

Springer Theses
Recognizing Outstanding Ph.D. Research

Édgar Roldán

Irreversibility and Dissipation in Microscopic Systems



Springer

Springer Theses

Recognizing Outstanding Ph.D. Research

For further volumes:
<http://www.springer.com/series/8790>

Aims and Scope

The series “Springer Theses” brings together a selection of the very best Ph.D. theses from around the world and across the physical sciences. Nominated and endorsed by two recognized specialists, each published volume has been selected for its scientific excellence and the high impact of its contents for the pertinent field of research. For greater accessibility to non-specialists, the published versions include an extended introduction, as well as a foreword by the student’s supervisor explaining the special relevance of the work for the field. As a whole, the series will provide a valuable resource both for newcomers to the research fields described, and for other scientists seeking detailed background information on special questions. Finally, it provides an accredited documentation of the valuable contributions made by today’s younger generation of scientists.

Theses are accepted into the series by invited nomination only and must fulfill all of the following criteria

- They must be written in good English.
- The topic should fall within the confines of Chemistry, Physics, Earth Sciences, Engineering and related interdisciplinary fields such as Materials, Nanoscience, Chemical Engineering, Complex Systems and Biophysics.
- The work reported in the thesis must represent a significant scientific advance.
- If the thesis includes previously published material, permission to reproduce this must be gained from the respective copyright holder.
- They must have been examined and passed during the 12 months prior to nomination.
- Each thesis should include a foreword by the supervisor outlining the significance of its content.
- The theses should have a clearly defined structure including an introduction accessible to scientists not expert in that particular field.

Édgar Roldán

Irreversibility and Dissipation in Microscopic Systems

Doctoral Thesis accepted by
Universidad Complutense de Madrid, Spain



Springer

Author
Dr. Édgar Roldán
Departamento de Física Atómica
Molecular y Nuclear and GISC
Universidad Complutense de Madrid
Madrid
Spain

Supervisor
Prof. Juan M. R. Parrondo
Departamento de Física Atómica
Molecular y Nuclear and GISC
Universidad Complutense de Madrid
Madrid
Spain

ISSN 2190-5053
ISBN 978-3-319-07078-0
DOI 10.1007/978-3-319-07079-7
Springer Cham Heidelberg New York Dordrecht London

ISSN 2190-5061 (electronic)
ISBN 978-3-319-07079-7 (eBook)

Library of Congress Control Number: 2014941093

© Springer International Publishing Switzerland 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

*Dedicated to my family,
and to my team, EMOTAMZOR*

Supervisor's Foreword

We learn in college physics courses that the entropy of an isolated system can only increase in time. This is one of the formulations of the Second Law of Thermodynamics. The immediate corollary is that a process where the entropy increases cannot be reversed in time, and it is accordingly called irreversible.

In the search for a microscopic rationale of the Second Law, Maxwell and Boltzmann found that the above statements have a probabilistic nature. In other words, there is a nonzero probability that entropy decreases and, consequently, the time reversal of a process where entropy increases can occur spontaneously with nonzero probability. For macroscopic systems this “non-zero probability” is overwhelmingly small. On the other hand, in the fluctuating microscopic world, one can easily observe those ephemeral violations of the Second Law. The uphill motion of a Brownian particle, where the energy from a thermal bath is converted into potential energy, is a well-known example. As early as 1904, Poincaré wrote that, in the Brownian motion, “we see under our eyes now motion transformed into heat by friction, now heat changed inversely into motion, and that without loss since the movement lasts forever. This is contrary to the principle of Carnot.”

In the last two decades there has been a systematic exploration of the fluctuations in the energetics of small systems, both from experimental and theoretical perspectives. Several techniques to manipulate Brownian particles and to measure minute energy transfers have been developed, boosted by an increasing interest in the transduction of energy carried out by proteins within living cells. On the theoretical side, the formulation of a stochastic thermodynamics for random trajectories and the so-called fluctuation theorems have yielded a better understanding of the statistical properties of the work and the dissipated heat in processes in small systems.

The present thesis by Édgar Roldán describes several significant and novel contributions to the field of stochastic thermodynamics. The main one is a technique that relates dissipation and irreversibility. If irreversibility is probabilistic, then we could try to quantify it by comparing the probability of observing a process with the probability of observing the reverse. This is in fact possible using a powerful concept from information theory, the Kullback–Leibler divergence, which measures how different two probability distributions are. More interestingly, it turns out that this measure of irreversibility is proportional to the increase of

entropy during the process. We then recover the well-known fact that a macroscopic increase of entropy during a process (much larger than the Boltzmann constant) implies that it is practically impossible to observe its time reversal.

In the first part of the book, the quantitative relationship between entropy production and irreversibility, as measured by the Kullback–Leibler divergence, is explored in full detail and applied to different situations. The results are relevant not only from a fundamental point of view. Quantifying irreversibility turns out to be useful in biology to discern whether the behavior of a biological system, for example the filaments in the ear cells, are spontaneous thermal fluctuations or, on the contrary, induced by some active, ATP consuming, mechanism.

The second part of the book is devoted to experiments with Brownian particles manipulated by optical tweezers. One of these experiments reproduces a symmetry breaking, i.e., a process where a system “makes a choice” among different mutually exclusive options. The choice is random and induced by thermal fluctuations. Consequently, the effect of the symmetry breaking on the energetics can be explored using fluctuation theorems. There is indeed an important signature of the symmetry breaking which is universal, that is, does not depend on the physical nature of the system or the process inducing the symmetry breaking. That effect is an apparent decrease of the entropy of isolated systems and it prompts us to reconsider some basic aspects of the Second Law of Thermodynamics.

In summary, the work of Édgar Roldán reported here is a major contribution to statistical physics: it provides us with new techniques for analyzing the energetics of small systems and tackles fundamental and subtle aspects of the Second Law. More than a century after the explanation of Brownian motion given by Einstein, fluctuations seem to remain crucial for understanding how the macroscopic world arises from microscopic dynamics.

Madrid, April 2014

Prof. Juan M. R. Parrondo

Abstract

The relationship between dissipation and irreversibility is at the core of statistical mechanics and thermodynamics [1]. The first studies on this relationship were done in the context of linear irreversible thermodynamics, where the entropy produced by macroscopic systems weakly perturbed away from equilibrium was considered. Near equilibrium, the entropy production is proportional to the macroscopic currents or flows of the system such as the particle current or the heat flow [2]. In this regime, time irreversibility is revealed as a presence of currents in the system. The connection between dissipation (or entropy production) and irreversibility (currents or flows) is merely qualitative in linear irreversible thermodynamics, and cannot be trivially generalized to microscopic systems driven arbitrarily far from equilibrium. In addition, entropy production and other thermodynamic quantities such as heat or work done by a microscopic system in nonequilibrium processes were not understood until the introduction of stochastic thermodynamics.

Recent works in the framework of nonequilibrium statistical mechanics have extended the laws of thermodynamics to small-scale (microscopic or smaller) systems, where thermal fluctuations play an important role in the dynamics. These new results are part of a new emerging field within statistical mechanics called *stochastic thermodynamics*. Some of its most relevant results are the generalization of the first law to microscopic systems by defining the work and heat associated to a single stochastic trajectory or realization of a nonequilibrium process [3]. The entropy of a microscopic system along a stochastic trajectory and the generalization of the second law to overdamped Brownian particles was done in [4]. Both first and second law of thermodynamics are valid in the microscopic scale when ensemble averages of stochastic work, heat, and entropy are considered.

Fluctuations in small systems give rise to rare events that do not happen in the macroscopic world, such as an entropy decrease in the universe or negative dissipation in a single realization. *Fluctuation theorems* are exact relationships that express universal properties of the probability distribution of thermodynamic functionals of the stochastic trajectories of the system, such as work, heat, or entropy [5]. Crooks [6] fluctuation theorem, Jarzynski's equality [7] or the steady-state fluctuation theorem [8] are some of the most celebrated examples. Unlike linear irreversible thermodynamics, fluctuation theorems are valid for processes where the system is driven arbitrarily away from equilibrium.

In a recent work by Kawai et al. [9], the dissipation along a nonequilibrium processes for isolated microscopic systems that are initially connected to a thermal bath is investigated. For such processes, the average work dissipation can be expressed in terms of a relationship between the phase space densities of the forward and the time-reversed (backward) process. Their formula, which along this thesis will be referred to as the *KPB theorem*, expresses the entropy production in terms of the microscopic properties of the system. The result also implies a connection between dissipation and the distinguishability between the forward and the backward process via a metric, the Kullback–Leibler divergence (KLD), that measures such distinguishability [9, 10]. The KLD between the forward and backward phase space densities is a measure of the arrow of time. The latter result is the first quantitative relationship between dissipation (or entropy production) and irreversibility (KLD between forward and backward phase space densities) derived for microscopic systems driven out of equilibrium.

The KPB theorem is found to be applicable to many different scenarios [10], and can be used to reproduce relevant results in the context of information thermodynamics. For example, Landauer’s principle, which establishes the minimum amount of work needed to erase one bit of information, is a particular case of the KPB theorem. The energetics of the Szilard engine, which is a simplified one-particle Maxwell demon, can also be reproduced using the KPB theorem. An application of the result to microscopic systems that reach a nonequilibrium stationary state (NESS) was also suggested in [10]. A first experimental test of the relationship between dissipation and the KLD in the NESS was done by dragging Brownian particles with an optical tweezer in [11].

In this thesis, we study the relationship between dissipation and irreversibility for microscopic systems that reach a NESS within the framework of stochastic thermodynamics. We are first interested in expressing this relationship in the NESS and secondly, we aim to apply this result to practical situations. The main goal of this work is to find a technique that allows one to estimate the average dissipation of a microscopic system in a NESS from the statistics of a single stationary trajectory produced by the system during the process. We want our technique to be able to estimate the average dissipation by using any data (one or several discrete or continuous degrees of freedom) sampled from the system and even ignoring any physical detail of the process. Our method uses the KLD to measure the distinguishability between a stationary trajectory and its time-reversal, providing a quantitative tool to measure the arrow of time in the NESS, i.e., the time irreversibility.

Our work is a continuation of the theoretical result first proved in [9] and of the experimental test done in [11]. We are interested in going one step further, by designing a tool to estimate the dissipation of any physical microscopic system that reaches a NESS, focusing on biological systems, where nonequilibrium processes play a fundamental role. We want our technique to be able to capture the irreversibility of a process even when having access only to partial information of the system (i.e., when only a subset of the phase space can be sampled), which is

the most common situation in data obtained from experiments. In addition, we aim to explore the relationship between irreversibility and dissipation in the NESS experimentally by designing new thermodynamic protocols using micromanipulation techniques.

We first clarify in Chap. 2 the relationship between dissipation and irreversibility for microscopic systems that reach a NESS, by expressing the average entropy production in a NESS in terms of the KLD rate between the probability to observe a trajectory and the probability to observe the time-reversed trajectory. In order to apply our technique to a variety of physical situations, we develop in Chap. 3 estimators of the KLD rate for both discrete and continuous data. We find techniques that allow estimating the KLD using finite data from a single stationary trajectory. For discrete time series, the best strategy consists on doing counting statistics of blocks or sequences of m-data as we describe in Sect. 3.1.1. In Sect. 3.1.2, we introduce a second estimator for discrete data based on *zippers* or compression algorithms that outperforms the estimator based on block entropies when there is not sufficient statistics. Additionally, we introduce two new different techniques to measure the KLD rate for real-valued or continuous time series. First, a novel estimator that transforms a time series into a directed graph whose in-going and out-going degree distributions capture the irreversibility of a time series (see Sect. 3.3). This method distinguishes between reversible and irreversible generic stochastic processes as well as between conservative and dissipative chaotic processes. We also introduce in Sect. 3.2.2 a second estimator of the KLD rate for continuous correlated time series which uses an asymmetric functional to produce a new uncorrelated series whose KLD is easier to calculate.

We first apply our technique to a discrete flashing ratchet model (see Chap. 4), analyzing the bound to the dissipation provided by the KLD rate when one has access to trajectories sampled with full or partial information of the physical system. We show that, even when we only measure *a single trajectory* containing partial information of the system, our technique can distinguish between equilibrium and NESS. When an external force stalls the ratchet, no net current of particles is observed and linear irreversible thermodynamics fails to predict that the system is dissipating energy. However, our method is able to correctly predict that the system is producing entropy despite no net current being observed.

Our next step is to apply our technique to real data obtained from a biological system that produces a real-values stationary time series. We aim to distinguish between NESS and equilibrium states, which is equivalent to distinguish between active (ATP consuming) and passive biological processes. In Chap. 5 we consider experimental data of the spontaneous oscillations of bullfrog's ear hair bundles. Our KLD rate estimator based on the use of residuals distinguishes between active and passive spontaneous oscillations of ear hair bundles from bullfrog's sacculus. Our technique only requires the measurement of spontaneous oscillations of the position of the top of the hair bundles during ~ 100 s, outperforming significantly the requirements of previous works [12]. In addition, our KLD rate estimator indicates the minimum irreversibility to sustain spontaneous oscillations in

bullfrog's ear hair bundles, which is of the order of the energy consumption rate from ATP hydrolyzation by a single adaptation motor.

The last part of the thesis is devoted to the theory and experiments of other interesting phenomena of stochastic thermodynamics. The energetics of a symmetry breaking and a symmetry restoration can also be calculated using the relationship between dissipation and irreversibility at the microscopic scale. When a microscopic system breaks a symmetry, the accessible phase space of the system is reduced to a subset of the full phase space. We will show in Chap. 6 that, when the symmetry is broken quasistatically, the total entropy decreases. We have derived a formula that relates the average entropy production when the system chooses an instance with the probability of the system to choose that option. Our formula is valid in the quasistatic limit and it is universal: it does not depend on the physical nature of the system or on the mechanism that biases the probability to choose any option. In addition, we have obtained an equivalent formula for the symmetry restoration process, which, together with the symmetry breaking result, can be used to derive Landauer's principle and the energetics of the Szilard engine as discussed in Chap. 6. For a complete validation of our formula, we have checked our results experimentally, by trapping a Brownian particle in a dual optical tweezer and using external electric fields to tune the probability of the bead to choose between the two traps.

Finally, as another experimental application of stochastic thermodynamics, we have designed an experiment to tune the amplitude of the Brownian fluctuations of a microscopic charged sphere. The setup consists of a Brownian particle trapped with an optical tweezer and an external random electric field of tunable amplitude. With our setup, we are able to control the kinetic temperature of a microscopic particle from room temperature to 3,000 K in a fast timescale (\sim ms). This technique opens the possibility of implementing a variety of thermodynamic processes and cycles at the microscale, which need a fine control of the temperature, like the Carnot engine.

To conclude, in this thesis, we show that the time reversal asymmetry of a stationary time series provides information about the entropy production of the physical mechanism generating the series, even if one ignores any detail of that mechanism. We have performed a complete study on the relationship between dissipation and irreversibility for microscopic system in a NESS. Our study includes analytical results, simulations and experimental validations and provide new insights in different fields such as: stochastic thermodynamics, time series analysis, information theory, or biophysics.

Our work has potential applications from an experimental point of view. Using our technique, one can, from the data of a single (discrete or continuous) stationary trajectory obtained sampling any physical variable of a microscopic system, estimate the average dissipation rate of the mechanism that generated the data. This opens the possibility to distinguish between different active and passive biological processes using the data from only one experimental sampling.

Our experimental results using optical tweezers provide a new insight into the thermodynamics of choice, finding exact relationships between entropy production

and the probability to choose an option in a symmetry breaking. This could be used to estimate the probability of a biomolecule to adopt a specific conformation from energetic measurements. On the other hand, the implementation of adiabatic processes in the microscopic scale—and therefore the construction of a Carnot engine—could also be possible by using active feedback when tuning the kinetic temperature of a Brownian particle with our experimental setup described in Chap. 7.

References

1. H.B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, vol. 1, 2nd edn. (Wiley-VCH, New York, August 1985), p. 512
2. D. Konddepudi, I. Prigogine, *From Heat Engines to Dissipative Structures*, (John Wiley & Son, Chichester, 1998)
3. K. Sekimoto, Progress Theoret. Phys. Suppl. **130**, 17–27 (1998)
4. U. Seifert, Phys. Rev. Lett. **95**, 40602 (2005)
5. U. Seifert, arXiv of preprint arXiv: 1205.4176 (2012)
6. G. Crooks, Phys. Rev. E **60**, 2721 (1999)
7. C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997)
8. D.J. Evans, E. Cohen, G. Morriss, Phys. Rev. Lett. **71**, 2401–2404 (1993)
9. R. Kawai, J.M.R. Parrondo, C. Van den Broeck, Phys. Rev. Lett. **98**, 80602 (2007)
10. J.M.R. Parrondo, C. Van den Broeck, R. Kawai, New J. Phys. **11**, 073008 (2009)
11. D. Andrieux et al., Phys. Rev. Lett. **98**, 150601 (2007)
12. P. Martin, A. Hudspeth, F. Jülicher, Proc. Nat. Acad. Sci. **98**, 14380–14385 (2001)

Acknowledgments

I want first to acknowledge the contribution of my Ph.D. supervisor, Juan M. R. Parrondo, to this work. This work would have never been possible without his help, his advice, his inspiration to solve complicated problems and the opportunities that he has given me to collaborate with other physicists in different institutions.

During my Ph.D., I was member of Departamento de Física Atómica Molecular y Nuclear in the faculty of physics of Universidad Complutense de Madrid. I acknowledge the support given by my department every time I needed their help and the opportunity to collaborate in teaching *Física Aplicada a la Biología* during 2 years.

I also had the chance to become a member of *Grupo Interdisciplinar de Sistemas Complejos* (GISC¹), with whom I participated in several stimulating workshops, conferences, and seminars during these years.

Within my department, I have been working in *Group of Statistical Mechanics*² together with Juan M. R. Parrondo, Ricardo Brito, Luis Dinís, Jordan M. Horowitz, Pablo Rodríguez, and Rahul Marathe. Ricardo has helped me to resolve many problems and difficulties in both administrative and scientific areas, providing many interesting suggestions to improve my work in the group seminars. I am very grateful to Luis Dinís, with whom I shared office along this years, for his inexhaustible help in both physics and IT-related problems. Jordan M. Horowitz is one of the most hard-working, brilliant, and perfectionist postdoctoral physicists I have had the chance to work with. I want to highlight his help in editing the manuscript published in [1] as well as his multiple interesting advices about statistical mechanics. Pablo Rodríguez's advices have been very valuable for me during this years as well as his suggestions in the group seminars. I also acknowledge the stimulating discussions with Rahul Marathe during his stay in our group and his help to use mathematica.

Thanks also to my collaborators from *Universidad Politécnica de Madrid*, Bartolo Luque, Lucas Lacasa and Àngel M. Nuñez for their ideas and their work in the paper about the visibility technique.

¹ GISC website: <http://valbuena.fis.ucm.es/gisc/>

² Group of Statistical Mechanics website: <http://seneca.fis.ucm.es/>

I am especially grateful to *Max-Planck Institut Für Physik Komplexer Systeme* (MPI-PKS, Dresden, Germany) and *Max-Planck Institute for Molecular Cell Biology and Genetics* (MPI-CBG, Dresden, Germany) for the opportunity that I was given to do several research stays in both institutes. Prof. Dr. Frank Jülicher has given me invaluable help every time I stayed in MPI-PKS in Biological Physics group, where I always feel in a very motivating scientific atmosphere. I emphasize his crucial contribution to our work related with hair bundles. Thanks to Dr. Stephan Grill, I could also work in Grill Lab within MPI-CBG. I am grateful to him for letting me participate in the scientific activities of his lab and for showing me an optical tweezers experimental setup for the first time in my life.

