytic formula in Eq. (4.42) (green line) and the overdamped approximation Eq. (4.43)

For this equation the explicit form of the probability distribution of work $p(W)$ can be obtained through the large deviation formalism. We start by recalling that the probability of observing a trajectory $\{X(s)\}_{s=0}^\tau$ is [52]

$$\mathcal{P}[\{X(s)\}_{s=0}^\tau] \propto \exp\left\{-\int_0^\tau ds \mathcal{L}(x(s), \dot{x}(s), s)\right\} \quad (4.45)$$

with

$$\mathcal{L}(X, \dot{X}, t) = \frac{M\gamma(t)}{4T(t)} \left(\dot{X} + \alpha(t)X - \frac{F(t)}{M\gamma(t)} \right)^2. \quad (4.46)$$

Let us note that the system's main peculiarity, i.e. to have a time-dependent noise amplitude and external driving, does not affect the derivation of Eqs. (4.45) and

(4.46). If the coefficient that multiplies the lagrangian,

$$\Gamma = \sqrt{\frac{mF_0^2}{T_0^3}}, \quad (4.47)$$

is large, we can use a large deviation approach and therefore derive the probability distribution of W . In particular, the large deviation function of W ,

$$I(w) = \lim_{\Gamma \rightarrow \infty} -\frac{1}{\Gamma} \log p(w) \quad (4.48)$$

can be obtained by using the *contraction principle*, i.e.

$$I(w) = \sup_{\{X\}_\tau: W[\{X\}] = w} \int_0^\tau ds \mathcal{L}(X(s), \dot{X}(s)). \quad (4.49)$$

In other words, asymptotically in Γ , the probability of observing a given value of work w corresponds to the probability of observing a specific path, i.e. the most probable among all the paths corresponding to the given value of w .

To evaluate $I[w]$ we must study a maximization problem with constraints: to do this we use the method of Lagrange multipliers, i.e. we add to the action a term $-\beta W[\{X\}, \tau]$, or, equivalently, we study the problem with a modified Lagrangian,

$$\mathcal{L}'(\dot{X}, X, \beta, t) = \mathcal{L}(\dot{X}, X, t) - M\beta\dot{F}(t)X, \quad (4.50)$$

when the β must be derived *a fortiori*, to impose the constraint on the desired value of the work.

To find the maximum of the action we use Euler-Lagrange equations,

$$\frac{d}{dt} \frac{\partial \mathcal{L}'}{\partial \dot{X}} - \frac{\partial \mathcal{L}'}{\partial X} = 0 \quad (4.51)$$

that, using Eq (4.50), consist in resolving the following differential equation

$$q(t)\ddot{X} + q'(t)\dot{X} + g(t)X = f(t), \quad (4.52)$$

where $q(t) = \gamma(t)/T(t)$, $g(t) = X(\alpha(t)q'(t) + \alpha'(t)q(t) - \alpha^2(t)q(t))$ and $f(t) = N(\gamma(t)\alpha(t)/F(t))' + N\alpha^2\gamma/F - \beta F'(t)$. To solve the Dirichlet problem associated with this (second order) differential equation, we must impose the boundary conditions $X(t=0) = X_0$ and $X(t=\tau) = X_1$. Again the solution of this equation can be found by expanding $X(t)$ in powers of ϵ and δ : however, in principle, since τ is a finite quantity, we are no longer authorized to focus only on the particular solution of the equation, and the solutions of the associated homogeneous equation should be used to impose the boundary conditions. However, such homogenous solutions, are exponentials with a characteristic time $\tau_o \sim mN^2T_0/F_0^2$: when $\tau \gg \tau_o$, we can safely assume that the contribution of the homogenous solution to the total work can be neglected.

The first order expansion of the particular solution $X_\beta(t)$ is, thus, a linear combination of $\cos(\omega t)$ and $\sin(\omega t)$ (its explicit expression is quite involved, so we do not report it here), in which the coefficient multiplying the sinusoidal term depends

linearly on β . This solution must be now plugged into Eq. (4.39), in order to obtain an expression for β as a function of w : it is quite easy to verify that this expression is again linear, i.e. $\beta = a_1 w + a_2$.