I have had the opportunity to work in the *Institute of Photonic Sciences* (ICFO, Castelldefels) in Petrov Lab, where all the experimental results of the thesis except the part of the ear hair bundles were obtained. Thanks to Dmitry Petrov I have been able to learn the methodology of the work in experimental physics and also to gain insight into the experimental approach or point of view of optical tweezers experiments. Ifñaki Martínez has also been crucial in all the experiments of our collaboration. I am very grateful to him because of his non-stop experimental work and endless motivation, especially in the symmetry breaking experiment. Thanks as well to the other members of the Optical Tweezers group (Raúl, Pau, Antonio) with whom I had the chance to work with in the best possible working environment.

I was funded by *Beca de Formación de Profesorado Universitario* (FPU, Spanish Government) during all my Ph.D. This thesis was also funded by projects by Grant MOSAICO (Spanish Government), ENFASIS (Spanish Government), and MODELICO (Comunidad de Madrid, Spain) and partially by the scholarship Becas para el estudio de Master en España. Thanks also to GEFENOL for letting me to participate in two very stimulating summer schools. Optical tweezers experiments were funded by Fundacio Privada Cellex Barcelona, Generalitat de Catalunya Grant No. 2009-SGR-159, and from the Spanish Ministry of Science and Innovation (MICINN FIS2008-00114, FIS2011-24409).

My thesis is dedicated to all the people who have shared my life experience during the last years. These words are a personal acknowledgment to all of them:

First, to my family, for their constant support. Especially to my family from San Sebastián, but mainly to my parents, my brother, and my aunt.

Secondly, to my long-lasting friends of Madrid, David, José Manuel *El Chío*, and Don Francisco de la Poza, with whom I expect to continue to have nice times in the future.

I am also in debt with all the people who have made my long stay in the UCM one of the best experiences in my life.

Thanks to my *classmates* for making physics more motivating: Cojoncio, Juantxo, Ricardo, Re, Mario, JJ, Hugo, Mariano Schneider, Josérra, Bea *Betapata*, Eduardo Toraño *Pecholobo*, Álex, Maru, Ismael, Rubén Pedro, Héctor, Bea, Laura, Clara, Jesús, Marta...

Special acknowledgments to “the pilots,” with whom I shared many kind of good experiences: Cris, Paula, Loreto, Gema, Ana, Óscar, Rocío, Jorge, Chechu, Rodri, Jónatan, Rafa, Sergio, Mercedes, Ester, Alberto ...

Also, a tribute to the “survivors” of the Ph.D. (and Postdoc) in UCM and special guests of Candi and Gañán: Ricardo Torres, Giovanni *Don G, Herr* Markus, Alex *Elbelga*, Ginevra, Jessica, Federica, Davide, Guillermo, Jacobo, José, JAR, David...

This line is to show RESPECT to Nikos *Zeos Fytas*.

It has been amazing to organize La Liga de Físicas during 3 years with Miguel *Villar* and Jose. Thanks a lot to all the teams and friends I made playing football.

Here are some of the most celebrated Dresdeners that I would like to mention as an important part of my life during the Ph.D.: Saúl, Manuel, Cris, Elena Q., Alba, Elena D., Edu, Pablo, Ana, Abi, Marcus, Stephan, JB (awesome biolympics), Nick, Mirjam, Birgit, Martin, Vero, Stoyno, AC.. and especially the *Indian gang*: Sundar, Sukhdeep, etc.

To all the starving scientists of Benasque, for the scientific and gastronomic moments: Oriol, Dani, Edoardo *el Sabio*, Svetlo, Virginia, Paula, Jordi, Ana, La Tofí, Juan, Esther, Ula, Fran, Elisa, Rubén, Alessio, Rafa, ... and to our friends of La Parrilla. *Pit i Collons*.

The *golden box* clearly deserves to be mentioned in this section. Thanks Paula, Loreto, Cris, Anita, but also to his special collaborators, Javi, Dr. Essalhi, Araceli, etc.

Of course, thanks to the awesome *DomiLab* for existing and making pleasant any hard working day: Juanxo *el Alcarreño*, Betapata, Joserra, Marco. To Dominicus, for helping me to solve countless problems and being an example to follow for me. To Viviana, even more specially if possible, for being always where and when needed.

Last but not least, enormous acknowledgments to my teammates of EMOTAMZOR. *Un día de primavera de hace ya 5 años soplaban viento de levante en el Paraninfo... Desde entonces 6 temporadas inenarrables con vosotros luchando por la conquista de numerosos cetros deportivos, y aún más importante, forjando grandísimas amistades y teniendo una experiencia inolvidable jugando con vosotros. Ha sido un honor ser vuestro capitán jugando a tope con EQTLI. Gracias Ré, Cordobé, Torres, Mariete Ley, Llorente, Carlas, Canilla, Maese, Vigo, Chacal, Iríbar, Mágico, Migue, Oliverio, Quiros, Discofarras, Futre, Rubinho, Héczor, Bardo, Cojoncio, Carlitos, Camarada, etc. Maese puso la puntilla.*

I am sure that I have forgotten people in this short chronicle. Thanks to the ones I forgot and are now reading these lines.

Édgar Roldán

Reference

1. É. Roldán, J.M.R. Parrondo, Phys. Rev. E **85**, 031129 (2012)

Contents

Part I Introduction

1 Small-Scale Thermodynamics	3
1.1 Review of Thermodynamics	4
1.2 Maxwell's Demon and Szilard Engine	7
1.3 Review of Statistical Mechanics	10
1.4 Stochastic Thermodynamics I: Stochastic Energetics	15
1.4.1 Brownian Motion and Langevin Equation	15
1.4.2 First Law of Thermodynamics in the Mesoscopic Scale.	18
1.4.3 Second Law of Thermodynamics in the Mesoscopic Scale	20
1.5 Stochastic Thermodynamics II: Fluctuation Theorems	22
1.5.1 Jarzynski's Equality	22
1.5.2 Crooks's Theorem.	23
1.5.3 Steady State and Transient Fluctuation Theorems	24
1.6 Experiments	26
1.7 Outline and Goals	31
References	33

Part II Irreversibility and Dissipation

2 Dissipation and Kullback–Leibler Divergence	37
2.1 What Is Time Irreversibility?	38
2.2 Average Dissipation in Irreversible Processes	38
2.2.1 Linear Irreversible Thermodynamics	38
2.2.2 Entropy Production in Microscopic Systems.	39
2.2.3 KPB Theorem	40
2.3 Kullback–Leibler Divergence and Irreversibility	44
2.4 Dissipation and Irreversibility in the Nonequilibrium Stationary State	50

2.5	Discrete Systems	52
2.5.1	Markov Chains Obeying Local Detailed Balance	53
2.5.2	Hidden Markov Processes	56
	References	59
3	Estimating the Kullback–Leibler Divergence	61
3.1	Discrete Case	62
3.1.1	Plug-in Estimators	62
3.1.2	Compression-Based Estimator	64
3.2	Continuous Case	68
3.2.1	Symbolization and Nearest-Neighbours	69
3.2.2	Autoregressive Models	70
3.3	Visibility Technique	73
3.3.1	The Method	73
3.3.2	Validation of the Method	77
	References	84
4	A Case Study: The Flashing Ratchet	87
4.1	Ratchets	88
4.2	The Model	90
4.3	Full Information	92
4.4	Partial Information	94
4.5	Detecting Irreversibility in the Absence of Currents	96
4.6	Estimating Dissipation with the Visibility Technique	98
4.7	Conclusions	99
	References	100
5	Application to Biology: The Ear Hair Bundle	103
5.1	Irreversibility in Biology: Ear Hair Bundles	104
5.2	Simulations	110
5.3	Application to Experimental Data	118
5.4	Conclusions	121
	References	122
Part III Experimental Tests and Applications of Stochastic Thermodynamics		
6	Energetics of Symmetry Breaking	125
6.1	Symmetry Breaking and Symmetry Restoration	126
6.2	Experimental Setup	132
6.3	Experimental Test	135
6.4	Building a Szilard Engine	144
6.5	Numerical Simulations	147

Contents	xxi
----------	-----

6.6 Conclusions	150
References	151
7 Effective Heating with Random Forces	153
7.1 Controlling the Temperature of Small Systems	154
7.2 Experimental Setup	156
7.3 Brownian Sphere in a Stationary Trap with Additional Noise	159
7.4 A Nonequilibrium Process	163
7.5 Applications of Our Setup	167
7.5.1 Sampling Energy Landscapes and Kramers Transitions	167
7.5.2 Building Microscopic Heat Engines	169
7.6 Conclusions	170
References	171

Part IV Conclusions

8 Conclusions and Outlook	175
References	178
Appendix A: Appendix to Chapter 1	179
Appendix B: Appendix to Chapter 2	185
Appendix C: Appendix to Chapter 3	193
Appendix D: Appendix to Chapter 5	197
Appendix E: Appendix to Chapter 6	199
Appendix F: Appendix to Chapter 7	205
Publications	211

Part I

Introduction

Chapter 1

Small-Scale Thermodynamics

The laws of thermodynamics apply to macroscopic systems, which are described by a large number of degrees of freedom ($N \sim 10^{23}$). Since the relative deviations from the mean behavior scale like $1/\sqrt{N}$, the deviations from the laws of thermodynamics are extremely unlikely to occur in the macroscopic scale [37]. In the last decade, it has been possible to manipulate and measure physical systems at the micro and nano scale, such as colloidal particles [22], DNA [10] or molecular motors [63]. At this scale, deviations from the mean behavior are more apparent. Moreover, the random forces exerted by thermal fluctuations are of the same order of magnitude of the forces that drive microscopic systems out of equilibrium. Notice that the order of magnitude of forces exerted by microscopic systems are of the order of piconewtons and their characteristic displacements of the order of nanometers, and the average energy of thermal fluctuations at room temperature is of the order of pN nm, $kT \simeq 4 \text{ pN nm}$ for $T = 25^\circ\text{C} = 298.15 \text{ K}$.

The need to describe the dynamics of small systems has led to the introduction of *stochastic thermodynamics* [57], which aims to extend thermodynamics to the regime where thermal fluctuations are relevant, that is, systems of small size (characteristic size of $\sim \mu\text{m}$) which undergo processes that can be sampled at submillisecond time resolution. By defining the stochastic heat and work associated to a stochastic trajectory [60] the first law can be extended to microscopic systems driven out of equilibrium. The second law of thermodynamics presents deviations at this scale: the total entropy production can decrease in a single realization of the process. However, the average entropy production over many realizations of the process is positive, being the second law true only in average [36].

In the last decade, there has been an increasing interest in describing laws that quantify the probability to observe deviations from the mean of thermodynamic quantities in nonequilibrium processes of microscopic systems. *Fluctuation theorems* provide quantitative relations that measure the probability to observe rare events in terms of the work done on the system [20, 35] or the entropy production [29]. In addition to that, fluctuation theorems open new possibilities in the microscopic scale,

such as providing tools to measure quantify time irreversibility or to design novel artificial small-scale engines.

This chapter reviews the basic concepts used along the rest of the work. The chapter is organized as follows: In Sect. 1.1, we review the basic concepts of macroscopic thermodynamics, focusing on the first and second law of thermodynamics. In Sect. 1.2 we study the first thought experiments that explored the limitations of the second law of thermodynamics, focusing on Maxwell's demon and Szilard's engine. In Sect. 1.3 we present some the main results of statistical mechanics that connect the thermodynamics of macroscopic systems with the statistics of their microscopic constituents. In Sect. 1.4, we review the fundamental aspects of stochastic thermodynamics that concern the energetics of microscopic particles, namely, the definition of stochastic work and heat and the expression of the first and second laws in the realm of stochastic processes. In Sect. 1.5 we discuss the theoretical framework of the fluctuation theorems connected to this work, and in Sect. 1.6 we show some of the most important experiments done in the recent years in stochastic thermodynamics.

1.1 Review of Thermodynamics

During a thermodynamic process, the total energy of a system changes in general. The *internal energy* of a system, U , is defined as the total energy of a system or the sum of the kinetic and the potential energy of all the atoms and electrons that form the system. During a process, part of the change of the internal energy of the system is due to the change of the control parameter by the external agent energy and the rest is exchanged with the thermal reservoir. The part of the energy that is transferred to the system that is due to the action of the external agent is called *work*, W , and the energy transferred from the thermal environment to the system is called *heat*, Q . The sum of these two contributions equals to the change in internal energy $\Delta U = U_{\text{fin}} - U_{\text{ini}}$,

$$\Delta U = Q + W. \quad (1.1)$$

The above relation is known as the *First Law of Thermodynamics*. In Fig. 1.1 we sketch the first law of thermodynamics. The arrows indicate the direction of the energy flow that corresponds with a positive sign of the heat and the work. When, for a closed system, an infinitesimal change of the control parameter is done, the internal energy of the system changes a small amount dU , and the first law is expressed as follows:

$$dU = \delta Q + \delta W, \quad (1.2)$$

δQ and δW being the amount of heat and work that are transferred to the system when a differential change in the control parameter is done. Notice that δQ and δW are *inexact differentials*, which do not satisfy the usual properties of a differential. For a process in which a system starts in equilibrium state A and ends in another equilibrium state B , the integral of δQ along the process is the total heat transferred

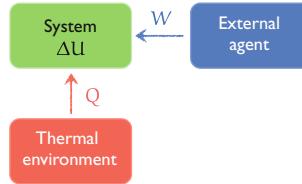


Fig. 1.1 Illustration of the first law of thermodynamics. The heat, Q , is defined positive when it is transferred from the environment to the system and the work W is defined positive when it is done from the external agent to the system. The change in internal energy in the system is equal to the sum of the heat transfer plus the work done along the process, $\Delta U = Q + W$

to the system $Q = \int_A^B \delta Q$. If the process is irreversible, this integral cannot be expressed as the value of a function (primitive) in B minus the value of the same function in A , since the heat Q depends not only on the initial and final states but also on the path that the system follows in the coordinate space. Notice that the work is also path dependent, at odds with the internal energy, which is a state function and $\int_A^B dU = U_B - U_A$. In a cyclic process, the change of internal energy is $\Delta U_{\text{cycle}} = 0$ since the initial and final states are the same irrespective of how the process is done, but not the heat and the work, which are trajectory dependent. In this case, by virtue of the first law $Q_{\text{cycle}} = -W_{\text{cycle}}$.

Not all the processes compatible with the first law of thermodynamics occur in nature. A glass falling to the ground and smashing into pieces is an example. This process is likely to be observed in nature, but not its time reversal, where the pieces retrace their path backwards until reaching their original position. The second law of thermodynamics expresses a restriction of some processes to occur in terms of a function called *entropy*. Let us consider a reversible thermodynamic process. The following quantity

$$\int_A^B \frac{\delta Q}{T}, \quad (1.3)$$

is the total heat (over T) transferred from the environment to the system during the reversible process. There are many possible reversible processes that drive the system from the initial equilibrium state A to the final equilibrium state B . It can be shown [28] that this quantity does not depend on the (reversible) process and it can therefore be expressed as

$$\int_A^B \frac{\delta Q}{T} = S_B - S_A = \Delta S, \quad (1.4)$$

that is, as a difference between the value of a function in the end of the process minus the value of the same function in the beginning of the process. That function is called

entropy of the system, S , which is defined as the heat exchanged in a reversible process $\Delta S = Q_{\text{rev}}/T$. When the same process starting in equilibrium state A and ending in equilibrium state B is done irreversibly, the heat exchanged during the process is smaller than the heat exchanged in the reversible case, or equivalently,

$$\begin{aligned} \int_A^B \frac{\delta Q}{T} &\leq \int_A^B \frac{\delta Q_{\text{rev}}}{T} \\ \int_A^B \frac{\delta Q}{T} &\leq \Delta S, \end{aligned} \quad (1.5)$$

where equal sign holds only for reversible processes. The above result is known as the *Second Law of Thermodynamics*.

Let us consider an irreversible isothermal process where a physical system is in contact with a single thermal environment that is at constant temperature, T . In this case, the integral in Eq. (1.5) equals to the heat transfer over the temperature, $\int_A^B \frac{\delta Q}{T} = \frac{Q}{T}$. The entropy change of the system can be expressed is in this case,

$$\Delta S = \frac{Q}{T} + S_{\text{prod}}, \quad (1.6)$$

where S_{prod} is called the *entropy production* of the system. According to the second law of thermodynamics (1.5), entropy production increases in irreversible processes,

$$S_{\text{prod}} \geq 0, \quad (1.7)$$

where the equal sign holds when the process is done reversibly. Entropy production can be interpreted as the sum of two terms following its definition in (1.6),

$$S_{\text{prod}} = \Delta S - \frac{Q}{T}, \quad (1.8)$$

where the first term, ΔS , is the entropy change in the system, and the second term, $-\frac{Q}{T}$, is the entropy change in the thermal bath (recall that Q is the heat absorbed by the system from the bath and therefore $-Q$ is the heat change in the bath). Therefore, for a system in contact with a single thermal bath at temperature T , entropy production can be seen as the total entropy production in a super-system or *universe* that contains the system and the environment. For this reason, entropy production S_{prod} is often called total entropy change, ΔS_{tot} , and the second law is expressed as $\Delta S_{\text{tot}} \geq 0$.

If we consider the first law (1.1) and we replace the heat by $Q = \Delta U - W$ in the definition of entropy production (1.8), the second law (1.6) implies that

$$W - (\Delta U - T \Delta S) \geq 0. \quad (1.9)$$

The following thermodynamic function

$$F = U - TS, \quad (1.10)$$

is called *Helmholtz Free Energy* of the system or *Free Energy* of the system [31]. The physical meaning of the free energy as a thermodynamic function can be done following Eq. (1.9),

$$W - \Delta F \geq 0. \quad (1.11)$$

The equal sign holds for reversible processes, $W_{\text{rev}} = \Delta F$, whereas for irreversible processes the work is $W_{\text{rev}} \geq \Delta F$. Therefore, for a process in which the system starts in an equilibrium state A and ends in another equilibrium state B , the free energy difference between the final and the initial state $\Delta F = F_B - F_A$ is the minimum amount of work that the external agent needs to perform to bring the system from the state A to the state B , $\Delta F = W_{\text{rev}}$. The *work dissipation* (or *dissipation*), W_{diss} , is defined as the excess work needed to perform a physical process,

$$W_{\text{diss}} = W - \Delta F. \quad (1.12)$$

The dissipation is always positive and zero if the process is reversible,

$$W_{\text{diss}} \geq 0. \quad (1.13)$$

Notice that in an isothermal process where the temperature of the thermal bath is T , dissipation equals to the entropy production times the temperature of the environment,

$$W_{\text{diss}} = TS_{\text{prod}}. \quad (1.14)$$

1.2 Maxwell's Demon and Szilard Engine

We have seen that the second law of thermodynamics imposes restrictions to some processes to occur spontaneously. For example, if we put in contact two gases at different temperatures, one at T_1 and the other at $T_2 < T_1$, the mixture of the gases will reach an intermediate temperature T_{fin} whose value is between the initial temperatures of the gases, $T_1 < T_{\text{fin}} < T_2$. The second law forbids that, after the mixture, the temperature of the colder gas decreases and the temperature of the hotter gas increases.