Finally, we must plug $X_{\beta(w)}(t, \tau)$ into Eq. (4.50) to obtain, thanks to the contraction principle, an explicit formula for $I(w)$:

$$I(w) = \frac{(w - w_0(\tau))^2}{2\sigma_w(\tau)^2}, \quad (4.53)$$

where the explicit expression of w_0 and σ_w^2 , again, is too involved to be reported here: the most important conclusion to draw is that, within our assumptions, $p(W)$ is gaussian. Moreover, we see from Fig. 4.13 that, at least at a numerical level for

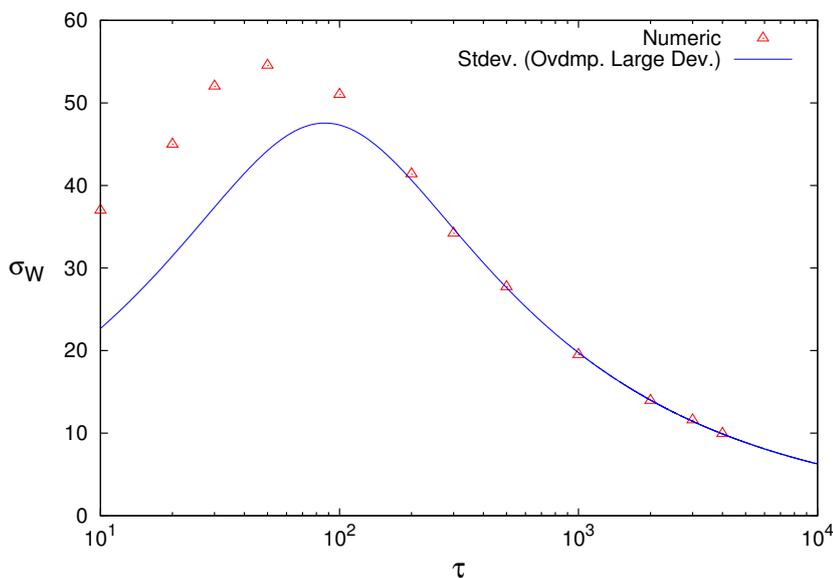


Figure 4.14. Comparison of the standard deviation of the work W obtained within the large deviation formalism (blue line) with the values obtained with a numerical integration of the 2V (underdamped) model.

our choice of parameters, w_0 coincides with the average value of W obtained without resorting to the large deviation formalism and that, in addition, it agrees with the underdamped expression at large τ and, interestingly for us, in the engine regime described in the previous sections. In addition, from Fig.4.14, we observe that for large values of τ , our analytic expression for $\sigma_w(\tau)$ coincides with the measurements of the same quantity in numerical simulations of the underdamped equation. It is clear, therefore, that, when comparing these results with the molecular dynamics results, see Fig. 4.4, one will observe the already mentioned discrepancies.

4.9 Conclusions

We have studied the thermodynamic properties of a model engine. The essential distinguishing features of our system are: 1) a realistic gas-like dynamics occurring in a spatially extended domain (i.e. the space between the moving piston and the

thermostat); 2) inertial effects which allow for a larger freedom in the choice of parameters (e.g. τ is not constrained by the piston's velocity) and a more rich phase diagram; 3) a cyclical protocol repeating in a finite time τ which is not related to the relative excursions of the pressure and temperature, δ and ϵ . The results of the simulations of the molecular model are compared to analytical and numerical solutions of a simplified, coarse-grained, equation, which yields a qualitatively similar picture. A clear scenario emerges from our study, where the relation between the fluxes (heat and work) and the thermodynamic forces do not depend trivially upon τ , as it appears, for instance, in the approximated expressions of the Onsager matrix, Eq. (4.31). Our model is appropriate to study the issues of finite-time thermodynamics in a case where the adiabatic limit ($\tau \rightarrow \infty$) and the linear regime (small thermodynamic forces) are disentangled. It would be interesting to check whether higher order terms in the expansion (in δ, ϵ) of the 2V model is able to reproduce the presence (observed both in the non-linear 2V model and the MM) of a global maximum of the power. In view of the finite number N of particles in the engine, we also studied the fluctuations of work from an analytical point of view: this can lead, in the future, to the definition and the analysis of a fluctuating efficiency [100].

4.A Details on the derivation of the Langevin Equation

In order to get a linear Langevin equation from kinetic theory we start from the conditional equilibrium distribution with fixed values of the macroscopic variables X, V, T , and then determine the average number of particles that, in the unit time Δt , collide with the piston or with the thermal wall. Using Eq. (4.2) and (4.3), one can determine post-collisional velocities and, accordingly, the average change of V and T , over a time Δt . In the following, to simplify the notation, we denote with $\langle \cdot \rangle$ the conditional average $\langle \cdot | X, V, T \rangle$.