In 1871, Maxwell designed a thought experiment where this situation, forbidden by the second law, was possible (see Fig. 1.2a). Imagine that two gases at different temperatures are put in contact via an adiabatic wall that can be open or closed at any time. Maxwell suggested that a “demon”, i.e. an intelligent being able to see the molecules of each gas, could in principle produce a violation of the second law. Maxwell's demon controls the opening of the wall in such a way that only the

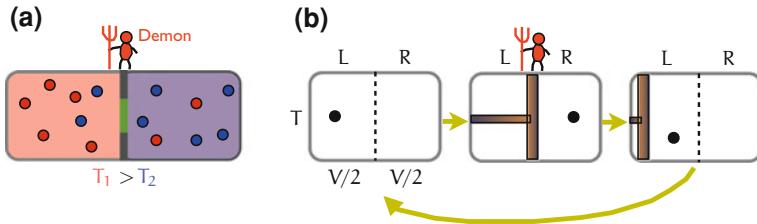


Fig. 1.2 Maxwell’s demon and the Szilard engine. **a** Illustration of Maxwell’s demon. Two gases are confined in separated containers and put in contact via an adiabatic wall. The gas on the *left* container is at a higher temperature T_1 than the gas on the *right* container, T_2 . A demon can see the individual molecules of the gas and selectively open or close the wall at any time. If the demon opens the door allowing only the slow molecules in the *left* container (*in blue*) to pass from *left* to *right*, and the fast molecules in the *right* container (*in red*) to pass from *right* to *left*, the temperature of the gas in the *left* container increases and the temperature of the gas in the *right* container decreases. **b** Illustration of the Szilard engine. A single molecule is confined in a container that is immersed in a thermal bath at temperature T . The particle moves randomly along the two different halves of the container, *L* (*left*) and *R* (*right*). A demon introduces a movable piston in the *middle* of the container and measures in which half of the container is the particle. Then the demon exerts a pressure on the piston equal and *opposite* to the pressure of the gas, making the gas expand reversibly

molecules that move slower than the average in the hotter container, and the ones that move faster than the average in the colder container, can pass through the wall. The result of the process is in principle forbidden by the second law: the colder gas gets colder and the hotter gets hotter.

A simplified version of Maxwell’s demon was introduced by Szilard [61] in another thought experiment (see Fig. 1.2b). *Szilard’s engine* is formed by a single particle confined in a container of volume V . The container is assumed to be immersed in a thermal reservoir of temperature T , therefore the particle thermalizes at temperature T after every collision with the walls of the container. After some time, a demon inserts a movable piston in the middle of the container and measures the position of the particle. Then the one particle gas is let to expand reversibly and the piston is removed [51].

The energetics of the Szilard engine reveals important aspects concerning the relationship between thermodynamics and information [46]. We can assume that the energy needed to insert the piston is negligible since the insertion of the piston can be done reversibly [51]. The only energetic contribution comes from the reversible expansion of the one particle gas against the piston from an initial volume $V/2$ to a final volume V . In a reversible expansion, the gas is described by the law of ideal gases throughout the process, $pV = kT$, and the expansion work is negative,

$$W = \int_{V/2}^V -pdV = -kT \int_{V/2}^V \frac{dV}{V} = -kT \ln 2. \quad (1.15)$$

Since the process is cyclic, $\Delta F = 0$, and therefore $W - \Delta F = -kT \ln 2 \simeq -0.69 kT < 0$, the particle is systematically extracting work cyclically from the thermal reservoir, which is in contradiction with the second law (1.11). Notice that the extraction of work is possible in the case of *heat engines*, where a system is put in contact with *two* thermal baths at different temperatures.

This apparent contradiction with the second law of thermodynamics was solved by Bennett [4], who investigated the energetics of the Szilard engine using Landauer's principle relating logical irreversibility and energy dissipation of computing processes [44]. Since the information is always stored in physical devices, the nature of information is physical. Because of this, simple logical operations have an energetic cost. In particular, the erasure of a bit is a logical irreversible operation, since a bit that is initially in one of the two possible states (0 or 1) is set to a reference state irrespective of its initial state. Landauer proved that the erasure of a bit requires *at least* $kT \ln 2$ of work dissipated to the thermal environment,

$$W_{\text{eras}} \geq kT \ln 2. \quad (1.16)$$

Where the bound is met in a quasistatic erasure. This statement is known as *Landauer's principle*. Landauer's principle was recently tested experimentally in [5] using optically-trapped Brownian particles as we discuss in Sect. 1.6.

Bennett reanalyzed the energetics of the Szilard engine taking into account Landauer's principle [4]. In every cycle, there are two contributions to the total work: First, the work done in the system is the expansion work of the one-molecule gas. Second, contrary to what Szilard suggested, the work needed to measure where the particle cannot be neglected. Since the measurement has only two possible outcomes (L and R), the measurement in the Szilard engine is equivalent to a measurement of the state of a bit. In every cycle, the state of the demon has to be restored to a reference state in order to be able to perform a measurement in the next cycle, which requires an amount of work that is at least $W_{\text{demon}} = W_{\text{eras}} \geq kT \ln 2$. The work per cycle of the Szilard engine is therefore

$$W_{\text{cycle}} = W_{\text{gas}} + W_{\text{demon}} \geq -kT \ln 2 + kT \ln 2 = 0, \quad (1.17)$$

which is in accordance with the second law of thermodynamics.

Notice that in the above proof we have assumed that a one particle gas behaves like a classical ideal gas. An alternative approach would be considering the particle as a quantum system. The “quantum version of the Szilard engine” was analyzed in [68], where Szilard's result $W_{\text{cycle}} = kT \ln 2$ as well as Landauer's principle are recovered in the quantum limit.

1.3 Review of Statistical Mechanics

The connection between the thermodynamic properties of microscopic systems and the physical properties of their microscopic constituents was done with the introduction of statistical mechanics. Before that, the first attempts to explain equilibrium properties of macroscopic systems from the molecular motion, were done in kinetic theory. Boltzmann's theory aimed to describe the equilibrium properties of a gas in terms of the probability to find a molecule with velocity v at a given position x at time t , $\rho(x, v, t)$. One of the most remarkable results of Boltzmann's formalism is the analytic expression of the velocity distribution of the molecules of an ideal gas in thermal equilibrium at temperature T . By assuming that there are only binary collisions between the molecules of the gas plus the celebrated *Stosszahlansatz* or molecular chaos hypothesis,¹ Boltzmann found an expression for the velocity distribution that is independent on the interaction between the molecules. In the simplest case, let us consider a gas in thermal equilibrium at temperature T formed by molecules of mass m . The gas is assumed to be in a one dimensional container. Under these assumptions, the probability to find a particle with velocity v at any position x and any time t is given by the *Maxwell-Boltzmann distribution*

$$\rho(v) = \sqrt{\frac{m}{2\pi kT}} e^{-mv^2/2kT}. \quad (1.18)$$

The probability distribution of the particle velocity is a Gaussian distribution with zero mean and dispersion

$$\langle v^2 \rangle = \frac{kT}{m}. \quad (1.19)$$

The above relationship is related to the *equipartition theorem*: In equilibrium, the average energy per degree of freedom of a physical system in contact with a thermal bath at temperature T is equal to $kT/2$. For an ideal gas in isothermal conditions, the only energetic contribution is kinetic. The average kinetic energy per molecule can be derived from (1.19),

$$\langle E \rangle = \left\langle \frac{mv^2}{2} \right\rangle = \frac{kT}{2}. \quad (1.20)$$

Statistical mechanics goes further than kinetic theory and describes not only the temperature, but also other thermodynamic quantities such as entropy or the free energy in terms of the statistics of the degrees of freedom of the molecular constituents. Equilibrium properties of not only dilute gases, but also of any macroscopic system can be studied within the framework of statistical mechanics [34].

¹ *Molecular chaos hypothesis* holds that at any binary collision between two molecules the distribution of position and momenta of the two particles after the collision are independent each other.

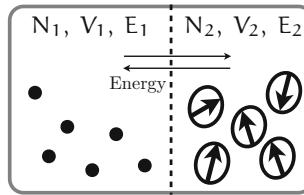


Fig. 1.3 Energy exchange between two isolated systems. Two different subsystems, a gas (subsystem 1) and a paramagnetic solid (subsystem 2) are isolated from the environment. The energy exchange is allowed between the two subsystems but the number of particle and the volume of the two subsystems does not change in time. Despite allowing energy exchange between the two subsystems, the total energy is constant $E = E_1 + E_2$

The first case studies of statistical mechanics are isolated systems. A physical system and together with the environment can be considered as a single bigger physical system (often called “universe”) whose total energy remains equal to a constant value E . The total number of particles N and the volume of the universe V are also considered as constant. The *macrostate* of the system is defined by the values of the thermodynamic parameters that characterize the state of the system, in this case (N, V, E) . The *microstate* of the system is given by the value of all the position and momenta of the molecules of the system, which in principle are three dimensional, $\{\vec{q}_i, \vec{p}_i\}$. The number of microstates that are compatible with a macrostate (N, V, E) is denoted by $\Omega(N, V, E)$. The following two postulates are the basis of statistical mechanics:

- **Equiprobability:** All the microstates compatible with a given macrostate (N, V, E) are equally probable.
- **Equilibrium:** The equilibrium state is such that the number of compatible microstates $\Omega(N, V, E)$ is maximum.

Let us consider the situation in which two macroscopic systems that might be of different nature (gases, paramagnetic solids, etc.) are put in contact via a wall that allows the exchange of energy between them such as the example depicted in Fig. 1.3. The first subsystem has N_1 particles, volume V_1 and its energy is labeled by E_1 , whereas the second subsystem is characterized by N_2, V_2 and E_2 . Since the total system is isolated, the total energy of the system does not change irrespective of the energy exchange, $E_1 + E_2 = E$. In this situation, the number of microstates of the total system compatible with a certain values of N, V and E is equal to

$$\Omega_T(N, V, E) = \Omega_1(N_1, V_1, E_1)\Omega_2(N_2, V_2, E_2), \quad (1.21)$$

since for every microstate of subsystem 1 there are Ω_2 microstates of 2 that are compatible with N, V, E . Because of the equilibrium postulate of statistical mechanics, in equilibrium Ω_T is maximum, or equivalently, $d\Omega_T(N, V, E)/dE = 0$. This condition, together with the constraint $E_1 + E_2 = E$ yields the following condition

$$\frac{\partial \ln \Omega_1(N_1, V_1, E_1)}{\partial E_1} = \frac{\partial \ln \Omega_2(N_2, V_2, E_2)}{\partial E_2}. \quad (1.22)$$

We know from thermodynamics that the physical parameter that is equal in the two subsystems when they are put in contact via a diathermic wall is the temperature. Therefore, the quantity $\frac{\partial \ln \Omega(N, V, E)}{\partial E}$ has to be related with temperature. More precisely, the definition of temperature in statistical mechanics is

$$\frac{\partial \ln \Omega(E)}{\partial E} \equiv \frac{1}{kT}. \quad (1.23)$$

From macroscopic thermodynamics it is also known [12] that $\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{N,V}$ where the subindex means at N and V constant. Equation (1.23) is consistent with the following definition of entropy,

$$S(N, V, E) = k \ln \Omega(N, V, E), \quad (1.24)$$

which was introduced by Boltzmann. The above result is one of the first quantitative relationships done between a macroscopic quantity (system entropy, S) with a microscopic property (the number of microstates, Ω) in statistical mechanics. Equation (1.24) shows that the entropy is larger when $\Omega(N, V, E)$ is larger, that is, when there are more states that are accessible by the system.

We notice that macroscopic quantities such as temperature are in general measured as a time average over a time window that is much larger than the atomic timescales. Statistical mechanics postulates that macroscopic parameters can also be obtained as averages over *ensembles* of physical systems. From now, we will denote by (q, p) all the position and momenta of the system, i.e., $(q, p) \equiv \{\vec{q}_i, \vec{p}_i\}_{i=1}^N$. We define by $\rho(q, p)$ the phase space density of the position of momenta of all the particles of the system, $i = 1, \dots, N$ to be $\{\vec{q}_i, \vec{p}_i\}_{i=1}^N$. $\rho(q, p)dqdp$ is the probability of the system at time t to be in $[p, p + dp]$ and $[q, q + dq]$. The ergodic hypothesis assumes that for any thermodynamic function A ,

$$\frac{1}{\Delta t} \int_t^{t+\Delta t} dt' A(t') = \int dqdp \rho(q, p) A(q, p). \quad (1.25)$$

The internal energy of the system can be obtained as the ensemble average of the Hamiltonian describing the system $H(q, p)$,

$$U = \langle H(q, p) \rangle = \int dqdp \rho(q, p) H(q, p). \quad (1.26)$$

Another particular case of thermodynamic function that can be expressed as an ensemble average is the system entropy. Gibbs proved that system entropy can be expressed as the average of the quantity $-k \ln \rho(q, p)$ [31],

$$S = \langle -k \ln \rho(q, p) \rangle = -k \int dq dp \rho(q, p) \ln \rho(q, p). \quad (1.27)$$

This result is consistent with Boltzmann's definition of entropy (1.24). Notice that the r.h.s. of (1.27) is proportional to a quantity that is commonly used in information theory *Shannon entropy* H , $H(\rho(x)) = \int dx \rho(x) \ln \rho(x)$, which is a measure of the uncertainty of the variable x distributed by $\rho(x)$ [18].

The time evolution of the phase space density $\rho = \rho(q, p; t)$ is governed by *Liouville's theorem* [53],

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \{\rho, H\} = 0, \quad (1.28)$$

where $\{\cdot, \cdot\}$ denotes the *Poisson bracket* and

$$\begin{aligned} \{\rho, H\} &= \sum_i \left(\frac{\partial \rho}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial \rho}{\partial p_i} \frac{\partial H}{\partial q_i} \right) \\ &= \sum_i \left(\frac{\partial \rho}{\partial q_i} \dot{q}_i + \frac{\partial \rho}{\partial p_i} \dot{p}_i \right). \end{aligned} \quad (1.29)$$

Here i denotes the sum over all the degrees of freedom of the system. Liouville's theorem implies that, for an observer that moves in time, the density of representative points in phase space is conserved. An ensemble of representative points in phase space move like an incompressible fluid in physical space [53]. An example of phase space densities that satisfy Liouville's theorem are those which depend on (q, p) via a Hamiltonian function $H(q, p)$, that is $\rho(q, p) = \rho(H(q, p))$. Depending on the boundary conditions (isothermal, isolated processes) the phase space density compatible with this constraint takes different forms as we now discuss.

For isolated systems described by a Hamiltonian $H(q, p)$, the phase space density is described by the *microcanonical ensemble*. Assuming that all the microstates that are compatible with a total energy E are equiprobable, the phase space density for all of them is proportional to

$$\rho(\{q_i, p_i\}) \propto \delta(H(q, p) - E). \quad (1.30)$$

In practical situations, the interactions between the system and the surroundings cannot be neglected, and isolated systems are only an ideal case that are useful for calculations. We now assume isothermal conditions, where a system exchanges energy with a thermal bath that does not change its temperature T even when exchanging energy with the system. In thermal equilibrium, the phase space density for the system is described by the *canonical distribution*,

$$\rho(q, p) = \frac{e^{-\beta H(q, p)}}{Z}. \quad (1.31)$$

where $\beta = 1/kT$, and the normalization constant

$$Z = \int dq dp e^{-\beta H(q, p)}, \quad (1.32)$$

is called the *partition function* of the system. The partition function contains all the thermodynamic information of the system. For example, the average energy of the system can be obtained from the partition function,

$$\langle H \rangle = -\frac{\partial \ln Z}{\partial \beta}. \quad (1.33)$$

A similar expression for the free energy in terms of the partition function can be found,

$$F = -kT \ln Z. \quad (1.34)$$

Statistical mechanics can be used to describe not only physical properties of macroscopic systems in equilibrium but also processes where the system is driven out of equilibrium [2, 21, 69]. The driving is in general modeled by the change of a control parameter λ according to a given protocol $\{\lambda(t)\}_{t=0}^{\tau}$ produced by the action of an external agent. In a physical process, the variation of the average energy of a system can be expressed as the sum of two terms

$$\delta \langle H \rangle = \int dq dp \delta H \rho(q, p; t) + \int dq dp H \delta \rho(q, p; t), \quad (1.35)$$

where the chain rule was used inside the ensemble average. The first term is the variation of the average energy due to the variation of the hamiltonian, which equals to the work done by the external agent. The second term only takes into account the variations of the phase space density, and it is equal to the heat transferred to the system. The work done by the external agent on a system defined by a Hamiltonian $H(q, p; \lambda)$ whose trajectory in phase space is $(q(t), p(t))$ and passes at time t through $(q(t), p(t)) = (q, p)$, $W(q, p; t)$, is given by

$$W(q, p; t) = \int_{\lambda(0)}^{\lambda(\tau)} \left. \frac{\partial H(q, p; \lambda)}{\partial \lambda} \right|_{q=q(t), p=p(t), \lambda=\lambda(t)} \cdot d\lambda(t). \quad (1.36)$$

If the system is in contact with a thermal bath at temperature T , and the process is isothermal and reversible, the probability distribution of the system is the canonical equilibrium distribution at any time $\rho(q, p; \lambda) = \frac{e^{-\beta H(q, p; \lambda)}}{Z_\lambda}$, where $Z_\lambda = \int dq dp e^{-\beta H(q, p; \lambda)}$. In this situation, the average work $\langle W \rangle = \int dq dp \rho(q, p; t) W(q, p; t)$ yields the expected result, $\langle W \rangle = \Delta F$.

1.4 Stochastic Thermodynamics I: Stochastic Energetics

As we have shown in the previous section, classical thermodynamics applies only to systems with a large number of degrees of freedom and establishes laws for macroscopic observables. When the size of the physical system under consideration is microscopic (or of smaller size) and the physical properties of the system can be sampled in a milisecond (or shorter) timescale, the dynamics of the system can be described with the equations of motion of few degrees of freedom. The paradigmatic example is the motion of a microscopic object (or *Brownian* particle) in a fluid at constant temperature. Because of the small size of the particle, the thermal fluctuations of the environment make the motion of the particle to be random or stochastic. Langevin equation [43] provides a mesoscopic² description of the dynamics of the Brownian particle by describing the thermal fluctuations of the environment with a stochastic force that takes into account the degrees of freedom of the thermal bath. Recently, it has been possible to extend the notions of heat and work to systems obeying a Langevin equation, and the first law [60] and the second law of thermodynamics [56] have been studied in the mesoscopic scale.

1.4.1 Brownian Motion and Langevin Equation

The *Brownian* particle is the paradigmatic case study of stochastic thermodynamics. In 1828, Brown observed in the microscope the motion of grains of pollen in water [7]. Brown was able to track the position of the grains, observing that their motion was apparently random. He checked this effect in different materials and claimed that this random motion had connection to life. In 1905, Einstein proved that the erratic motion of such colloidal particles is due to random impacts of water molecules of the environment, postulating that thermal fluctuations drive the erratic motion of Brownian particles [23]. He first developed a theory for an ensemble of Brownian particles moving in one dimension, showing that their average position vanishes at any time $\langle x(t) \rangle = 0$, and that the average mean square displacement depends linearly on time

$$\langle x^2(t) \rangle = 2Dt, \quad (1.37)$$

being D the *diffusion coefficient*. The above formula is the basis of the *diffusive* motion, in which, at odds with ballistic motion, the distance travelled by the particles does not scale with time but with the square root of time. Einstein showed that the value of the diffusion coefficient in equilibrium is

$$D = \frac{kT}{\gamma}, \quad (1.38)$$

² The *mesoscopic* description lies between micro and macroscopic description and consists on describing the motion of a microscopic system with few degrees of freedom.

where k is Boltzmann's constant, T is the temperature of the surrounding fluid and γ is the friction coefficient of the particle in the fluid. In the case of an spherical particle of radius R immersed in a fluid of viscosity η , the value of the friction coefficient for laminar flow (low Reynolds number) is given by Stokes' law $\gamma = 6\pi\eta R$ [3].