The equation for X reads $\dot{X} = V$. The total average force acting on the piston due to collisions is:

$$\begin{aligned}
 & \lim_{\Delta t \rightarrow 0} \frac{\langle \Delta V \rangle_{coll}}{\Delta t} = \\
 &= \frac{N}{X} \int_0^\infty dv \frac{2m}{m+M} (v-V) |V-v| \rho_T(v) \Theta(V-v) = \\
 &= \frac{2Nm}{(M+m)X} \sqrt{\frac{m}{2\pi T}} \int_V^\infty dv (v-V)^2 e^{-\frac{mv^2}{2T}} = \\
 &= \frac{N}{(m+M)X} \left((T+mV^2) \operatorname{erfc} \left(\sqrt{\frac{m}{2T}} V \right) + \right. \\
 & \quad \left. - \sqrt{\frac{2mTV^2}{\pi}} e^{-\frac{mV^2}{2T}} \right), \tag{4.54}
 \end{aligned}$$

where $\operatorname{erfc}(x) = 2/\sqrt{\pi} \int_x^\infty dt \exp(-t^2)$. To obtain the total force, a term $-F/M$ must be added. The elastic collisions with the piston also affect the kinetic energy of the gas, through the term

$$\begin{aligned}
\langle \dot{T} \rangle \Big|_{coll} &= \frac{m}{X} \int dv (v'^2 - v^2) |v - V| \Theta(v - V) \rho_T(v) = \\
&= \frac{m}{X} \sqrt{\frac{m}{2\pi T}} \int_V^\infty dv \left[\left(v + 2 \frac{M}{m+M} (V - v) \right)^2 - v^2 \right] (v - V) e^{-m \frac{v^2}{2T}} = \\
&= -\frac{2M}{(m+M)^2 X} \left(\sqrt{\frac{2mT}{\pi}} (2T - MV^2) e^{-\frac{mV^2}{2T}} + \right. \tag{4.55}
\end{aligned}$$

$$\left. + V (MT + mMV^2 - 2mT) \operatorname{erfc} \left(\sqrt{\frac{m}{2T}} V \right) \right). \tag{4.56}$$

Finally, the average change of temperature in a time interval Δt , due to the collision with the thermal wall is simply given by the term

$$\begin{aligned}
\langle \dot{T} \rangle \Big|_{ther} &= \frac{m}{X} \int dv dv' (v'^2 - v^2) |v| \Theta(-v) \rho_T(v) \rho_o(v') = \\
&= -\frac{m^2}{XT_o} \sqrt{\frac{m}{2\pi T}} \int_{-\infty}^0 dv \int_0^\infty dv' (v'^2 - v^2) vv' e^{-m \frac{v^2}{2T}} e^{-m \frac{v'^2}{2T_o}} = \\
&= \sqrt{\frac{2}{\pi m}} \frac{\sqrt{T}(T_o - T)}{X}. \tag{4.57}
\end{aligned}$$

The equilibrium value of X, V and T for which $\langle \dot{X} \rangle = 0$, $\langle \dot{V} \rangle = 0$ and $\dot{T} = \langle \dot{T} \rangle \Big|_{coll} + \langle \dot{T} \rangle \Big|_{ther} = 0$ are

$$X_{eq} = N \frac{T_o}{F}, \tag{4.58}$$

$$V_{eq} = 0, \tag{4.59}$$

$$T_{eq} = T_o, \tag{4.60}$$

where terms $\mathcal{O}(m/M)$ are neglected. We can obtain a linear equation by expanding the expressions above up to first order around the equilibrium values: this can be done only if fluctuations are small, i.e. when $N \gg 1$ and $M \gg m$. The sum of Eq. (4.54) and $-F/M$ yields

$$\langle \dot{V} \rangle = -k(X - X_{eq}) - \gamma V + \mu(T - T_{eq}), \tag{4.61}$$

with $k = F^2/MNT_o$, $\gamma = 2F\sqrt{2m/(M^2\pi T_o)}$ and $\mu = F/(MT_o)$. Similarly the sum of Eq. (4.55) and Eq. (4.57) yields

$$\langle \dot{T} \rangle = -\frac{2MT_o}{N} \mu V - \alpha(T - T_{eq}), \tag{4.62}$$

with $\alpha = F\sqrt{2/(mN^2\pi T_o)}$. The coefficients k, γ, μ and α vary in time according to the time evolution of F and T_o . In order to take into account the fluctuations of this variables one must add three independent gaussian terms $\xi_X \equiv \xi_0$, $\xi_V \equiv \xi_1$ and $\xi_T \equiv \xi_2$, with an appropriate weight matrix b_{ij} with $i, j = 0, 1, 2$. In this particular

case the matrix is diagonal, $b_{ii} = b_i$, with $b_0 = 0$, $b_1 = \sqrt{2\gamma\sigma_V^2}$ and $b_2 = \sqrt{2\gamma\sigma_T^2}$. The final form of the linear Langevin equation thus reads

$$\begin{aligned}\dot{X} &= V, \\ \dot{V} &= -k(X - X_{eq}) - \gamma V + \mu(T - T_{eq}) + \sqrt{\frac{2\gamma T_o}{M}}\xi_1, \\ \dot{T} &= -\frac{2MT_o}{N}\mu V - \alpha(T - T_{eq}) + \sqrt{\frac{4\alpha T_o^2}{N}}\xi_2.\end{aligned}\quad (4.63)$$

Let us note that this equation, with fixed F and T_o , satisfies detailed balance [43].

4.B Details on the analytic solution of the 2V model

Let us consider Eq. (4.13), and, to shorten the notation, let $Y = \langle X \rangle$:

$$\ddot{Y} + k(t)Y + \gamma(t)\dot{Y} = f(t), \quad (4.64)$$

where

$$\begin{aligned}\gamma(t) &= \frac{2F(t)}{M} \sqrt{\frac{2m}{\pi T_o(t)}}, \\ k(t) &= \frac{F(t)^2}{MNT_o(t)},\end{aligned}\quad (4.65)$$

and $f(t) = F(t)/M$. In the Ericsson cycle F and T depend on time in a too complicated manner in order to perform analytic calculations. Therefore, in order to obtain an explicit result, in the following we will assume

$$\begin{aligned}f(t) &= f_0(1 + \epsilon \cos(\omega t)), \\ T_o(t) &= T_0(1 + q\epsilon \sin(\omega t)),\end{aligned}\quad (4.66)$$

where $\omega = 2\pi/\tau$, $\epsilon \ll 1$ and $q \sim \mathcal{O}(1)$. We will now sketch the derivation of the non-homogeneous solution of Eq. (4.64) as an asymptotic expansion in $\epsilon \ll 1$. Let us expand in power of ϵ all the coefficients appearing in Eq. (4.64) up to $\mathcal{O}(\epsilon)$

$$k(t) \approx \frac{Mf_0^2}{NT_0} [1 + (2\cos\omega t - q\sin\omega t)\epsilon] \quad (4.67)$$

$$\begin{aligned}\gamma(t) &\approx \sqrt{\frac{4f_0^2 m}{\pi T_0}} \left[1 + \left(\cos\omega t - \frac{q}{2} \sin\omega t \right) \epsilon \right] \\ Y(t) &\approx Y_0(t) + \epsilon Y_1(t).\end{aligned}\quad (4.68)$$

Plugging these expressions into Eq. (4.64), for $\epsilon = 0$ one gets

$$\ddot{Y}_0 + \frac{Mf_0^2}{NT_0} Y_0 + 2f_0 \sqrt{\frac{2m}{\pi T_0}} \dot{Y}_0 = f_0, \quad (4.69)$$

leading to

$$Y_0(t) = \frac{NT_0}{F_0}. \quad (4.70)$$

At the following order $\mathcal{O}(\epsilon)$, Eq. (4.64) gives

$$\begin{aligned} \ddot{Y}_1 + \frac{Mf_0^2}{NT_0}(Y_1 + (2\cos\omega t - q\sin\omega t)Y_0) + \\ 2f_0\sqrt{\frac{2m}{\pi T_0}}\left(\dot{Y}_1 + \left(\cos\omega t - \frac{q}{2}\sin\omega t\right)\dot{Y}_0\right) = \\ = f_0\cos\omega t \end{aligned} \quad (4.71)$$

or, if we plug the value of Y_0 ,

$$\ddot{Y}_1 + \omega_0^2 Y_1 + \nu \dot{Y}_1 = -f_0\cos\omega t + f_0q\sin\omega t, \quad (4.72)$$

where $\omega_0 = \sqrt{Mf_0^2/(NT_0)}$ and $\nu = 2f_0\sqrt{2m/(\pi T_0)}$. A solution of this equation can be found in the form

$$Y_1(t) = A(\omega)(\cos(\omega t + \phi(\omega)) + q\sin(\omega t + \phi(\omega))), \quad (4.73)$$

where

$$A(\omega) = \frac{-f_0}{\sqrt{(\omega_0^2 - \omega^2)^2 + \nu^2\omega^2}}, \quad (4.74)$$

$$\phi(\omega) = \arctan\left(\frac{\nu\omega}{\omega_0^2 - \omega^2}\right). \quad (4.75)$$

Bibliography of the Author

- [Cerino et al., 2014] Cerino, L., Gradenigo, G., Sarracino, A., Villamaina, D., and Vulpiani, A. (2014). Fluctuations in partitioning systems with few degrees of freedom. *Physical Review E*, 89(4):042105.
- [Cerino and Puglisi, 2015] Cerino, L. and Puglisi, A. (2015). Entropy production for velocity-dependent macroscopic forces: The problem of dissipation without fluctuations. *EPL (Europhysics Letters)*, 111(4):40012.
- [Cerino et al., 2015a] Cerino, L., Puglisi, A., and Vulpiani, A. (2015a). A consistent description of fluctuations requires negative temperatures. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(12):P12002.
- [Cerino et al., 2015b] Cerino, L., Puglisi, A., and Vulpiani, A. (2015b). Kinetic model for the finite-time thermodynamics of small heat engines. *Physical Review E*, 91(3):032128.
- [Cerino et al., 2016a] Cerino, L., Cecconi, F., Cencini, M., and Vulpiani, A. (2016a). The role of the number of degrees of freedom and chaos in macroscopic irreversibility. *Physica A: Statistical Mechanics and its Applications*, 442:486–497.
- [Cerino et al., 2016b] Cerino, L., Puglisi, A., and Vulpiani, A. (2016b). Linear and nonlinear thermodynamics of a kinetic heat engine with fast transformations. *Physical Review E*, 93(4):042116.