Einstein's relation (1.38) is a particular case of the *fluctuation-dissipation relationship* (FDR) that is introduced in linear response theory [41]. Suppose that a system described by a single degree of freedom x is weakly perturbed from equilibrium by a force that depends on time, $f(t)$, which is applied from $t = 0$ when the system is in equilibrium. If the system is not perturbed far away from equilibrium, the average value of x at any time t depends linearly on the applied force via the *susceptibility* χ_x ,

$$\langle x(t) \rangle = \int_0^t dt' \chi_x(t - t') f(t'). \quad (1.39)$$

The fluctuations of the variable x in time can be measured by the autocorrelation function,

$$C_x(t) = \langle x(t)x(0) \rangle. \quad (1.40)$$

Close to equilibrium, the following relationship between the Fourier transform of the autocorrelation and the imaginary part of the susceptibility holds

$$\tilde{C}_x(\omega) = \frac{2kT}{\omega} \text{Im } \tilde{\chi}(\omega), \quad (1.41)$$

which is known as the fluctuation-dissipation relationship [41]. In linear response, the FDR can be used as a test to check whether or not a system is equilibrium with the thermal environment. Einstein's relation (1.38) can be derived from the FDR for a free Brownian particle described by considering the equation of motion of the position of the Brownian particle [41].

A theory of the motion of an individual free Brownian particle was first done by Langevin [43]. Langevin described the dynamics of the free Brownian particle of mass m starting with usual Newtonian equations of motion for the position $x(t)$ and momentum $p(t)$ of the particle. He expressed the total force that the Brownian particle feels as the sum of two terms. The first is the viscous force that the surrounding fluid exerts on the particle which is assumed to be in accordance with Stokes' law $F(t) = -\gamma v(t)$, being $v(t) = p(t)/m$ the instantaneous velocity of the particle. The second component of the force is a stochastic term $\xi(t)$ that plays the role of the fluctuating force exerted to the Brownian particle by the molecules of the environment. By considering these two forces, the equations of motion of the Brownian particle are

$$\frac{dp}{dt} = -\gamma \frac{p}{m} + \xi(t), \quad (1.42)$$

$$\frac{dx}{dt} = \frac{p}{m}. \quad (1.43)$$

Here $\xi(t)$ is a *Gaussian white noise* (see Appendix A.1) of zero mean and variance that depends on the diffusion coefficient, that is, a random force that is Gaussian distributed, its average vanishes and it is delta correlated with intensity given by (1.38),

$$\langle \xi(t) \rangle = 0, \quad (1.44)$$

$$\langle \xi(t)\xi(t') \rangle = 2D\delta(t-t'), \quad (1.45)$$

with $D = kT/\gamma$. Notice that Langevin equation is valid when all the degrees of freedom (position and momentum) that describe the surrounding fluid are rapidly varying degrees of freedom in comparison with the position and momentum of the colloidal particle. This assumption is true when the mass of the Brownian particle is several orders of magnitude larger than the mass of the molecules of the environment. We notice that Langevin's description is mesoscopic since it lies between the macroscopic approach and a microscopic description containing all the degrees of freedom of the Brownian particle and the surrounding fluid.

If the time resolution in which one measures the position of the Brownian particle Δt is much larger than the momentum relaxation time $\Delta t \gg \tau_p = m/\gamma$, Eqs. (1.42, 1.43) can be replaced by the following equation

$$\gamma \dot{x} = \xi(t), \quad (1.46)$$

where $\xi(t)$ is the same as in the underdamped case. The above equation is known as the *overdamped Langevin equation* for a free Brownian particle. It can be obtained as the overdamped limit of Eq. (1.42), that is, when inertia is negligible.

A typical example of a system that obeys Langevin equation is a microscopic sphere of polystyrene immersed in water. For a spherical polystyrene particle of radius $R \sim 1 \mu\text{m}$, the mass of the particle m is of order of picograms and the friction coefficient in water ($\eta = 0.89 \text{ mPas}$), assuming Stokes' law, is around $\gamma = 6\pi\eta R \sim 10^{-9} \frac{\text{Ns}}{\text{m}}$. This yields a characteristic time $\tau_m \sim 1 \mu\text{s}$ which is much shorter than the usual experimental time resolution $\Delta t \sim \text{ms}$, being the overdamped limit valid in this case.

If additional external forces are applied to the Brownian particle, an additional term modeling the external forces has to be taken into account in Langevin's model. If the external force is controlled via a control parameter $\lambda(t)$ and it is caused by a potential energy $V(x, \lambda)$, the underdamped Langevin equations for the particle are

$$\frac{dp}{dt} = -\frac{\partial V(x, \lambda)}{\partial x} - \gamma \frac{p}{m} + \xi(t), \quad (1.47)$$

$$\frac{dx}{dt} = \frac{p}{m}. \quad (1.48)$$

Notice that in general, the total external force applied on the particle $F(x, \lambda)$ has two contributions: Firstly, a term that derives from a conservative potential $\frac{-\partial V(x, \lambda)}{\partial x}$, and

secondly, a term that models the application of forces directly on the particle $f(x, \lambda)$. Therefore, in general, we should take into account both contributions and replace $-\frac{\partial V(x, \lambda)}{\partial x}$ by $F(x, \lambda) = -\frac{\partial V(x, \lambda)}{\partial x} + f(x, \lambda)$. In the overdamped limit, Eqs. (1.47, 1.48) yield the overdamped Langevin equation in the presence of external conservative forces is

$$\gamma \frac{dx}{dt} = -\frac{\partial V(x, \lambda)}{\partial x} + \xi(t). \quad (1.49)$$

This equation describes the random motion of colloidal microscopic particles but it can also be applied to magnetic or electric systems (see e.g. [16]). Langevin equation has also been applied to model the dynamics of molecular motors [9, 39].

1.4.2 First Law of Thermodynamics in the Mesoscopic Scale

In a pioneering work, Ken Sekimoto extended the notions of heat and work to thermodynamic processes affecting small systems described by Langevin equation, providing a new framework to measure the energetics of processes that occur at the mesoscale [59]. Let us consider Langevin equation as a force balance equation in which the sum of all forces acting on the particle are zero,

$$0 = -\frac{\partial V(x, \lambda)}{\partial x} - \gamma \frac{dx}{dt} + \xi(t). \quad (1.50)$$

In the force balance equation, there are two forces that are exerted by the environment on the particle, the friction force $-\gamma \frac{dx}{dt}$ and the force of the environment $\xi(t)$. Therefore $(-\gamma dx/dt + \xi(t)) \circ dx(t)$ is the energy transfer from the environment to the particle, that is, the heat transferred to the particle. Notice that we have used the *Stratonovich product* \circ (see Appendix A.2) for reasons that we explain below [60]. We therefore define

$$\delta Q \equiv \left(-\gamma \frac{dx}{dt} + \xi(t) \right) \circ dx(t), \quad (1.51)$$

as the heat transferred to an overdamped Brownian particle when it is displaced a differential amount $dx(t)$. In the underdamped limit, the above result can be expressed as

$$\delta Q = \left(-\gamma \frac{p}{m} + \xi(t) \right) \circ dx(t) = \left(\frac{dp}{dt} + \frac{\partial V(x, \lambda)}{\partial x} \right) \circ dx(t), \quad (1.52)$$

where we have used Eq. (1.47) in the second equality. We now focus on the energy balance of the particle. First, we notice that, since we use the Stratonovich product the usual rules of calculus apply, and we can express the two terms of $\delta Q = \left(\frac{dp}{dt} + \frac{\partial V(x, \lambda)}{\partial x} \right) \circ dx(t)$ in the underdamped limit as follows

$$\frac{dp}{dt} \circ dx(t) = \frac{dp}{dt} \circ \frac{p}{m} dt = d\left(\frac{p^2}{2m}\right), \quad (1.53)$$

$$\frac{\partial V(x, \lambda)}{\partial x} \circ dx(t) = dV(x, \lambda) - \frac{\partial V(x, \lambda)}{\partial \lambda}. \quad (1.54)$$

By summing Eqs. (1.53) and (1.54), we obtain

$$\delta Q = d\left(\frac{p^2}{2m} + V(x, \lambda)\right) - \frac{\partial V(x, \lambda)}{\partial \lambda} d\lambda. \quad (1.55)$$

The first term in (1.55) is the internal energy of the particle (kinetic and potential) $U(x, \lambda) = \frac{p^2}{2m} + V(x, \lambda)$. Therefore, by defining the second term in (1.55) as the work done by the external agent on the particle when it moves a small amount $dx(t)$,

$$\delta W \equiv \frac{\partial V(x, \lambda)}{\partial \lambda} d\lambda, \quad (1.56)$$

we recover the first law of thermodynamics (1.1) in the microscopic scale $dU = \delta Q + \delta W$. We notice that in the overdamped limit, the heat (1.51) can be expressed as

$$\delta Q \equiv \frac{\partial V(x, \lambda)}{\partial x} \circ dx(t). \quad (1.57)$$

In this case, the first law is equivalent to a chain rule for the internal energy, since $dU = (\partial U / \partial \lambda) \circ d\lambda + (\partial U / \partial x) \circ dx(t)$ where the first term can be identified as the work and the second as the heat.

The path dependence of the work and heat is now revealed in stochastic thermodynamics. The total work done by the particle and the total heat transferred to the particle depend on the specific stochastic trajectory and they are expressed as functionals of the microscopic trajectory. The total heat and work transferred from $t = 0$ to $t = \tau$ when the external agent controls the state of the system via $\lambda(t)$ and the trajectory of the particle is $x_\tau \equiv \{x(t)\}_{t=0}^\tau$ is obtained by integrating Eqs. (1.51, 1.56) with respect to time,

$$\mathcal{Q}[x_\tau] = \int_{x(0)}^{x(\tau)} \frac{\partial V(x, \lambda)}{\partial x} \Big|_{(x, \lambda) = (x(t), \lambda(t))} \circ dx(t), \quad (1.58)$$

$$W[x_\tau] = \int_{\lambda(0)}^{\lambda(\tau)} \frac{\partial V(x, \lambda)}{\partial \lambda} \Big|_{(x, \lambda) = (x(t), \lambda(t))} \circ d\lambda(t). \quad (1.59)$$

Notice that the expression for the work (1.59) is analogous to the general formula of the work given in statistical mechanics (1.36).

For a Brownian particle described by the overdamped Langevin equation, the free energy change in an isothermal process is equal to the work done by the external

agent when the process is done quasistatically, i.e. when the system is at any time described by the canonical ensemble,

$$\Delta F = \int_{\lambda(0)}^{\lambda(\tau)} \left\langle \frac{\partial V(x, \lambda)}{\partial \lambda} \right\rangle_{\text{eq}} \circ d\lambda(t). \quad (1.60)$$

As it is proved in [60], the stochastic definition of work is consistent with the usual statement of the second law of thermodynamics when considering the ensemble average (1.11) $\langle W \rangle - \Delta F \geq 0$.

1.4.3 Second Law of Thermodynamics in the Mesoscopic Scale

The generalization of the second law of thermodynamics to the mesoscopic scale can be done by extending the notion of entropy to systems where thermal fluctuations are relevant. Seifert [56] introduced the “trajectory entropy” for systems obeying an overdamped Langevin equation. His definition was previously proposed by Crooks [20] for stochastic Markovian systems [32] obeying the microscopic reversible condition, and it was extended to colloidal particles in the overdamped limit and discrete systems described by a master equation in [56]. Let us consider a colloidal particle that moves in one dimension x from $t = 0$ to $t = \tau$. We assume that the particle is subject to a total external force $F(x, \lambda)$. In the overdamped limit, Langevin equation describes the motion of the particle,

$$\gamma \dot{x} = F(x, \lambda) + \xi, \quad (1.61)$$

where $\langle \xi(t)\xi(t') \rangle = 2\gamma kT\delta(t-t')$. We define the mobility by $\mu \equiv 1/\gamma$, and the diffusion coefficient is $D = kT/\gamma$. An equivalent description of the dynamics is given by the *Fokker-Planck equation* (see Appendix A.3)

$$\partial_t \rho(x, t) = -\partial_x j(x, t) = -\partial_x (\mu F(x, \lambda) - D\partial_x) \rho(x, t), \quad (1.62)$$

where $\rho(x, t)$ is the probability of the particle to be at x at time t . The Shannon entropy associated to the position of the particle at time t can be interpreted as a nonequilibrium entropy:

$$S(t) = -k \int dx(t) \rho(x(t), t) \ln \rho(x(t), t), \quad (1.63)$$

which suggests the introduction of the following trajectory dependent entropy

$$S[x(t)] = -k \ln \rho(x(t), t). \quad (1.64)$$

With this definition, the Shannon entropy at any time t can be seen as the ensemble average of the trajectory (nonequilibrium) entropy, $S(t) = \langle S[x(t)] \rangle$, where the average is done over all trajectories weighed with $\rho(x(t), t)$.

Using the trajectory dependent entropy, one can recover a second law of thermodynamics in the mesoscopic scale. Let us consider the entropy production of the system, i.e. $\dot{S}[x(t)]$. The time derivative is [8]

$$\begin{aligned} \frac{\dot{S}[x(t)]}{k} &= -\left. \frac{\partial_t \rho(x, t)}{\rho(x, t)} \right|_{x(t)} - \left. \frac{\partial_x \rho(x, t)}{\rho(x, t)} \right|_{x(t)} \dot{x} \\ &= -\left. \frac{\partial_t \rho(x, t)}{\rho(x, t)} \right|_{x(t)} + \left. \frac{j(x, t)}{D\rho(x, t)} \right|_{x(t)} \dot{x} \\ &\quad - \left. \frac{F(x, \lambda)}{kT} \right|_{x(t)} \dot{x}, \end{aligned} \quad (1.65)$$

where we have used (1.62) in the second equality. The last term in (1.65) can be identified as an entropy change in the reservoir $\dot{S}_{\text{res}} = -\dot{Q}/T$, being Q calculated following (1.51). Since $-\gamma \dot{x} + \xi = -F(x, \lambda)$ the heat transferred to the particle is $\dot{Q}/T = -\left. \frac{F(x, \lambda)}{kT} \right|_{x(t)} \dot{x}$ and the entropy change in the reservoir $\dot{S}_{\text{res}} = \left. \frac{F(x, \lambda)}{kT} \right|_{x(t)} \dot{x}$.

The entropy production rate is (1.6)

$$\dot{S}_{\text{prod}} = \dot{S} + \dot{S}_{\text{res}} = k \left. \frac{\partial_t \rho(x, t)}{\rho(x, t)} \right|_{x(t)} + k \left. \frac{j(x, t)}{D\rho(x, t)} \right|_{x(t)} \dot{x}. \quad (1.66)$$

When averaging over $\rho(x, t)$, the first term in (1.66) vanishes, yielding [56]

$$\langle \dot{S}_{\text{prod}}(t) \rangle = k \int dx \frac{j(x, t)^2}{D\rho(x, t)} \geq 0. \quad (1.67)$$

Equation (1.67) is equivalent to the relationship between entropy and currents derived in linear irreversible thermodynamics [40]. For a physical system obeying the over-damped Langevin equation, the entropy production is a positively defined quadratic form of the currents, as that obtained in linear irreversible thermodynamics for non-equilibrium processes in linear regime (see Sect. 2.2.1). The equality in (1.67) is obtained in equilibrium, where the current vanishes $j(x, t) = 0$.

We have seen that thermodynamic laws hold when considering averages of thermodynamic functions over many realizations of a process. Because of the importance of thermal fluctuations at the microscopic scale, fluctuations can produce deviations from the mean behavior. One can for example observe negative entropy production in a single experiment of a microscopic system, such as stretching a DNA molecule. The introduction of the *fluctuation theorems* has provided a theoretical framework to measure the probability of the deviations from the mean behavior in microscopic systems.

1.5 Stochastic Thermodynamics II: Fluctuation Theorems

In mesoscopic systems, the importance of fluctuations can lead to rare events such as a negative entropy production—or negative dissipation—in a single realization. As it was seen in the previous section, thermodynamic laws only apply to ensemble averages in the mesoscopic scale. It is interesting to know the probability of thermodynamic functions to take values that are different from the average value. *Fluctuation theorems* are exact relationships that express universal properties of the probability distribution $\rho(\Omega)$ for thermodynamic functionals of the position $\Omega[x(t)]$ such as work, heat or entropy [58]. Unlike the first results of nonequilibrium statistical physics that are valid only in near-equilibrium conditions (small external forces), fluctuation theorems are valid for systems that are perturbed arbitrarily far from equilibrium. During the last decade, several fluctuation theorems have been derived and tested in simulations and experiments. We review some of the most relevant results have been obtained in the context of fluctuation theorems in the last decade.

1.5.1 Jarzynski's Equality

One of the first results concerning the probability distributions of thermodynamic functions in nonequilibrium processes was introduced by Jarzynski in [35]. Let us consider the following process: A system is initially in an canonical equilibrium state A . An external agent drives the system arbitrarily far from equilibrium by modifying a control parameter λ according to a given protocol $\lambda(t)$. At the end of the process, the system is allowed to relax to a final canonical equilibrium state B . In both isolated or isothermal conditions, the work done by the external agent in this process is stochastic and the average over all the possible realizations satisfies the following *equality*

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}. \quad (1.68)$$

Notice that, since e^x satisfies Jensen's inequality, $\langle e^x \rangle \geq e^{\langle x \rangle}$ since the exponential is a convex function.³ As a result, Jarzynski's equality implies the second law $\langle W \rangle \geq \Delta F$. One of the most important implications of Jarzynski's equality is that we can estimate an equilibrium property, the free energy difference between two equilibrium states, from a nonequilibrium average,

$$\Delta F = -kT \ln \langle e^{-\beta W} \rangle. \quad (1.69)$$

This result is valid for any protocol followed by the control parameter and it holds for isolated systems as well as for systems weakly coupled to a single thermal bath. The equality has been proved analytically [35], in simulations [52] and tested experimentally [17] as we will discuss in Sect. 1.6.

³ A convex function $f(x)$ is such that for any $x_1 \neq x_2$, $\frac{1}{2}(f(x_1) + f(x_2)) \geq f(\frac{x_1+x_2}{2})$.

1.5.2 Crooks's Theorem

Let us consider a system that is driven out of equilibrium in a finite time process defined by the evolution of a control parameter $\lambda(t)$ that changes from $t = 0$ to $t = \tau$. We call this process *forward process* and denote it by F . Let us now consider the time-reversed or *backward* (B) process in which the system is driven in the time-reversed manner $\tilde{\lambda}(t) = \lambda(\tau - t)$. We also assume that the system is in contact with a single thermal bath of temperature T during the process. The initial and final states of the system are assumed to be equilibrium states in both forward and backward processes. We notice that the initial (equilibrium) state in the backward process is obtained from the same distribution as the final (equilibrium) state of the forward process, and vice versa. In [20], Crooks proved that the following fluctuation theorem holds for any nonequilibrium driving,

$$\frac{\rho_F(W)}{\rho_B(-W)} = e^{\beta(W - \Delta F)}. \quad (1.70)$$

Here $\rho_F(W)$ is the work distribution of the forward process and $\rho_B(-W)$ is the distribution of $-W$ in the reverse process. This result was originally proved for processes where the dynamics is Markovian [32] and the entropy production is odd under time reversal. Such condition is met when the dynamics is *microscopically reversible* [19] which is equivalent to say

$$\frac{\mathcal{P}_F[\{x(t)\}_{t=0}^\tau]}{\mathcal{P}_B[\{x(\tau - t)\}_{t=0}^\tau]} = e^{-\beta Q[\{x(t)\}_{t=0}^\tau]}, \quad (1.71)$$

where $Q[\{x(t)\}_{t=0}^\tau]$ is the heat transferred to the system from the reservoir when the trajectory of the system is $\{x(t)\}_{t=0}^\tau$. Systems described by a Langevin equation or Metropolis Monte Carlo satisfy all these conditions. However, the condition of the entropy production to be odd under time reversal is more restrictive, but it is valid for systems that are initially in a thermal equilibrium state, and for systems that are in a time symmetric nonequilibrium steady state (e.g. a gas in a piston that is driven sinusoidally or a fluid under a constant shear). In a more recent work, the validity of Crooks's fluctuation theorem has also been proved for an example of an isolated system following Hamiltonian dynamics provided that the initial condition of both forward and backward processes is canonical at temperature T [15]. Crooks's fluctuation theorem has also been tested experimentally using RNA hairpins as we discuss in Sect. 1.6.