Bibliography

- [1] ANDRESEN, B., BERRY, R., NITZAN, A., AND SALAMON, P. Thermodynamics in finite time. I. The step-Carnot cycle. *Physical Review A*, **15** (1977), 2086.
- [2] ANDRESEN, B., BERRY, R. S., NITZAN, A., AND SALAMON, P. Thermodynamics in finite time. i. the step-carnot cycle. *Physical Review A*, **15** (1977), 2086.
- [3] BARNUM, H., ET AL. Is Boltzmann entropy time's arrow's archer? *Physics Today*, **47** (2008), 11.
- [4] BERDICHEVSKY, V., KUNIN, I., AND HUSSAIN, F. Negative temperature of vortex motion. *Phys. Rev. A*, **43** (1991), 2050.
- [5] BERDICHEVSKY, V., KUNIN, I., AND HUSSAIN, F. Reply to "Comment on 'Negative temperature of vortex motion'". *Phys. Rev. E*, **47** (1993), 2968.
- [6] BLICKLE, V. AND BECHINGER, C. Realization of a micrometre-sized stochastic heat engine. *Nature Physics*, **8** (2012), 143.
- [7] BOFFETTA, G., DEL CASTILLO-NEGRETE, D., LÓPEZ, C., PUCACCO, G., AND VULPIANI, A. Diffusive transport and self-consistent dynamics in coupled maps. *Physical Review E*, **67** (2003), 026224.
- [8] BRANDNER, K., SAITO, K., AND SEIFERT, U. Thermodynamics of micro-and nano-systems driven by periodic temperature variations. arXiv:1505.07771,(2015).
- [9] BRAUN, S., RONZHEIMER, J. P., SCHREIBER, M., HODGMAN, S. S., ROM, T., BLOCH, I., AND SCHNEIDER, U. Negative absolute temperature for motional degrees of freedom. *Science*, **339** (2013), 52.
- [10] BRICMONT, J. Science of chaos or chaos in science. *Pysicalia Mag.*, **17** (1995), 159.
- [11] BRITO, R., RENNE, M., AND VAN DEN BROECK, C. Dissipative collapse of the adiabatic piston. *EPL (Europhysics Letters)*, **70** (2005), 29.
- [12] BUONSANTE, P., FRANZOSI, R., AND SMERZI, A. On the dispute between boltzmann and gibbs entropy. arXiv:1506.01933 (2014).

- [13] BURIONI, R., CASSI, D., CECCONI, F., AND VULPIANI, A. Topological thermal instability and length of proteins. *Proteins*, **55** (2004), 529.
- [14] BUSTAMANTE, C., LIPHARDT, J., AND RITORT, F. The nonequilibrium thermodynamics of small systems. *Physics Today*, **58** (2005), 43.
- [15] CALLEN, H. B. *Thermodynamics and an Introduction to Thermostatistics*. John Wiley & Sons (2006).
- [16] CAMPISI, M. Construction of microcanonical entropy on thermodynamic pillars. *Physical Review E*, **91** (2015), 052147.
- [17] CAMPISI, M. AND KOBE, D. H. Derivation of Boltzmann Principle. *American Journal of Physics*, **78** (2009), 608.
- [18] CARNOT, S. *Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance*. ... (1824).
- [19] CARR, L. D. Negative temperatures? *Science*, **339** (2013), 42.
- [20] CELANI, A., BO, S., EICHHORN, R., AND AURELL, E. Anomalous thermodynamics at the microscale. *Physical review letters*, **109** (2012), 260603.
- [21] CELANI, A., BO, S., EICHHORN, R., AND AURELL, E. Anomalous Thermodynamics at the Microscale. *Physical Review Letters*, **109** (2012), 260603.
- [22] CENCINI, M., PALATELLA, L., PIGOLOTTI, S., AND VULPIANI, A. Macroscopic equations for the adiabatic piston. *Physical Review E*, **76** (2007), 051103.
- [23] CERCIGNANI, C. *Ludwig Boltzmann: the man who trusted atoms*. Oxford University Press (2006).
- [24] CHIBBARO, S., RONDONI, L., AND VULPIANI, A. *Reductionism, Emergence and Levels of Reality*. Springer (2014).
- [25] COSTANTINI, G., MARCONI, U. M. B., AND PUGLISI, A. Noise rectification and fluctuations of an asymmetric inelastic piston. *EPL (Europhysics Letters)*, **82** (2008), 50008.
- [26] COSTENIUC, M., ELLIS, R. S., TOUCHETTE, H., AND TURKINGTON, B. The generalized canonical ensemble and its universal equivalence with the microcanonical ensemble. *Journal of statistical physics*, **119** (2005), 1283.
- [27] CROSIGNANI, B., DI PORTO, P., AND SEGEV, M. Approach to thermal equilibrium in a system with adiabatic constraints. *American Journal of Physics*, **64** (1996), 610.
- [28] CURZON, F. AND AHLBORN, B. Efficiency of a carnot engine at maximum power output. *American Journal of Physics*, **43** (1975), 22.
- [29] DE GROOT, S. AND MAZUR, P. *Non-equilibrium Thermodynamics*. Dover (1984).

- [30] DELRE, E., CROSIGNANI, B., DI PORTO, P., AND DI SABATINO, S. Built-in reduction of statistical fluctuations of partitioning objects. *Phys. Rev. E*, **84** (2011), 21112.
- [31] DUNKEL, J. AND HILBERT, S. Consistent thermostatics forbids negative absolute temperatures. *Nat. Phys.*, **10** (2013), 67.
- [32] DÜRR, D., GOLDSTEIN, S., AND LEBOWITZ, J. A mechanical model of brownian motion. *Communications in Mathematical Physics*, **78** (1981), 507.
- [33] EHRENFEST, P. AND EHRENFEST, T. *The conceptual foundation of the statistical approach in mechanics*. Cornell University Press, New York (1956). (trasl. from 1911 German version).
- [34] EINSTEIN, A. *Investigations on the Theory of the Brownian Movement*. DoverPublications. com (1956).
- [35] ESPOSITO, M., KAWAI, R., LINDENBERG, K., AND VAN DEN BROECK, C. Efficiency at Maximum Power of Low-Dissipation Carnot Engines. *Physical Review Letters*, **105** (2010), 150603.
- [36] ESPOSITO, M., LINDENBERG, K., AND VAN DEN BROECK, C. Universality of Efficiency at Maximum Power. *Physical Review Letters*, **102** (2009), 130602.
- [37] FALCIONI, M., VILLAMAINA, D., VULPIANI, A., PUGLISI, A., AND SARRACINO, A. Estimate of temperature and its uncertainty in small systems. *American Journal of Physics*, **79** (2011), 777.
- [38] FALCIONI, M. AND VULPIANI, A. *Meccanica Statistica Elementare*. Springer (2014).
- [39] FERMI, E., PASTA, J., AND ULAM, S. Studies of nonlinear problems. *Los Alamos Report LA-1940*, **978** (1955).
- [40] FRENKEL, D. AND WARREN, P. B. Gibbs, Boltzmann, and negative temperatures. *Am. J. Phys.*, **83** (2015), 163.
- [41] FRULEUX, A., KAWAI, R., AND SEKIMOTO, K. Momentum transfer in nonequilibrium steady states. *Physical review letters*, **108** (2012), 160601.
- [42] GALLAVOTTI, G. AND COHEN, E. Dynamical ensembles in nonequilibrium statistical mechanics. *Phys. Rev. Lett.*, **74** (1995), 2694.
- [43] GARDINER, C. W. *Handbook of stochastic methods*. Springer Berlin (1985).
- [44] GASPARD, P. Hamiltonian dynamics, nanosystems, and nonequilibrium statistical mechanics. *Phys. A: Math. Gen.*, **369** (2006), 201.
- [45] GNOLI, A., PETRI, A., DALTON, F., GRADENIGO, G., PONTUALE, G., SARRACINO, A., AND PUGLISI, A. Brownian ratchet in a thermal bath driven by coulomb friction. *Phys. Rev. Lett.*, **110** (2013), 120601.