We remark that Jarzynski's equality is direct consequence of Crooks's theorem. The average $\langle e^{-\beta W} \rangle$ reads,

$$\langle e^{-\beta W} \rangle = \int dW \rho_F(W) e^{-\beta W} = \int dW \rho_B(-W) e^{-\beta \Delta F} = e^{-\beta \Delta F}, \quad (1.72)$$

where the second equality is consequence of (1.70). Like Jarzynski's equality, Crooks theorem can be used to estimate the free energy difference between two equilibrium states if the system starts and ends in two different equilibrium states. The estimation can be done from the intersection of the work distributions of any forward and backward nonequilibrium process that drive the system from state A to state B ,

$$\frac{\rho_F(W^*)}{\rho_B(-W^*)} = 1 \Rightarrow W^* = \Delta F. \quad (1.73)$$

1.5.3 Steady State and Transient Fluctuation Theorems

Let us now consider a process where a microscopic system is driven to a nonequilibrium steady state (NESS) due to nonequilibrium constraints. One example is a system in contact with two thermal baths at different temperatures, such as a metallic bar with its two ends connected to baths at different temperatures. After some transient, the bar reaches a NESS where heat flows through the bar (it is out of equilibrium) from the hot to the cold reservoir and the temperature distribution does not change in time (it is stationary). Another example is a fluid between two parallel walls which move relative to each other at constant velocity and produce a constant shear on the fluid. In such processes, one cannot define a protocol $\lambda(t)$ since the nonequilibrium constraints do not change along the process. In the case of a fluid under constant shear, the NESS ensemble of the system is identical to the ensemble that is reached with the time reversed process but with a simple reflection of momenta.

In the NESS, the entropy production of the system (and also the total entropy production) is odd under time reversal and Crooks's theorem is satisfied. Since forward and time reversal processes are identical, we can drop off the index F, B and use ρ for distributions of the process in the NESS. Let us consider the observable defined by the time average of the entropy production rate \dot{S}_{prod} in the interval $t \in [0, \tau]$ given by $S_\tau \equiv \frac{1}{\tau} \int_0^\tau \dot{S}_{\text{prod}}(t) dt$, where $t = 0$ denotes an arbitrary initial time of an interval of duration τ in the steady state. In the NESS, the following fluctuation theorem holds

$$\ln \left(\frac{\rho_\tau(\dot{S})}{\rho_\tau(-\dot{S})} \right) = \frac{\dot{S}\tau}{k}, \quad (1.74)$$

where $\rho_\tau(\dot{S})$ is the probability to observe a time average of the entropy production rate (in the stationary state) of value $S_\tau = \dot{S}$. Equation (1.74) is known as the *steady state fluctuation theorem* (SSFT) for entropy production. The SSFT is often called the *Gallavotti-Cohen theorem* since it was first formally proved by Gallavotti and Cohen for fluids under shear stress in which the molecular chaos hypothesis holds [29].

Entropy consumption in small systems was first observed in numerical simulations of a hard-disk fluid under a constant shear stress [25]. The average entropy production from $t = 0$ to τ was found to be negative in several (but not the majority) of

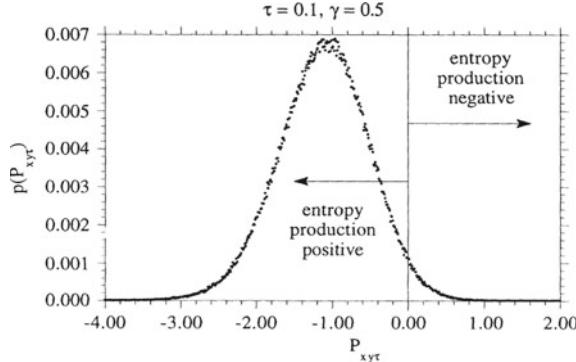


Fig. 1.4 Entropy fluctuations in a hard-disk fluid under shear stress. The probability distribution of averaged stress tensor of a hard-disk fluid under shear stress obtained from simulations in [25] is plotted as a function of the value of the stress tensor. The *vertical line* separates the data where entropy was produced or consumed by the system. Picture taken from [25]

the realizations of the model, for a suitable short time τ (see Fig. 1.4). Therefore, despite observing entropy consuming trajectories, the second law of thermodynamics remains valid in the microscopic scale when averaging to all the possible trajectories of the system.

Steady-state fluctuation theorem is only true in the long time limit [26]. When τ is not very large compared to the transient time of the system, the left hand side of (1.74) is also a linear function of the entropy but the slope is not equal to $1/k$. In the long time limit, (1.74) can be rewritten using the entropy production rate, defined by $\dot{S} \equiv \lim_{\tau \rightarrow \infty} S_\tau = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau \dot{S}_{\text{prod}}(t) dt$,

$$\frac{\dot{S}}{k} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \ln \left(\frac{\rho_\tau(S)}{\rho_\tau(-S)} \right), \quad (1.75)$$

where $\rho_\tau(S)$ is the probability to observe an entropy production of value $S_{\text{prod}} = S$ in the steady state in the interval $t \in [0, \tau]$. Equation (1.75) is obtained as the limit of (1.74) when $\tau \rightarrow \infty$. The SSFT (1.75) is valid in both linear and non linear regime [55] and also applies for systems obeying a Langevin equation [42], for more general stochastic dynamics [48] and for general Markov processes [45]. In [64, 65] the SSFT for both work and heat is found to be true for long τ for dragged Brownian particles that are described by the overdamped Langevin equation.

Notice that the SSFT predicts that in the NESS it is possible to observe negative entropy production in finite time processes as well as in the long time limit. However, the probability to observe a negative entropy production, $-S$, is exponentially smaller than the probability to observe a entropy production S , since $\rho(-S) = \rho(S)e^{-S/k}$. The second law of thermodynamics is recovered when averaging to the trajectory ensemble,

$$\frac{\langle \dot{S} \rangle}{k} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \left\langle \ln \frac{\rho_\tau(S)}{\rho_\tau(-S)} \right\rangle, \quad (1.76)$$

where the right hand side is positive as we will see in Chap. 2, which yields the second law $\langle \dot{S} \rangle \geq 0$. Equation (1.76) also provides a connection between the average entropy production (\dot{S}) and a measure of time irreversibility of the process, as we will study in Chap. 2.

Notice that a *transient fluctuation theorem* (TFT) was introduced by Evans and Searles [26, 55] by proving the same result as (1.74) but considering probability distributions of $S_\tau \equiv \frac{1}{\tau} \int_0^\tau \dot{S}_{\text{prod}}(t) dt$, $t = 0$ being the initial time of the process and obtaining $\rho(S_\tau)$ from an ensemble average over many trajectories. In [26, 55] the TFT is proved for finite τ for isolated systems initially described by an equilibrium microcanonical ensemble that evolve towards a stationary state. Also in [64, 65], the TFT for work and heat transferred to an overdamped Brownian particle is proved analytically to be true for any τ .

We remark that, unlike entropy production, system and reservoir entropy changes along a nonequilibrium process do not satisfy a transient or steady state fluctuation theorem in general. On the other hand, the total entropy production can be decomposed as $S_{\text{prod}} = \Delta S_{\text{na}} + \Delta S_{\text{a}}$, where ΔS_{na} and ΔS_{a} are called the nonadiabatic and adiabatic components of the entropy production, respectively, and they both separately satisfy a TFT and a SSFT [24].

1.6 Experiments

The development of microscopy, micro-manipulation and optical trapping techniques has allowed to test theoretical results of the thermodynamics of small systems in different scenarios [11]. We now review some of these experiments, in particular those who concern the implementation of Szilard's engine, the verification of Landauer's principle and the test of fluctuation theorems.

Optical traps are the main experimental device used in the majority small-scale experiments, since they can be used to trap a single microscopic particle (or even a molecule) and to manipulate it, measuring forces of the order of piconewtons and distances of the order of nanometers. *Optical tweezers* were first developed by Ashkin in the 80s. In 1986, Ashkin et al. were able to trap dielectric particles in water with particle diameters ranging from 25 nm to 10 μm [1]. Their experimental setup consisted on a single highly focused argon-laser beam (wavelength 541 nm) as sketched in Fig. 1.5a that was used to trap dielectric beads immersed in a flow cell that contained water. A microscope was added to the setup to observe in real time the trapped particles as well as the diffraction pattern of the light. The trapping mechanism can be explained rigorously using Mie's scattering theory, which describes the interaction between an electromagnetic plane wave and a dielectric sphere of a size comparable to the wavelength of the electromagnetic wave [50]. Using ray optics we can describe the scattering and momentum transfer to the particle [1] as we show in

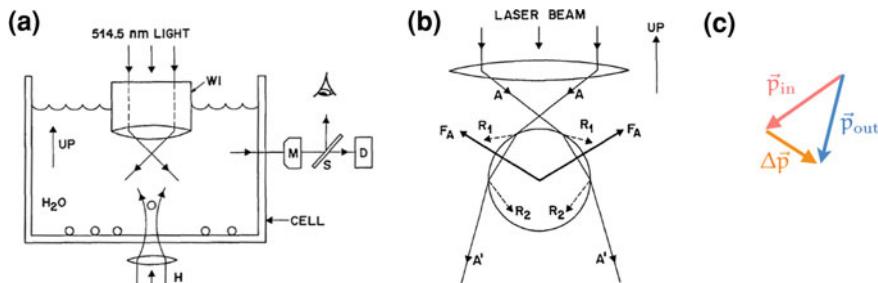


Fig. 1.5 First single-beam optical trap. **a** Sketch of the experimental setup. A 514.5 nm argon laser light passes through a microscope that focuses the light in a cell where microscopic dielectric beads are immersed in water. A microscope (M) allows to observe the trapped particles. **b** Ray optics of a spherical particle that is slightly after the optical focus of the microscope. The change in momentum of the light is indicated in the figure, as well as the direction of the reaction force of the bead (F_A). The net gradient force points to the focus. **c** Momentum of the incident \vec{p}_{in} and scattered \vec{p}_{out} light and momentum change in the light $\Delta \vec{p}$ corresponding to the ray optics sketch of **(b)**. Pictures taken from [1]

Fig. 1.6 Optical forces in an optical tweezer. **a** When the bead deviates from the propagation axis, the intensity profile is such that the bead pushes the light to the *left* and the light pushes the bead to the *right*, that is, through the propagation axis. **b** The gradient force pushes the bead towards the focus of the microscope

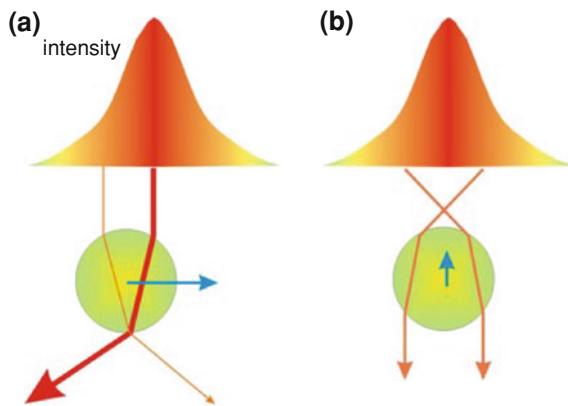


Fig. 1.5b, c. Because of third Newton's law, the force exerted to the bead is opposite to the momentum change in the light. The net force in Fig. 1.5b points to the focus of the microscope, that is, in the direction of the gradient of intensity. This *gradient force* is compensated by the *scattering force* in the equilibrium position of the particle in the trap, which lies slightly after the focus of the microscope. On the other hand, the intensity profile of the laser is generally Gaussian. Because of this, when the particle moves perpendicular to the propagation axis, the light exerts a force on the bead pushing the bead towards the propagation axis as shown in Fig. 1.6. As a result, the particles are trapped in three dimensions by the optical tweezer as sketched in Fig. 1.6. The optical force exerted on the particle is Hookean (linear) for small deviations of the bead from the trap center, \vec{x}_0 , in any direction, $F_i = \kappa_i(x_i - x_{0,i})$, for $i = x, y, z$ and κ_i being the stiffness of the trap measured in the i axis.

The development of trapping techniques has allowed to build the first experimental microscopic-sized heat engine formed by a single Brownian particle trapped with a single optical tweezer [6]. The temperature of the environment is switched between two different values using laser heating, which makes the Brownian particle to work between two different thermal baths, and therefore as a heat engine.

In a different experiment, a Szilard engine formed by a single Brownian particle was designed [62]. A microscopic particle is made to climb up a spiral-staircase-like potential created by an electric field using feedback control. A dimeric particle formed by two polystyrene spherical beads (both with diameter $D = 287\text{ nm}$) is attached to the top glass surface of a fluid chamber. One of the particles of the dimer is held fixed and linked to the top surface of the chamber and the other performs rotational Brownian motion. Using elliptically rotating electric fields, they managed to induce a tilted periodic potential to the particle. A feedback control loop of total time $t = \tau$ is implemented as follows: At $t = 0$, the angular position of the particle is measured. If the position lies within a specific interval, the phase of the potential is inverted at $t = \epsilon$ and the particle is let to relax to a new equilibrium position until $t = \tau$ in which the feedback cycle is repeated. When ϵ is comparable to the relaxation time of the particle in the periodic potential, the particle is made to climb upwards in the energy landscape, as sketched in Fig. 1.7a.

Landauer's principle has been experimentally verified recently using optically trapped microscopic particles [5]. A single silica spheric bead (diameter $D = 2\text{ }\mu\text{m}$) immersed in water was trapped with optical tweezers. In their experiment, a bistable potential is created by switching the trapping laser position between two positions at a fast rate (faster than the relaxation time of the bead on the trap, τ_r). By modulating the laser intensity, Berut et al. were able to modulate the height of the barrier that separates the two minima of the double well potential. Using this setup, they implemented a restore-to-zero cyclic process (shown in Fig. 1.7b) in which the bead is initially in one of the two equilibrium positions. Then, the barrier is lowered in 1s ($\gg \tau_r$) and the barrier is kept low for a time τ during which an external force that increases with time is applied. The cycle is finished lowering the barrier to its original height during 1s. In [5] they measure the stochastic heat using (1.58) and show that in the limit of slow process ($\tau \simeq 40\text{ s}$), Landauer's bound (1.16) is experimentally checked, yielding $\langle Q \rangle \simeq kT \ln 2$ within experimental errors.

Several single-molecule experiments using DNA hairpins of microscopic colloidal systems have allowed to measure deviations from the second law of thermodynamics in small systems and to test fluctuation theorems in different scenarios [11]. We now review some of the most relevant experiments done to test some of the fluctuation theorems introduced in Sect. 1.5.

- **Jarzynski's equality:** In Ref. [47], Liphardt et al. tested Jarzynski's equality in a single molecule experiment using a RNA hairpin. They mechanically stretched a RNA molecule between two different conformations. One of the ends of the RNA molecule was attached to a bead that was held fixed with an optical tweezer whereas the other end was attached to a second bead whose position could be controlled using a micropipette. Using direct measurements of the force exerted to

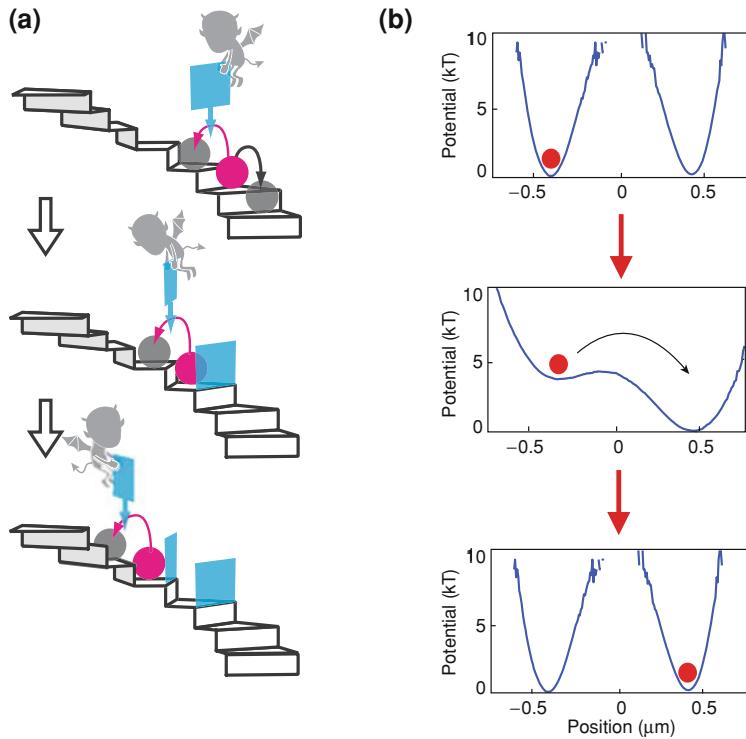


Fig. 1.7 Experimental Szilard engine and experimental test of Landauer’s principle. **a** Sketch of the feedback control of Toyabe et al. experiment used to mimic a microscopic Szilard engine. **b** Restore-to-zero protocol of Berut et al. experiment done to test Landauer’s principle. Pictures taken from [62] and [5]

the hairpin, they measured the work done on the molecule in fast (nonequilibrium) extensions of the molecule (1.59). They estimated the free energy difference from the averaged exponential work over nonequilibrium experiments (1.69), $\Delta F_{JE} = -kT \ln\langle e^{-\beta W}\rangle$, finding that this their estimation (based on Jarzynski’s equality) did not significantly differ from the actual equilibrium value of ΔF measured in a reversible (slow) stretching.

- **Crooks’s fluctuation theorem:** With a similar setup as that used in [47], Collin et al. tested Crooks’s fluctuation theorem (1.70) experimentally by unfolding and refolding a RNA hairpin [17]. A short interfering RNA hairpin which folds irreversibly but not very far from equilibrium (dissipation of the unfolding is around $6kT$) was stretched and unfolded at different pulling speeds.

In Fig. 1.8a we show the RNA hairpin used in their experiment as well as the force extension curves (force applied as a function of the distance between the pulling ends). The pulling is done irreversibly, which produces hysteresis in force-extension cycles as shown in Fig. 1.8a. The area under the force extension curves

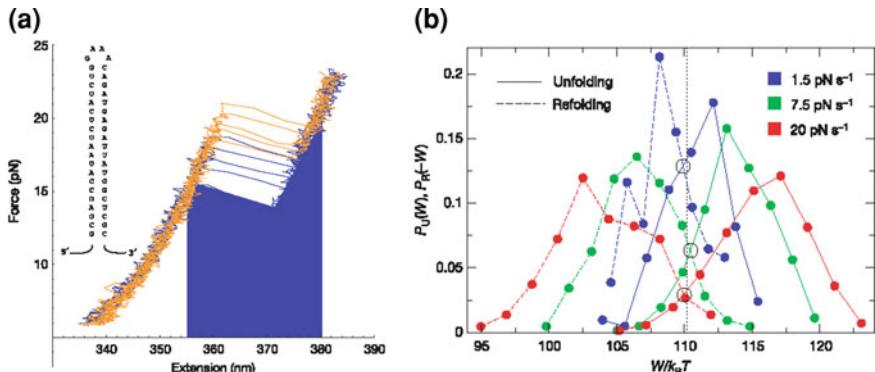


Fig. 1.8 Experimental verification of Crooks's fluctuation theorem. **a** RNA hairpin used in pulling experiments and five force extension curves of the unfolding (orange) and refolding (blue) curves at a load rate of 7.5 pN s^{-1} . The shaded area is the work returned to the external agent in refolding processes. **b** Unfolding (solid) $\rho_U(W)$ and refolding (dashed) $\rho_R(-W)$ work histograms at different pulling speeds. The distributions at different speeds cross around $110 kT$. Data and pictures taken from [17]

of the unfolding (refolding) process is the work done along the unfolding (refolding) process. The distribution of the work of the unfolding $\rho_U(W)$ and refolding $\rho_R(-W)$ is shown in Fig. 1.8b. As seen in Fig. 1.8b, the work distributions $\rho_U(W)$ and $\rho_R(-W)$ cross at the same value of the work for different pulling speeds. The value of the work at which the two work distributions cross is equal to the free energy difference between the folded and unfolded states of the RNA hairpin measured with molecular dynamics simulations, reproducing Crooks's prediction (1.73).