- [46] GOLDSTEIN, S. Boltzmann's approach to statistical mechanics. In *Chance in physics*, pp. 39–54. Springer (2001).
- [47] GOLDSTEIN, S. Typicality and notions of probability in physics. In *Probability in physics* (edited by Y. Ben-Menahem and M. Hemmo), pp. 59–71. Springer (2012).
- [48] GONG, Z. AND QUAN, H. Jarzynski equality, crooks fluctuation theorem, and the fluctuation theorems of heat for arbitrary initial states. *Physical Review E*, **92** (2015), 012131.
- [49] GRUBER, C., PACHE, S., AND LESNE, A. Two-time-scale relaxation towards thermal equilibrium of the enigmatic piston. *Journal of statistical physics*, **112** (2003), 1177.
- [50] HILBERT, S., HÄNGGI, P., AND DUNKEL, J. Thermodynamic laws in isolated systems. *Phys. Rev. E*, **90** (2014), 062116.
- [51] HUANG, K. *Statistical mechanics*. Wiley, New York (1988).
- [52] HUNT, K. L. C. Path integral solutions of stochastic equations for nonlinear irreversible processes: The uniqueness of the thermodynamic Lagrangian. *The Journal of Chemical Physics*, **75** (1981), 976.
- [53] HURTADO, P. I. AND REDNER, S. Simplest piston problem. i. elastic collisions. *Physical Review E*, **73** (2006), 016136.
- [54] IZUMIDA, Y. AND OKUDA, K. Molecular kinetic analysis of a finite-time Carnot cycle. *Europhys. Lett.*, **83** (2008), 60003.
- [55] IZUMIDA, Y. AND OKUDA, K. Onsager coefficients of a finite-time Carnot cycle. *Physical Review E*, **80** (2009), 021121.
- [56] JARZYNSKI, C. Comparison of far-from-equilibrium work relations. *Comptes Rendus Physique*, **8** (2007), 495.
- [57] JARZYNSKI, C. Comparison of far-from-equilibrium work relations. *C. R. Physique*, **8** (2007), 495.
- [58] KAC, M. *Probability and related topics in physical sciences*, vol. 1. American Mathematical Soc. (1957).
- [59] KHINCHIN, A. I. *Mathematical foundations of statistical mechanics*. Dover (1949).
- [60] KLAGES, R., JUST, W., AND JARZYNSKI, C. (eds.). *Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond*. Wiley (2013).
- [61] KOSKI, J. V., MAISI, V. F., PEKOLA, J. P., AND AVERIN, D. V. Experimental realization of a szilard engine with a single electron. *Proceedings of the National Academy of Sciences*, **111** (2014), 13786.

- [62] KUBO, R. *Statistical Mechanics. An Advanced Course with Problems and Solutions*. Elsevier (1965).
- [63] KUBO, R., TODA, M., AND SAITO, N. *Statistical Physics I: Equilibrium Statistical Mechanics*. Springer (1992).
- [64] LANDSBERG, P. T. Heat engines and heat pumps at positive and negative absolute temperatures. *J. Phys. A: Math. Gen.*, **10** (1977), 1773.
- [65] LANDSBERG, P. T. *Thermodynamics and Statistical Mechanics*. Dover (2014).
- [66] LEBOWITZ, J. AND SPOHN, H. A Gallavotti-Cohen-type symmetry in the large deviation functional for stochastic dynamics. *Journal of Statistical Physics*, (1999), 333.
- [67] LEBOWITZ, J. L. Boltzmann's entropy and time's arrow. *Physics today*, **46** (1993), 32.
- [68] LEBOWITZ, J. L., PERCUS, J., AND VERLET, L. Ensemble dependence of fluctuations with application to machine computations. *Physical Review*, (1967).
- [69] LIPHARDT, J., DUMONT, S., SMITH, S. B., TINOCO, I., AND BUSTAMANTE, C. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science (New York, N.Y.)*, **296** (2002), 1832.
- [70] MACHLUP, S. AND ONSAGER, L. Fluctuations and irreversible process. ii. systems with kinetic energy. *Physical Review*, **91** (1953), 1512.
- [71] MARCONI, U. M. B., PUGLISI, A., RONDONI, L., AND VULPIANI, A. Fluctuation-dissipation: response theory in statistical physics. *Physics Reports*, **461** (2008), 111.
- [72] MARTÍNEZ, I. A., ROLDÁN, É., DINIS, L., PETROV, D., PARRONDO, J. M., AND RICA, R. A. Brownian carnot engine. *Nature Physics*, **12** (2016), 67.
- [73] MAZUR, P. AND VAN DER LINDEN, J. Asymptotic form of the structure function for real systems. *Journal of Mathematical Physics*, **4** (1963), 271.
- [74] NORRIS, J. R. *Markov chains*. 2. Cambridge university press (1998).
- [75] O'NEIL, K. AND CAMPBELL, L. J. Comment on "Negative temperature of vortex motion". *Phys. Rev. E*, **47** (1993), 2966.
- [76] ONSAGER, L. Statistical hydrodynamics. *Nuovo Cimento, Suppl. VI Ser. IX* (1949), 279.
- [77] PACI, E. AND MARCHI, M. Intrinsic compressibility and volume compression in solvated proteins by molecular dynamics simulation at high pressure. *Proceedings of the National Academy of Sciences*, **93** (1996), 11609.

- [78] PERRIN, J. *Les atomes*. Librairie Félix Alcan (1913).
- [79] POLETTINI, M., VERLEY, G., AND ESPOSITO, M. Efficiency statistics at all times: Carnot limit at finite power. *Phys. Rev. Lett.*, **114** (2015), 050601.
- [80] PROESMANS, K., CLEUREN, B., AND VAN DEN BROECK, C. Linear stochastic thermodynamics for periodically driven systems. *Journal of Statistical Mechanics: Theory and Experiment*, **2016** (2016), 023202.
- [81] PROESMANS, K. AND VAN DEN BROECK, C. Onsager coefficients in periodically driven systems. *Physical review letters*, **115** (2015), 090601.
- [82] PUGLISI, A. *Transport and Fluctuations in Granular Fluids: From Boltzmann Equation to Hydrodynamics, Diffusion and Motor Effects*. Springer (2014).
- [83] PUGLISI, A., RONDONI, L., AND VULPIANI, A. Relevance of initial and final conditions for the fluctuation relation in markov processes. *J. Stat. Mech.*, (2006), P08010.
- [84] PUGLISI, A. AND VILLAMAINA, D. Irreversible effects of memory. *Europhys. Lett.*, **88** (2009), 30004.
- [85] RAMSEY, N. F. Thermodynamics and statistical mechanics at negative absolute temperatures. *Phys. Rev.*, **103** (1956), 20.
- [86] RUELE, D. *Statistical Mechanics, Rigorous results*. World Scientific (1969).
- [87] RUGH, H. H. Dynamical approach to temperature. *Phys. Rev. Lett.*, **78** (1997), 772.
- [88] SANO, T. G. AND HAYAKAWA, H. Roles of dry friction in the fluctuating motion of an adiabatic piston. *Physical Review E*, **89** (2014), 032104.
- [89] SARRACINO, A., GNOLI, A., AND PUGLISI, A. Ratchet effect driven by coulomb friction: the asymmetric rayleigh piston. *Physical Review E*, **87** (2013), 040101.
- [90] SCHMIEDL, T. AND SEIFERT, U. Efficiency at maximum power: An analytically solvable model for stochastic heat engines. *EPL (Europhysics Letters)*, **81** (2008), 20003.
- [91] SCHNEIDER, U., MANDT, S., RAPP, A., BRAUN, S., WEIMER, H., BLOCH, I., AND ROSCH, A. Comment on “Consistent thermostatics forbids negative absolute temperatures”. arXiv:1407.4127 (2014).
- [92] SEIFERT, U. Entropy production along a stochastic trajectory and an integral fluctuation theorem. *Physical review letters*, **95** (2005), 040602.
- [93] SEIFERT, U. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics*, **75** (2012), 126001.
- [94] SEKIMOTO, K. *Stochastic energetics*. Springer Verlag (2010).

- [95] SWENDSEN, R. H. Explaining irreversibility. *American Journal of Physics*, **76** (2008), 643.
- [96] TALKNER, P., LUTZ, E., AND HÄNGGI, P. Fluctuation theorems: Work is not an observable. *Physical Review E*, **75** (2007), 050102.
- [97] TEHVER, R., TOIGO, F., KOPLIK, J., AND BANAVAR, J. R. Thermal walls in computer simulations. *Phys. Rev. E*, **57** (1998), R17.
- [98] TOUCHETTE, H. The large deviation approach to statistical mechanics. *Phys. Rep.*, **478** (2009), 1.
- [99] VAN DEN BROECK, C. Thermodynamic Efficiency at Maximum Power. *Physical Review Letters*, **95** (2005), 190602.
- [100] VERLEY, G., ESPOSITO, M., WILLAERT, T., AND VAN DEN BROECK, C. The unlikely Carnot efficiency. *Nature Communications*, **5** (2014), 4721.
- [101] VILAR, J. M. G. AND RUBI, J. M. Communication: System-size scaling of Boltzmann and alternate Gibbs entropies. *J. Chem. Phys.*, **140** (2014), 1.
- [102] VILLAMAINA, D. AND TRIZAC, E. Thinking outside the box: fluctuations and finite size effects. *European Journal of Physics*, **35** (2014), 035011.
- [103] VON HELMHOLTZ, H. Principien der Statik monocyclischer Systeme. *Journal für die reine und angewandte Mathematik*, **1884** (1884), 111.
- [104] VON SMOLUCHOWSKI, M. Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen. *Annalen der physik*, **326** (1906), 756.
- [105] VULPIANI, A., CECCONI, F., CENCINI, M., PUGLISI, A., AND VERGNI, D. (eds.). *Large Deviations in Physics*. Springer (2014).
- [106] WANG, J.-S. AND SWENDSEN, R. H. Negative temperatures and the definition of entropy. *Physica A: Statistical Mechanics and its Applications*, **453** (2016), 24.
- [107] ZANGHÌ, N. I fondamenti concettuali dell'approccio statistico in fisica. In *La natura delle cose: Introduzione ai fondamenti e alla filosofia della fisica* (edited by V. Allori, M. Dorato, F. Laudisa, and N. Zanghì), pp. 202–247. Carocci (2005).