- **Steady state (SSFT) and transient fluctuation theorems (TFT):** The steady state fluctuation theorem (1.75) was first tested in numerical simulations of a shearing fluid [25] and later on, with a molecular dynamics simulation of a colloidal particle immersed in water translated by an optical trap [66]. Since then, several experiments have been done in order to check the validity of both SSFT and TFT described in Sect. 1.5.3. Ciliberto et al. checked a SSFT (1.75) for the heat in the long time limit in a Rayleigh-Bernard cell made of water [13] and also in turbulent flows [14]. Feitosa et al. studied the energy fluctuations in a granular medium formed by 1mm-diameter spherical glass beads that are contained in a rectangular cage, finding that the SSFT holds in the long time limit [27]. In [30], the SSFT is verified for the power fluctuations in an electric circuit with a resistor in parallel with a capacitance. Schuler et al. [54] studied the validity of the SSFT in a defect center in a diamond that was periodically excited by a laser, which is an example of a two-level system under a nonthermal noise. This experiment proves that fluctuation theorems also hold for nonlinear dynamics where entropy fluctuations are not Gaussian. Jop et al. [38] studied the fluctuations of the work and heat for a colloidal particle in an optical tweezer immersed in water. The optical

tweezer is switched faster than the relaxation time of the bead to generate a double well potential, and the intensity is modulated sinusoidally, driving the particle to a steady state in which the SSFT for both work and heat is true in the long time limit. Wang et al. [67] checked the TFT for a colloidal particle driven horizontally by an optical tweezer and the steady state fluctuation theorem when driving the tweezer circularly. In Ref. [33], the limitations of the SSFT are analyzed in two different scenarios: a colloidal particle trapped with an optical tweezer whose center moves stochastically and a atomic force microscopy cantilever close to a metallic surface.

1.7 Outline and Goals

Understanding and exploiting the relationship between irreversibility and dissipation in nonequilibrium processes is one of the main goals of thermodynamics and statistical mechanics. Until the development of stochastic thermodynamics, the first quantitative study of this relationship was done in linear irreversible thermodynamics, where entropy production was found to be directly proportional to macroscopically observable flows or currents. This result, however, only applies to macroscopic systems that are weakly perturbed out of equilibrium and it does not explain the microscopic origin of the entropy production.

During the last years, stochastic thermodynamics has become an emergent field in physics, focusing on the understanding the dynamics and the energetics of systems of small scale. Fluctuation and work theorems have provided a theoretical framework that allows one to quantify the probability of rare events that only occur in the microscopic (or sub-microscopic) scale, such as the observation of negative entropy production in particular realizations of nonequilibrium processes. Recent results in this framework suggest that irreversibility and dissipation can be *quantitatively* connected for microscopic systems driven arbitrarily far from equilibrium.

This thesis is devoted to analyze the relationship between irreversibility and dissipation for microscopic systems that reach a nonequilibrium steady state (NESS). We are first interested in expressing this relationship in the NESS and secondly, we aim to apply this result to practical situations. The main goal of this work is to find a technique that allows one to estimate the average dissipation of a microscopic system in a NESS from the statistics of a *single* stationary trajectory produced by the system during the process. We want our technique to be able to estimate the average dissipation of the system by using any data (one or several discrete or continuous degrees of freedom) sampled from the system and even ignoring any physical detail of the system. Our method uses the Kullback-Leibler divergence (KLD) to measure the distinguishability between a stationary trajectory and its time-reversal, providing a quantitative tool to measure the arrow of time in the NESS, i.e. the time irreversibility. We outline the content of the thesis according to the main goals of our research:

- 1. Estimation of the KLD from single stationary trajectories.** In order to apply our technique to a variety of physical situations, we need to develop estimators

of the KLD for both discrete and continuous data. We search for techniques that allow estimating the KLD using finite data from a single stationary trajectory, in order to apply our method to experimental data. This issue is addressed in Chap. 3.

2. **Application to discrete systems.** As a first case study, we will focus on discrete systems or coarse-grained descriptions of continuous systems. Using the existing techniques to estimate Shannon entropy of discrete data, we will extend this techniques to estimate KLD rate for discrete data obtained from a microscopic system in a NESS. We will check if we can distinguish between equilibrium and NESS by means of the value of the KLD and if we can infer or bound the dissipation of the physical mechanism that generated the discrete data. We will use a discrete ratchet model as a case study, where dissipation can be analytically calculated and compared to the KLD estimation using trajectories from simulations of the model. This goal is addressed in Chap. 4.
3. **Application to continuous systems. Distinction between active and passive biological processes.** We extend the estimation technique to continuous data, in particular, to data produced by biological systems. More precisely, our goal is to use the KLD to distinguish between active and passive biological processes. Ear hair bundles are an excellent biological system to study this problem since previous works succeeded to detect active processes using data from spontaneous and forced oscillations of the bundles [49]. Our aim is to improve the method to discriminate between the active and passive case by applying our method only to data from spontaneous oscillations. Both simulations and experimental data will be considered in this study. This issue is addressed in Chap. 5.
4. **Study of the energetics of the symmetry breaking.** We will show that the energetics of a generic symmetry breaking, as well as the energetics of a symmetry restoration, can be reproduced with the relationship between irreversibility and dissipation for microscopic systems. We are interested in studying the relationship between entropy production and the probability to choose an instance in a symmetry breaking, i.e., to study the thermodynamics of a choice of a generic probability p . This goal is addressed in Chap. 6.

The energetics of symmetry breaking, as well as other interesting results of stochastic thermodynamics, can be experimentally tested using optical tweezers. Optical traps allow to measure physical properties of microscopic systems, where the energetics of the physical processes are of the order of the characteristic energy of thermal fluctuations, kT . In this thesis we are also interested in checking theoretical results in the framework of stochastic thermodynamics experimentally, using optical tweezers, which we now outline.

1. **Measurement of the signature of the energetics of symmetry breaking.** Our aim is to design an experiment that allows us to check the relationship between the probability to choose an option in a symmetry breaking with the entropy produced when a microscopic system chooses that option. A Brownian particle in a double-well potential with an external force that tunes the probability to stay in any of the two wells can be used as a case study. This situation might be implemented

experimentally by trapping a colloidal particle using a dual optical tweezer. This goal is addressed in Chap. 6.

2. **Temperature control in the microscopic scale.** In a recent work, Blickle et al. used optical tweezers to design a microscopic heat engine [6], where a microscopic Brownian particle operates between two thermal baths at different temperatures. We are interested in designing an experimental technique to control the kinetic temperature of a Brownian particle and therefore to build new artificial micro and nano heat engines in the spirit of [6]. This issue is addressed in Chap. 7.

References

1. A. Ashkin, J. Dziedzic, J. Bjorkholm, S. Chu, Opt. lett. **11**, 288–290 (1986)
2. R. Balescu, NASA STI/Recon Technical Report A **76**, 32809 (1975)
3. G.K. Batchelor, in *An Introduction to Fluid Dynamics* (Cambridge university press, Cambridge, 2000)
4. C.H. Bennett, Int. J. Theoret. Phys. **21**, 905–940 (1982)
5. A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, E. Lutz, Nature **483**, 187–189 (2012)
6. V. Blickle, C. Bechinger, Nat. Phys. **8**(2), 143–146 (2012)
7. R. Brown, Philos. Mag. Ann. Chem. Math. Astron. Nat. Hist. Gen. Sci. **4**, 161–173 (1828)
8. C. Van den Broeck, M. Esposito, Phys. Rev. E **82**, 011144 (2010)
9. C. Bustamante, D. Keller, G. Oster, Acc. Chem. Res. **34**, 412–420 (2001)
10. C. Bustamante, Z. Bryant, S. Smith et al., Nature, 423–426 (2003)
11. C. Bustamante, J. Liphardt, F. Ritort Farran, Phys. Today **58**(7), 43–48 (2005)
12. H.B. Callen, in *Thermodynamics and an Introduction to Thermostatistics*, 2nd edn. ed by H.B. Callen, (Wiley-VCH, New York, 1985), p. 512. ISBN 0-471-86256-8
13. S. Ciliberto, C. Laroche, Le Journal de Physique. IV, **8**(6), 215–219 (1998)
14. S. Ciliberto, N. Garnier, S. Hernandez, C. Lacpatia, J.-F. Pinton, G. Ruiz Chavarria, Physica A: Stat. Mech. Appl. **340**, 240–250 (2004)
15. B. Cleuren, C. Van den Broeck, R. Kawai, Phys. Rev. Lett. **96**, 050601 (2006)
16. W. Coffey, Y.P. Kalmykov, J.T. Waldron, in *The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry, and Electrical Engineering*, vol 14 (World Scientific Publishing Company Incorporated, New York, 2004)
17. D. Collin, F. Ritort, C. Jarzynski, S. Smith, I. Tinoco, C. Bustamante, Nature **437**, 231–234 (2005)
18. T.M. Cover, J. A. Thomas, in *Elements of Information Theory* (Wiley-Interscience, New York, 2006)
19. G.E. Crooks, J. Stat. Phys. **90**, 1481–1487 (1998)
20. G. Crooks, Phys. Rev. E **60**, 2721 (1999)
21. S.R. De Groot, P. Mazur, in *Non-equilibrium Thermodynamics* (Dover Publications, New York, 2011)
22. W. Ducker, T. Senden, R. Pashley, Nature **353**, 239–241 (1991)
23. A. Einstein, Ann. Phys. Lpz **17**, 549 (1905)
24. M. Esposito, C. Van den Broeck, Phys. Rev. Lett. **104**, 090601 (2010)
25. D.J. Evans, E. Cohen, G. Morriss, Phys. Rev. Lett. **71**, 2401–2404 (1993)
26. D.J. Evans, D.J. Searles, Phys. Rev. E **53**, 5808 (1996)
27. K. Feitosa, N. Menon, Phys. Rev. Lett. **92**, 164301 (2004)
28. E. Fermi, *Thermodynamics* (Dover, New York, 1956)
29. G. Gallavotti, E. Cohen, Phys. Rev. Lett. **74**, 2694–2697 (1995)

30. N. Garnier, S. Ciliberto, Phys. Rev. E **71**, 060101 (2005)
31. W. Greiner, L. Neise, H. Stöcker, D. Rischke, in *Thermodynamics and Statistical Mechanics*, vol 1 (Springer, New York, 1995)
32. G. Grimmett, D. Stirzaker, in *Probability and Random Processes*, vol 2 (Clarendon press Oxford, Oxford, 1992)
33. J.R. Gomez-Solano, L. Bellon, A. Petrosyan, S. Ciliberto, EPL (Europhys. Lett.) **89**, 60003 (2010)
34. K. Huang, Stat. Mech. **18**, 3 (1987)
35. C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997)
36. C. Jarzynski, Eur. Phys. J. B. **64**, 331–340 (2008)
37. C. Jarzynski, Annu. Rev. Condens. Matter Phys. **2**, 329–351 (2011)
38. P. Jop, A. Petrosyan, S. Ciliberto, EPL (Europhys. Lett.) **81**, 50005 (2008)
39. F. Jülicher, A. Ajdari, J. Prost, Rev. Mod. Phys. **69**, 1269 (1997)
40. D. Kondepudi, I. Prigogine, in *From Heat Engines to Dissipative Structures* (Wiley, New York, 1998)
41. R. Kubo, Reports on Progress in Physics **29**, 255 (1966)
42. J. Kurchan, J. Phys. A: Math. Gen. **31**, 3719 (1999)
43. P. Langevin, CR Acad. Sci. Paris **146**, 530–533 (1908)
44. R. Landauer, IBM j. Res. Dev. **5**, 183–191 (1961)
45. J.L. Lebowitz, H. Spohn, J. Stat. Phys. **95**, 333–365 (1999)
46. H.S. Leff, A.F. Rex, D.L. Hogenboom, Am. J. Phys. **60**, 282 (1992)
47. J. Liphardt, S. Dumont, S.B. Smith, I. Tinoco Jr, C. Bustamante, Science **296**, 1832–1835 (2002)
48. C. Maes, J. Stat. Phys. **95**, 367–392 (1999)
49. P. Martin, A. Hudspeth, F. Jülicher, Proc. Natl. Acad. Sci. **98**, 14380–14385 (2001)
50. G. Mie, Annalen der Physik **330**, 377–445 (1908)
51. J.M.R. Parrondo, Chaos Interdisc. J. Nonlinear Sci. **11**, 725–733 (2001)
52. S. Park, F. Khalili-Araghi, E. Tajkhorshid, K. Schulten, J. Chem. Phys. **119**, 3559 (2003)
53. R. Pathria, in *Statistical Mechanics: International Series in Natural Philosophy*, vol 45 (Pergamon Press, New York, 1988)
54. S. Schuler, T. Speck, C. Tietz, J. Wrachtrup, U. Seifert, Phys. Rev. Lett. **94**, 180602 (2005)
55. D.J. Searles, D.J. Evans, Phys. Rev. E **60**, 159 (1999)
56. U. Seifert, Phys. Rev. Lett. **95**, 40602 (2005)
57. U. Seifert, Eur. Phys. J. B **64**, 423–431 (2008)
58. U. Seifert, arXiv preprint [arXiv:1205.4176](https://arxiv.org/abs/1205.4176) (2012)
59. K. Sekimoto, Prog. Theoret. Phys. Suppl. **130**, 17–27 (1998)
60. K. Sekimoto, in *Stochastic Energetics*, vol 799 (Springer, Berlin, 2010)
61. L. Szilard, Zeitschrift für Physik A Hadrons and Nuclei **53**, 840–856 (1929)
62. S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, M. Sano, Nature Phys. **6**, 988–992 (2010)
63. K. Visscher, M. Schnitzer, S. Block, Nature **400**, 184–189 (1999)
64. R. Van Zon, E. Cohen, Phys. Rev. Lett. **91**, 110601 (2003)
65. R. Van Zon, E. Cohen, Phys. Rev. E **67**, 046102 (2003)
66. G. Wang, E.M. Sevick, E. Mittag, D.J. Searles, D.J. Evans, Phys. Rev. Lett. **89**, 50601 (2002)
67. G. Wang, J. Reid, D. Carberry, D. Williams, E.M. Sevick, D.J. Evans, Phys. Rev. E **71**, 046142 (2005)
68. W.H. Zurek: arXiv preprint [quant-ph/0301076](https://arxiv.org/abs/quant-ph/0301076) (2003)
69. R. Zwanzig, J. Stat. Phys. **9**, 215–220 (1973)

Part II

Irreversibility and Dissipation

Chapter 2

Dissipation and Kullback–Leibler Divergence

In this chapter, we introduce the theoretical framework of the first part of our work, in which we study of the relationship between dissipation and irreversibility quantitatively in microscopic systems in the stationary state.

The relationship between entropy production (dissipation) and irreversibility forms the core of thermodynamics and statistical mechanics. The first studies in dissipation and irreversibility in nonequilibrium processes were done in the context of linear irreversible thermodynamics [16]. In linear regime, the entropy production is linear with the force that drives the system out of equilibrium. However, this relation holds only in linear regime and does not provide a quantitative description of entropy production in terms of the microscopic properties of the system.

With the introduction of fluctuation theorems (see Sect. 1.5), it is possible to derive exact relationships that connect the entropy production of a microscopic system in the NESS with its microscopic properties and, moreover, provide a quantitative tool to measure the time irreversibility of the process. Our work is devoted to clarify this relation and to provide tools to estimate time irreversibility in the NESS from a single stationary trajectory [18, 24, 25]. In this chapter, we describe the theoretical framework of our approach, whereas the estimation techniques and applications to simulations and experimental data are described in further chapters.

This chapter is organized as follows: In Sect. 2.1 we define the notion of time irreversibility in stochastic processes. In Sect. 2.2 we review the notion of dissipation in irreversible processes from the approach of linear irreversible thermodynamics to the new insights provided by fluctuation theorems. In Sect. 2.3 we introduce the concept of relative entropy or Kullback–Leibler divergence and show how it is related to the arrow of time. In Sect. 2.4 we show how dissipation and time irreversibility are quantitatively connected in the NESS. In Sect. 2.5 we study discrete systems and obtain exact expressions for the Kullback–Leibler divergence and the dissipation for two specific stochastic processes.

2.1 What Is Time Irreversibility?

Time irreversibility is the ability to distinguish between a process and its time reversal. In irreversible processes, one can ascertain from the observation of the process if the time is running forward or backwards. The fingerprint of an irreversible processes is therefore the ability to guess the direction of the *arrow of time*.

In the macroscopic world, time irreversibility occurs in different scenarios. One example is a magnetic hysteresis cycle. In such a cycle, the system does not recover its original demagnetized state after a periodic change of the external field. Another example is a glass falling to the ground and smashing into pieces, where the time-reversal process is never observed.

In the microscopic world, the arrow of time is blurred because of thermal fluctuations. Guessing if a process is time reversible or not from a single realization of the process is challenging: an irreversible process can look time reversible when sampled at different frequencies or using insufficient statistics. However, because of the fast relaxation times in the microscale, one can measure the probability to observe a path or its time reversal from the statistics of different trials. This probability can be used to *quantify* the time irreversibility of a process.

To quantify time irreversibility in the microscopic world one therefore needs a metric to compare the probability distributions of forward and backward trajectories. Such metric is the Kullback–Leibler divergence or relative entropy. In this chapter we show how one can quantify the time irreversibility using the Kullback–Leibler divergence and how this quantification is related to the dissipation of the process.

2.2 Average Dissipation in Irreversible Processes

In nonequilibrium processes, an irreversible process is accompanied by a positive entropy production. In linear regime, entropy production in macroscopic systems is a bilinear quadratic form on the macroscopic flows or currents of the systems.

At small scales, an analogous relation between entropy production and currents is found for systems obeying an overdamped equation as shown in Sect. 1.4.3. Using novel results derived in the context of fluctuation theorems, it is possible to measure the entropy production of microscopic systems that are driven arbitrarily far from equilibrium under an arbitrary external protocol.

2.2.1 Linear Irreversible Thermodynamics

Linear irreversible thermodynamics studies systems that are driven not far from equilibrium in the context of linear response theory [16]. Close to equilibrium, entropy production can be expressed as a linear combination of all the different

thermodynamic forces or gradients F_i that are exerted on the system. In this limit, the entropy production per unit volume, σ , equals to

$$\sigma = \sum_i F_i J_i, \quad (2.1)$$

where i runs over all the different forces on the system and J_i is the flux associated to the force F_i . One example is the heat flux J_Q , which is produced by a force that is proportional to the gradient of temperature, $F_Q \propto \nabla \frac{1}{T}$. Another example is the electric current $J_e = I$ that is driven by an electric field E , $F_e \propto E$. In linear regime, the forces are proportional to the fluxes

$$F_j = \sum_k L_{jk} J_k, \quad (2.2)$$

where L_{jk} are the *phenomenological coefficients*. The above relation expresses that, for example, it is possible to induce a heat flow from an electric current, or vice versa. Taking into account (2.2), entropy production per unit volume equals to

$$\sigma = \sum_{i,k} L_{ik} J_i J_k. \quad (2.3)$$

We notice that in linear response, entropy production is a positively defined quadratic form of the currents and it is therefore related to the presence of macroscopically observable flows, and since if $J > 0$, then $\sigma > 0$. This formulation however does not connect the entropy production with the microscopic properties of the system. Using fluctuation theorems, it is possible to obtain a formula that expresses the entropy production for microscopic systems driven arbitrarily far from equilibrium and do a connection between the work dissipated and the microscopic properties of the system.

2.2.2 Entropy Production in Microscopic Systems

As we showed in Sect. 1.4.3, the definition of the entropy associated to a trajectory of a microscopic system allows one to introduce the notion of entropy production in the case of a Brownian particle obeying an overdamped Langevin equation. The following expression relating the ensemble average of the entropy production and the probability flux was derived by Seifert [26],

$$\langle \dot{S}_{\text{prod}}(t) \rangle = k \int dx \frac{j(x, t)^2}{D\rho(x, t)} \geq 0. \quad (2.4)$$

Notice that this expression was also obtained for Brownian chemical motors described by Langevin equation [21]. We notice that (2.4) expresses a relationship between

entropy production and the probability current, $j(x, t)$. According to (2.4), the entropy production a Brownian particle vanishes if $j(x, t)$ does, which means that irreversibility for these kind of systems is revealed in the flows of the system. Moreover, the entropy production depends on the current as j^2 , which is in accordance with linear response theory, as shown in (2.3).

2.2.3 KPB Theorem

Recently, a quantitative relationship between dissipation and irreversibility in non-equilibrium processes for microscopic systems has been derived. The introduction of fluctuation theorems has allowed to express the entropy production of an isolated microscopic system in terms of the microscopic properties of the system [11, 15, 22]. The main result was derived by Kawai et al. [15] and it is known in literature as the KPB (Kawai, Parrondo and van den Broeck) theorem, which we now discuss.

In Refs. [15, 22], the dissipation of microscopic isolated systems that are brought from an initial equilibrium state at temperature T to a final equilibrium state at the same temperature is investigated. An expression relating the average dissipation (or entropy production) in such processes with the distinguishability between the process and its time reversal is found.

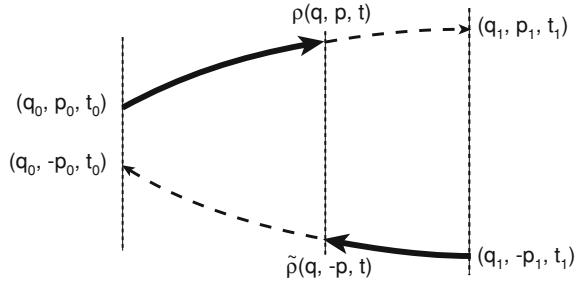
Consider an isolated physical system described by a Hamiltonian $H(q, p; \lambda)$, where (q, p) denotes a point in phase space, and λ is a parameter of the system controlled by an external agent. The system is initially in a canonical equilibrium state at temperature T . Then the system is disconnected from the thermal bath and driven out of equilibrium according to a protocol in which the external agent modifies λ from $\lambda(0) = \lambda_A$ to $\lambda(\tau) = \lambda_B$, following a protocol $\{\lambda(t)\}_{t=0}^\tau$. At $t = \tau$ the system is weakly coupled to a thermal bath and relaxes to a canonical state at temperature T .

The dissipation of the process described above (which we call *forward* process) is related to the distinguishability between phase space densities of the forward process and its time reversal (or *backward* process). In the backward process, the system is initially in a canonical equilibrium state at temperature T and driven by the time-reversed protocol $\{\tilde{\lambda}(t)\}_{t=0}^\tau = \{\lambda(\tau - t)\}_{t=0}^\tau$. For a given trajectory in the forward process that starts in $(q_0, p_0; 0)$ and ends in $(q_1, p_1; \tau)$, the corresponding time reversed trajectory is obtained by reversing the position and changing the sign of the momenta, i.e., the time reversed trajectory starts in $(q_1, -p_1; \tau)$ and ends in $(q_0, -p_0; 0)$ (see Fig. 2.1). Notice that the time t is taken in the forward process.

The KPB theorem relates the average dissipation in the forward process, $\langle W_{\text{diss}} \rangle$, with the distinguishability between the forward and the backward phase space densities measured at the same but arbitrary time during the process, $t \in [0, \tau]$,

$$\langle W_{\text{diss}} \rangle = \langle W \rangle - \Delta F = kT \int dq dp \rho(q, p; t) \ln \frac{\rho(q, p; t)}{\tilde{\rho}(q, -p; t)}. \quad (2.5)$$

Fig. 2.1 Forward and backward trajectories in phase space. Picture taken from [15]



This result shows that the dissipation of a nonequilibrium process is revealed in the phase space. The right hand side is often called relative entropy or *Kullback–Leibler divergence* (KLD) [4, 17] between the probability distributions $\rho(q, p; t)$ and $\tilde{\rho}(q, -p; t)$, which is denoted by the letter D ,

$$\langle W_{\text{diss}} \rangle = kT D[\rho(q, p; t) || \tilde{\rho}(q, -p; t)]. \quad (2.6)$$

The right hand side of (2.6) measures the difficulty to distinguish whether the microstate of the system at any time is generated from the forward or the backward experiment. Equation (2.6) links directly the average dissipation with the time irreversibility of the process, which can be *quantified* using the KLD. As we will show in Sect. 2.3, the value of the KLD increases when the two probability distributions are more different each other, indicating that the more different are the forward and reverse process the more work is dissipated. The KLD is positive, which ensures that the second law of thermodynamic holds in average, $\langle W_{\text{diss}} \rangle = kT D[\rho || \tilde{\rho}] \geq 0$. The dissipation of microscopic systems is revealed in the phase space and one can get tighter bounds to the dissipation than in the macroscopic case, where $\langle W_{\text{diss}} \rangle \geq 0$.

The KPB theorem can be extended to more general initial (equilibrium) conditions. In Ref. [22] it is proved that the change of the entropy in the system plus the bath averaged over many realizations of the process is equal to

$$\langle S_{\text{prod}} \rangle = k D[\rho(q, p; t) || \tilde{\rho}(q, -p; t)]. \quad (2.7)$$

Equation (2.7) is valid for a variety of initial equilibrium conditions, as shown in Ref. [22]: canonical, multi-canonical (several uncoupled systems at different temperatures), and grand-canonical distributions, as well as for different types of baths equilibrating the system at the end of the process. In all these cases, the formula holds when the evolution is isolated and the control parameter follows any arbitrary nonequilibrium protocol. In particular, for canonical initial conditions in the forward and in the backward processes, both at the same temperature T , entropy production equals the average dissipated work $\langle W_{\text{diss}} \rangle = \langle W \rangle - \Delta F$ divided by the temperature T and (2.7) becomes (2.6), $\langle W_{\text{diss}} \rangle = kT D[\rho(q, p; t) || \tilde{\rho}(q, -p; t)]$.

We now reproduce the proof of Eq. (2.7) for the case where both initial conditions of the forward and backward processes are canonical at temperature T . We use the

notation $z = (q, p)$ to denote all the variables in phase space, $\tilde{z} = (q, -p)$ being the corresponding microstate obtained by changing the sign of all the momenta. We first consider that the KLD $D[\rho(z; t) || \tilde{\rho}(\tilde{z}; t)]$ can be rewritten as

$$D[\rho(z; t) || \tilde{\rho}(\tilde{z}; t)] = \int dz \rho(z; t) \ln \rho(z; t) - \int dz \rho(z; t) \ln \tilde{\rho}(\tilde{z}; t). \quad (2.8)$$

Secondly, since the system is isolated during its evolution, the phase space density evolves according to Liouville's equation [cf. Eq.(1.28)]

$$\frac{\partial \rho(z; t)}{\partial t} = \mathcal{L}\rho(z; t). \quad (2.9)$$

In addition, the backward phase space density $\tilde{\rho}(\tilde{z}; t)$ obeys the same Liouville equation considering the derivative with respect to the forward time t [22],

$$\frac{\partial \tilde{\rho}(\tilde{z}; t)}{\partial t} = \mathcal{L}\tilde{\rho}(\tilde{z}; t). \quad (2.10)$$

The fact that both ρ and $\tilde{\rho}$ obey the same Liouville equation implies that the two terms in Eq. (2.8) are invariant in time, which can be proved by using both Liouville's equations and partial integration. A proof of the time invariance of the first term in Eq. (2.8) can be found in Appendix B.1, and the proof for the invariance of the second term is analogous. Consequently, we express (2.8) by evaluating the two terms at any time t , in particular,

$$D[\rho(z; t) || \tilde{\rho}(\tilde{z}; t)] = \int dz \rho(z; 0) \ln \rho(z; 0) - \int dz \rho(z; \tau) \ln \tilde{\rho}(\tilde{z}; \tau). \quad (2.11)$$

The first term in (2.11) corresponds to (minus) the system entropy in the beginning of the process, in k units. The second term in (2.11) can be interpreted as the system entropy at the end of the process plus the change in the entropy of the environment, as we shall see. We notice the difference between $\rho(z; \tau)$, which is the phase space density at the end of the forward process, which is not at equilibrium, and $\tilde{\rho}(\tilde{z}; \tau)$, which is the initial (equilibrium) distribution of the backward process.

Let us consider a process where a system that is initially in contact with a thermal bath at temperature T is disconnected from the bath and driven out of equilibrium by an external agent following a protocol $\{\lambda(t)\}_{t=0}^{\tau}$ from $\lambda(0) = \lambda_A$ to $\lambda(\tau) = \lambda_B$. After the process, the system is put in contact with a thermal bath at temperature T' and let relax to an equilibrium state. We assume that during the nonequilibrium driving, the Hamiltonian of the system is $H(z, \lambda)$. In this case, the initial distributions of the forward and backward processes are

$$\rho(z; 0) = \frac{e^{-\beta H(z; \lambda_A)}}{Z(T, \lambda_A)}, \quad \tilde{\rho}(\tilde{z}; \tau) = \frac{e^{-\beta H(\tilde{z}; \lambda_B)}}{Z(T', \lambda_B)}, \quad (2.12)$$

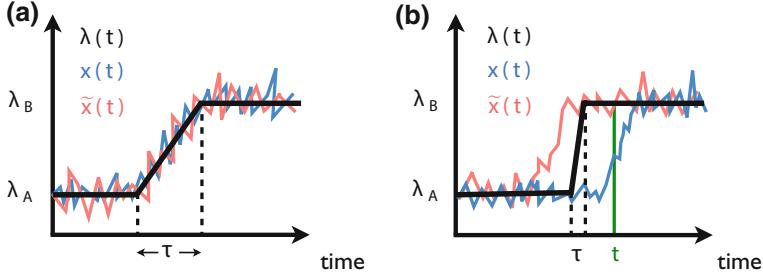


Fig. 2.2 KPB theorem in an example. A system described by a single degree of freedom x is driven by an external agent following a protocol in which a control parameter is changed linearly from $\lambda(0) = \lambda_A$ to $\lambda(\tau) = \lambda_B$. **a** Protocol (black line), and the value of x in the forward (blue) and backward process (red) when the total time of the process is much larger to the relaxation time of the system $\tau \gg \tau_r$. The dynamics is reversible and the phase space densities of forward and backward processes coincide at any time. **b** Same graphs when the time of the process is much smaller than the relaxation time of the system, $\tau \ll \tau_r$. The dynamics is irreversible and the phase space densities of forward and backward processes do not coincide at every time. In this case, at time t indicated in green in the figure, $\rho(x, t) \neq \tilde{\rho}(\tilde{x}, t)$ and we can measure the average dissipation of the forward process with $\langle W_{\text{diss}} \rangle = kT D[\rho(x, t) || \tilde{\rho}(x, t)]$

where $Z(t, \lambda) = \int dz e^{-H(z; \lambda)/kT}$. By replacing the above distributions in (2.11) and taking into account that $F(T, \lambda) = -kT \ln Z(T, \lambda)$,

$$k D[\rho(z; t) || \tilde{\rho}(\tilde{z}; t)] = -S(0) + \frac{\langle H \rangle_\tau - F(T', \lambda_B)}{T'}. \quad (2.13)$$

Here $\langle H \rangle_\tau$ is an average over $\rho(z, \tau)$, which does not coincide with the average energy on the initial (equilibrium) state of the backward process, $\langle H \rangle_{\text{eq}, \tau}$. On the other hand, when the system relaxes to equilibrium after it is connected to the thermal bath at temperature T' at time $t = \tau$, the system transfers a heat to the environment $Q_{\text{env}} = \langle H \rangle_\tau - \langle H \rangle_{\text{eq}, \tau}$. Taking into account this, Eq. (2.13) can be rewritten as

$$\begin{aligned} k D[\rho(z; t) || \tilde{\rho}(\tilde{z}; t)] &= -S(0) + \frac{\langle H \rangle_{\text{eq}, \tau} - F(T', \lambda_B)}{T'} + \frac{Q_{\text{env}}}{T'} \\ &= -S(0) + S(\tau) + \Delta S_{\text{bath}} = S_{\text{prod}}, \end{aligned} \quad (2.14)$$

which proves Eq. (2.7) for the case of initial and final canonical equilibrium states at any temperatures T and T' . In particular, for the case of initial and final conditions at the same temperature T , we recover the expression relating work dissipation and irreversibility (2.6).

In Fig. 2.2 we show an illustrative example of the applicability of KPB theorem. A system is initially in equilibrium at temperature T with an externally-controlled control parameter λ fixed at value λ_A . Then the system is disconnected from the thermal bath and the control parameter is kept constant and equal to λ_A for a time longer than the relaxation time of the system, τ_r . Then the control parameter is changed from

λ_A to λ_B linearly during a time τ in isolated conditions. In the end of the process, the control parameter is held fixed at λ_B for a time longer than τ_r , and the system is let to relax by putting the system in contact with a thermal bath of temperature T . If the relaxation time of the system τ_r is very small compared to τ , $\tau_r \ll \tau$, the system relaxes to the equilibrium state at any time t and the process is done reversibly. In this case, the forward and reverse trajectories are indistinguishable and $\rho(x; t) \simeq \tilde{\rho}(x; t)$ at any time t along the process. If the process is done much faster than the characteristic time scale of the system, $\tau_r \gg \tau$ then the system does not relax to the equilibrium state during the process as shown in Fig. 2.2b. In this case, forward and backward trajectories are distinguishable, and in general $\rho(x; t) \neq \tilde{\rho}(x; t)$. The KLD between these two distributions is an estimation of the dissipation of the process.

We remark that the formulas (2.7) and (2.6) connecting dissipation and irreversibility in the microscopic scale were first proved for the case of systems driven out of equilibrium in isolated conditions. The relationship can be extended to any system immersed in a thermal bath at temperature T , described by the overdamped Langevin equation. When such a system is in contact with a thermal bath at temperature T , the system plus the bath can be viewed as an isolated “super system”. The KPB theorem implies that the average dissipation of the system plus the bath equals to $\langle W_{\text{diss}} \rangle = kT D(\rho || \tilde{\rho})$ where $D(\rho || \tilde{\rho})$ is calculated in the full phase space (system plus bath) [13]. This relation was tested experimentally using a Brownian particle dragged by an optical tweezer at constant speed and with an electric circuit with an imposed mean current [1, 3]. Notice that both the position of the Brownian particle and the charge inside the resistor of the circuit obey an overdamped Langevin equation with different physical parameters [1].

The KPB theorem can be generalized to other stochastic processes in the microscopic scale using previous results obtained in the framework of fluctuation theorems—which are valid not only in isolated conditions—such as Crooks’s theorem [6]. The connection between fluctuation theorems and the KPB theorem, as well as its generalization to other stochastic processes such as nonequilibrium steady states was done in [11]. The relation between irreversibility and entropy production in a generic nonequilibrium steady state is of particular interest in this work, and it is discussed in the next sections.

2.3 Kullback–Leibler Divergence and Irreversibility

The Kullback–Leibler divergence (KLD), or relative entropy, measures the distinguishability of two probability distributions or two random processes [4]. Let us consider a random variable X and let p and q be two different probability distributions of the random variable X . We denote by $p(x)$ and $q(x)$ the probability of the variable X to take the value x when it is distributed according to p and q , respectively. The KLD between the probability distributions p and q is defined as [17]

$$D[p(x) || q(x)] = \int dx p(x) \ln \frac{p(x)}{q(x)}. \quad (2.15)$$

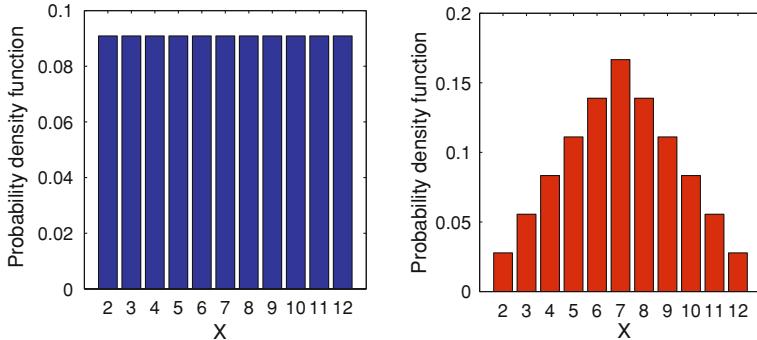


Fig. 2.3 Dice and lottery example of the KLD. *Left* probability density function of a random variable X that can take the values $2, 3, \dots, 12$ with equal probability as if it were drawn from a lottery. *Right* probability distribution of a random variable whose value is obtained as the sum of the outcome of two dice. One can measure the KLD between these discrete distributions $D(p||q) = \sum_i p_i \ln(p_i/q_i)$, where i runs from 2 to 12. The KLD between the two distributions is not symmetric, since $D(p_{\text{dice}}(x)||p_{\text{lottery}}(x)) = 0.128$ and $D(p_{\text{lottery}}(x)||p_{\text{dice}}(x)) = 0.152$

The KLD is always positive, $D[p(x)||q(x)] \geq 0$, and vanishes if and only if $p(x) = q(x)$ for all x . Therefore, D follows two of the main properties of a mathematical *distance*. However, D is not symmetric with respect to a change of arguments, $D[p(x)||q(x)] \neq D[q(x)||p(x)]$ as it can be seen in the simple example of Fig. 2.3.

The interpretation of the KLD as a measure of distinguishability is a consequence of *Chernoff-Stein lemma* [4]: the probability of incorrectly guessing (via hypothesis testing) that a sequence of n data is distributed according to p when the true distribution is q is asymptotically equal to $e^{-nD[p(x)||q(x)]}$. Therefore, when p and q are similar—in the sense that they overlap significantly—the likelihood of incorrectly guessing the distribution, p or q , is large [4]. In the example of Fig. 2.3, $D(p_{\text{dice}}||p_{\text{lottery}}) < D(p_{\text{lottery}}||p_{\text{dice}})$, meaning that it is easier to incorrectly guess—or equivalently harder to distinguish between distributions—that a sequence is generated from the dice, when true origin is the lottery. In other words, the smaller is the KLD, the more similar are the two distributions and it is harder to distinguish between them using hypothesis testing.

Chernoff-Stein lemma implies that the KLD in (2.6) and (2.7) can be considered as a measure of the arrow of time, since it measures the difficulty to distinguish whether the state (q, p) of the system at time t was generated in the forward or backward experiment [22]. Mathematically speaking, the relative entropy in KPB theorem (2.6),

$$D[\rho(q, p; t)||\tilde{\rho}(q, -p; t)] = \int dq dp \rho(q, p; t) \ln \frac{\rho(q, p; t)}{\tilde{\rho}(q, -p; t)}, \quad (2.16)$$

has to be understood as a relative entropy between two probability distributions of the *same* random variables (q, p) at time t : $\rho(q, p; t)$, and a second distribution $\sigma(q, p; t)$.

$$D[\rho(q, p; t) || \sigma(q, p; t)] = \int dq dp \rho(q, p; t) \ln \frac{\rho(q, p; t)}{\sigma(q, p; t)}, \quad (2.17)$$

such that the value of $\sigma(q, p; t)$ for (q, p) satisfies $\sigma(q, p; t) = \tilde{\rho}(q, -p; t)$.

Let us recall a property of the KLD that we use throughout our work [4]. If we have two random variables X, Y and two joint probability distributions $p(x, y)$ and $q(x, y)$, then the *chain rule* holds (see Appendix B.2),

$$\begin{aligned} D[p(x, y) || q(x, y)] &= D[p(x) || q(x)] + D[p(y|x) || q(y|x)] \\ &\geq D[p(x) || q(x)]. \end{aligned} \quad (2.18)$$

The inequality in Eq. (2.18) implies that it is harder to distinguish between p and q when we consider only the marginal distributions, $p(x)$ and $q(x)$, instead of the full joint distributions, $p(x, y)$ and $q(x, y)$. If X, Y describe the state of a physical system, Eq. (2.18) indicates that the KLD is smaller when calculated using a partial description of the system, given by the variable X , than when using full information (X and Y). The bound in (2.18) is an equality when the variable Y carries redundant information with respect to the variable X , for example, when Y is obtained as a function of X , $Y = f(X)$ for any function f .

When not all the degrees of freedom of the system can be sampled, we say that *partial information* of the physical system is available. Let x be *any* collection of m position and n momenta of the system $x = (q_1, \dots, q_m; p_1, \dots, p_n)$, where m and n can be different and $3n + 3m$ is smaller than the total number of degrees of freedom of the system. Since x describes in general only a part of the physical system we say that x contains *partial information* of the system. Because of the chain rule, when the state of the system is described with partial information given by x , the KPB theorem (2.6) turns into an inequality

$$\langle W_{\text{diss}} \rangle \geq kT D[\rho(x; t) || \tilde{\rho}(\tilde{x}; t)], \quad (2.19)$$

where $\tilde{x} = (q_1, \dots, q_m; -p_1, \dots, -p_n)$. We notice that even ignoring the full information of phase space (2.19) $D[\rho(x; t) || \tilde{\rho}(\tilde{x}; t)]$ still gives at least a lower bound to the average dissipation that is in accordance with the second law of thermodynamics, $\langle W_{\text{diss}} \rangle \geq kT D[\rho(x; t) || \tilde{\rho}(\tilde{x}; t)] \geq 0$. Consequently, when the system is described using only a reduced set of variables of the phase space, the KLD in (2.19) provides a lower bound to the dissipation.¹

¹ However, a recent work shows that, if the neglected information contains an external driving, the entropy production estimated in the coarse grained system can be bigger than the real entropy production [8].

A particular set of *redundant* variables do not provide any statistical information about the direction of the arrow of time of the process, and the value of the KLD remains the same if these variables are not sampled [11, 24, 25]. There are two groups of variables that provide *redundant* information to measure the irreversibility with the KLD. First, variables which are time reversible whose distribution is the same in the forward and backward processes. Second, system variables that are obtained as a function of other variables.

Now suppose that at every time t , we cannot measure the microstate $(q, p; t)$ of the system but we can only detect that the system is in a specific subset of the phase space. In this situation, the state of the system can be described using a *coarse-grained* random variable X that indicates in which subset the system is at every time t . For example, the position of a Brownian particle in one dimension x can be coarse-grained by introducing a new variable α that indicates if the value of the position is positive or negative, for example $\alpha = 0$ if $x \leq 0$ and $\alpha = 1$ if $x > 0$. If we phase space is partitioned in K non overlapping subsets $\{\mathcal{X}_j\}_{j=1}^K$ the coarse-grained forward and backward phase space densities are

$$\rho_j(t) = \int_{\mathcal{X}_j} dq dp \rho(q, p; t); \quad \tilde{\rho}_j(t) = \int_{\tilde{\mathcal{X}}_j} dq dp \tilde{\rho}(q, -p; t), \quad (2.20)$$

where $\tilde{\rho}_j$ is identical to ρ_j except a change of sign in all the momenta. These distributions measure the probability of the system to be in the region j of the phase space. At an arbitrary time t during the nonequilibrium process, the KLD between the forward and backward coarse-grained distributions is

$$D[\rho(t)||\tilde{\rho}(t)] = \sum_{j=1}^K \rho_j(t) \ln \frac{\rho_j(t)}{\tilde{\rho}_j(t)}, \quad (2.21)$$

Since the coarse-grained description of the state of the system is a partial description of its microstate, the KLD in Eq. (2.21) bound from below the KLD using full information of the phase space, by virtue of the chain rule. Therefore using a coarse-grained description of the microstate of the system, we can only bound from below the average dissipation [15],

$$\langle W_{\text{diss}} \rangle \geq kT D[\rho(t)||\tilde{\rho}(t)]. \quad (2.22)$$

The chain rule allows one to rewrite the KPB theorem using the KLD between the forward and reverse distributions of *trajectories* in phase space. For isolated systems, the evolution is deterministic, except for the last stage where the system is connected to the bath, and the point $z = (q, p)$ at time t determines the whole trajectory of the system $\{z(t)\}_{t=0}^\tau$. Then $z(t)$ and $\{z(t)\}_{t=0}^\tau$ carry the same information and the KLD of their respective probability densities are equal by virtue of the chain rule. Equation (2.6) can be rewritten in terms of *path probabilities*. Let $\mathcal{P}(\{z(t)\}_{t=0}^\tau)$ be the probability to observe a trajectory $\{z(t)\}_{t=0}^\tau = \{q(t), p(t); t\}_{t=0}^\tau$

in the forward process. The corresponding time-reversed path in the backward process starts in $(q(\tau), -p(\tau); \tau)$ and ends in $(q(0), -p(0); 0)$ as shown in Fig. 2.1. In a more compact notation, the time-reversed trajectory is defined as $\{\tilde{z}(\tau - t)\}_{t=0}^{\tau} = \{q(\tau - t), -p(\tau - t); \tau - t\}_{t=0}^{\tau}$. The probability to observe such a trajectory in the backward process is denoted by $\tilde{\mathcal{P}}(\{\tilde{z}(\tau - t)\}_{t=0}^{\tau})$. The KLD between forward and backward trajectory distributions equals to the KLD between phase space densities at every time t during the process,

$$D[\rho(z; t) || \tilde{\rho}(\tilde{z}; t)] = D[\mathcal{P}(\{z(t)\}_{t=0}^{\tau}) || \tilde{\mathcal{P}}(\{\tilde{z}(\tau - t)\}_{t=0}^{\tau})]. \quad (2.23)$$

Because of this, the dissipation in isolated microscopic systems can be expressed in terms of the distinguishability between the trajectory distributions of forward and backward processes. Equation (2.6) can be rewritten as [11, 24, 25]

$$\langle W_{\text{diss}} \rangle = kT D[\mathcal{P}(\{z(t)\}_{t=0}^{\tau}) || \tilde{\mathcal{P}}(\{\tilde{z}(\tau - t)\}_{t=0}^{\tau})]. \quad (2.24)$$

The above KLD has to be understood as a KLD between two distributions of a stochastic process [cf. (B.18) in Appendix B.2]. As noticed before in (2.17), the right hand side in (2.24) is a KLD between \mathcal{P} and a different trajectory distribution \mathcal{Q} ,

$$\langle W_{\text{diss}} \rangle = kT \int \mathcal{D}(\{z(t)\}_{t=0}^{\tau}) \mathcal{P}(\{z(t)\}_{t=0}^{\tau}) \ln \frac{\mathcal{P}(\{z(t)\}_{t=0}^{\tau})}{\mathcal{Q}(\{z(t)\}_{t=0}^{\tau})}, \quad (2.25)$$

where \mathcal{Q} is such that $\mathcal{Q}(\{z(t)\}_{t=0}^{\tau}) = \mathcal{P}(\{\tilde{z}(\tau - t)\}_{t=0}^{\tau})$.

We now recall that using Crooks's fluctuation theorem, we can arrive to an expression of the average dissipative work in terms of the forward and backward work distributions. Integrating Crooks's theorem (1.70), $W - \Delta F = \ln \frac{\rho(W)}{\rho(-W)}$, where $\rho(W)$ [$\tilde{\rho}(W)$] is the probability density of the work done on the system along the forward (backward) process [6, 11], one immediately gets

$$\langle W_{\text{diss}} \rangle = kT D[\rho(W) || \tilde{\rho}(-W)]. \quad (2.26)$$

Notice that the work W is a functional of the trajectory $\{z(t)\}_{t=0}^{\tau}$ [see (1.59)] containing less information than the trajectory itself. As indicated by the chain rule (2.18), the KLD of work distributions should in principle be smaller than the KLD of trajectory distributions, and therefore bound the dissipation from below. On the contrary, the KLD is the same, indicating that all the irreversibility of the process is captured by the dissipative work [11]. Equation (2.26) indicates that the irreversibility in the work reveals the entropy production.

If the microstate of the system is not known are every time t but only a trajectory containing the evolution of a subset of variables of the phase space, $\{x(t)\}_{t=0}^{\tau}$, the KLD between forward and backward trajectories yields a lower bound to the dissipation,

$$\langle W_{\text{diss}} \rangle \geq kT D[\mathcal{P}(\{x(t)\}_{t=0}^{\tau}) || \tilde{\mathcal{P}}(\{\tilde{x}(\tau - t)\}_{t=0}^{\tau})]. \quad (2.27)$$

The bound saturates when the variables that are sampled capture the same information about the irreversibility as the work. The information of the irreversibility is contained in the variables that interact with the work-performing device which are the *footprints* of irreversibility [11].

When the state of the system cannot be sampled at every time t along the process but only every finite time interval, Δt , the sampled trajectory contains less information than the path of the system along the process (which coincides with the limit $\Delta t \rightarrow 0$). In this case, even if the microstate of the system $(q(t), p(t); t)$ is known every Δt , the KLD gives a lower bound to the dissipation. In Ref. [10], Gomez-Marin et al. studied the dynamics of an overdamped Brownian particle that moves in one dimension dragged by a harmonic potential that moves at constant velocity. They measured the KLD between trajectories of the position of the particle $\{x(t + i\Delta t)\}_{i=0}^{\tau/\Delta t}$ in the forward process and the backward process, the latter consisting in moving the trap at the opposite velocity as in the forward process. When reducing the sampling time Δt , the KLD approaches asymptotically to the entropy production in the system.

In summary, when considering the KLD between forward and backward distributions of trajectories, the irreversibility is partially captured by the KLD when one or more of the following phenomena occur:

- **Coarse graining:** Only a coarse-grained description of the system is available, that is, we only know in which of an ensemble of phase space subsets the system is.
- **Partial information:** Not all the variables of phase space can be sampled but only a subset of variables in phase space.
- **Finite time sampling:** The trajectory cannot be sampled at any time but only at a finite sampling frequency.

When the information about the system is not full in the sense that one of the three above mentioned shortcomings occur, the KLD between forward and backward trajectories is a lower bound to the average entropy production in the system. In Fig. 2.4 we show an illustrative example where the three types of lack of information can occur.

In the majority of experimental contexts, only system variables are measurable, neglecting the variables of the bath and therefore using partial information of the system. In this case, the KLD yields a lower bound to the dissipation $\langle W_{\text{diss}} \rangle \geq kT D(\rho_F || \rho_B)$. In [13], this result is illustrated with an overdamped Brownian particle in a dragged harmonic trap immersed in a thermal bath at temperature T . It is shown that in isothermal conditions, a lower bound to the dissipation is obtained using the KLD $\langle W_{\text{diss}} \rangle \geq kT D(\rho_F || \rho_B)$. The bound is tighter when decreasing the friction coefficient γ , and therefore the when the coupling between the Brownian particle and the reservoir is weaker. When $\gamma \rightarrow 0$, the system is uncoupled to the bath and therefore isolated, and the equality is met.

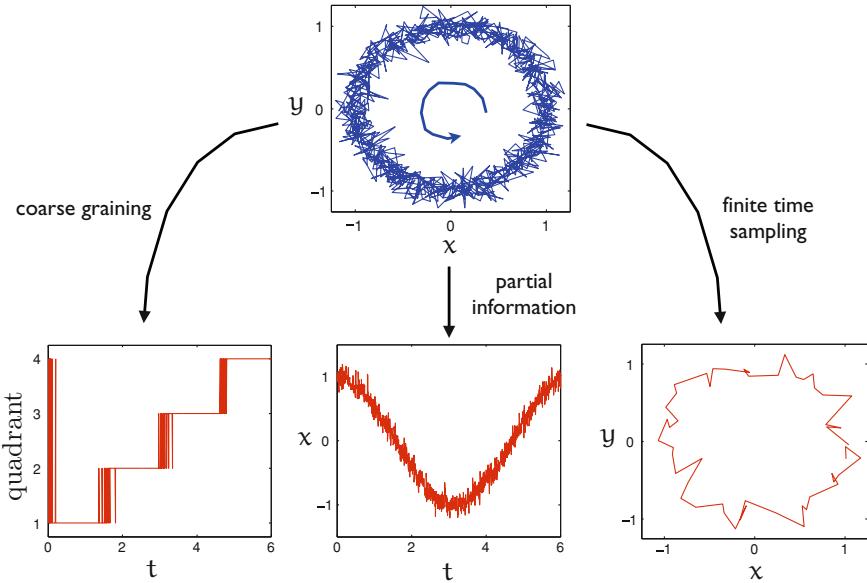


Fig. 2.4 Illustrative example of possible information shortcomings. In the *top panel*, we show the trajectory of a microscopic system that reaches a limit cycle in the xy plane. In the *bottom panels* we show the trajectory of the system when using partial descriptions of the microstate of the system: measuring the quadrant in which the system is at any time (*bottom left panel*), measuring only the variable x at any time (*bottom center panel*) and sampling the trajectory in xy plane every 20 data (*bottom right panel*)

2.4 Dissipation and Irreversibility in the Nonequilibrium Stationary State

The main goal of our work is to explore the existing quantitative relation between dissipation and irreversibility for microscopic systems for the case of nonequilibrium processes that reach a nonequilibrium stationary state (NESS). As we revised in Sect. 1.5.3, a fluid under a constant shear, a gas in a piston that is moved sinusoidally or a molecular motor driven by the ATP hydrolysis are only a few examples of nonequilibrium processes that reach a NESS.

We now proceed to apply the above results to stationary trajectories. Consider a long process where a microscopic system reaches a nonequilibrium stationary state (NESS) after a possible initial transient. In the NESS, the external parameter is held fixed, $\lambda(t) = \lambda$ or it is time-symmetric; the system is kept out of equilibrium due to the existence of baths at different temperatures (a possibility that is included in the hypothesis used in [22] to prove (2.7)) or different chemical potentials, external constant forces, etc. In the steady state, the protocol and its time reversal are identical $\lambda(t) = \bar{\lambda}(t) = \lambda$. For simplicity, we will denote by \mathcal{P} the trajectory probability

density in the stationary state. In the long time limit, $\tau \rightarrow \infty$, we can neglect the contribution of the transient to the entropy production and rewrite (2.7) for the entropy production per unit of time \dot{S}_{prod} in the NESS [20] as

$$\langle \dot{S}_{\text{prod}} \rangle = k \lim_{\tau \rightarrow \infty} \frac{1}{\tau} D \left[\mathcal{P} \left(\{z(t)\}_{t=0}^{\tau} \right) \middle| \mathcal{P} \left(\{\tilde{z}(\tau-t)\}_{t=0}^{\tau} \right) \right]. \quad (2.28)$$

As we showed in Sect. 1.5.3, a similar expression can be obtained from the steady state fluctuation theorem, which holds in the long time limit (1.75),

$$\langle \dot{S}_{\text{prod}} \rangle = k \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \left\langle \ln \frac{\rho_{\tau}(S)}{\rho_{\tau}(-S)} \right\rangle, \quad (2.29)$$

$$= k \lim_{\tau \rightarrow \infty} \frac{1}{\tau} D[\rho_{\tau}(S) || \rho_{\tau}(-S)]. \quad (2.30)$$

where $\rho_{\tau}(S)$ is the probability to observe an entropy production $S_{\text{prod}} = S$ in the interval $[0, \tau]$. Notice that the average in (2.29) is done over all possible values of the entropy production by averaging with $\rho_{\tau}(S)$, which yields the KLD in (2.30). Comparing (2.28) and (2.30) we arrive at

$$D \left[\mathcal{P} \left(\{z(t)\}_{t=0}^{\tau} \right) \middle| \mathcal{P} \left(\{\tilde{z}(\tau-t)\}_{t=0}^{\tau} \right) \right] = D[\rho_{\tau}(S) || \rho_{\tau}(-S)], \quad (2.31)$$

for $\tau \rightarrow \infty$. Consequently, although S is another observable that is obtained as a function of the microstate of the system, the KLD calculated with S yields the same value as the one calculated with full information of the system. Therefore entropy production captures all the information about the time irreversibility of the NESS.

When one does not observe the entire microscopic trajectory $\{z(t)\}_{t=0}^{\tau}$ in (2.28) but the trajectory followed by one or several observables of the system $x(t)$, the KLD only provides a lower bound to the entropy production, as we discussed in Sect. 2.3. Equation (2.31) indicates that the equality is recovered if the observables determine in a unique way the entropy production or the dissipated work.

We are interested in exploring this formula in simulations and experiments where a microscopic system reaches a NESS and we are given a single stationary trajectory or time series produced by the system. In an experimental context, the observables are usually sampled at a finite frequency. The output is then a time series of data or discrete trajectory, $\mathbf{x} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)$, where \hat{x}_i can be the value of a single or several observables of the system. In this case, we are interested in estimating the entropy production *per data* produced by the underlying physical process, which we denote by $\langle \dot{S}_{\text{prod}} \rangle$ in the rest of the chapter. Entropy production per data is related to the KLD rate *per data*, which we define below.

For the sake of simplicity, we now consider random discrete processes, but the discussion below holds also for continuous random processes. Given an infinitely long realization or time series sampled from a random discrete process X_i ($i = 1, 2, \dots$), which can be multi-dimensional, we define by $p(x_1^m)$ the probability that a given sequence of m consecutive data is equal to $x_1^m = (x_1, x_2, \dots, x_m)$. We define

the m -th order KLD for this random process X_i by the distinguishability (KLD) between $p(x_1^m)$ and the probability $p(x_m^1)$ to observe the reverse sequence of data $x_m^1 = (x_m, x_{m-1}, \dots, x_1)$.

$$D_m^X = D[p(x_1^m) || p(x_m^1)] = \sum_{x_1, \dots, x_m} p(x_1^m) \ln \frac{p(x_1^m)}{p(x_m^1)}. \quad (2.32)$$

The *KLD rate* for the process X_i is defined as the growth rate of D_m^X with the number of data,

$$d^X = \lim_{m \rightarrow \infty} \frac{D_m^X}{m}. \quad (2.33)$$

Because of the finite time sampling and given that x may not contain the information of the entropy production (2.31), the chain rule (2.18) implies that the KLD rates bounds from below the entropy production per data

$$\frac{\langle \dot{S}_{\text{prod}} \rangle}{k} \geq d^X. \quad (2.34)$$

The above bound is saturated if the random variable is the microstate of the system $X = \{\mathbf{q}, \mathbf{p}\}$ and the sampling rate is infinite or X determines uniquely the entropy production in the process.

Equation (2.34) is our basic result. It reveals a striking connection between physics and the statistics of a time series. The left-hand side, $\langle \dot{S}_{\text{prod}} \rangle/k$, is a purely physical quantity, whereas the right-hand side, d^X , is a statistical magnitude depending solely on the observed data, but not on the physical mechanism generating the data. This means that if we are given a stationary time series of any random variable X produced by a microscopic system we can bound from below the average entropy production rate in the physical mechanism that generated the data. In particular (2.34) can be used to study the minimum amount of entropy produced in a symmetry restore such as the erasure of a bit, which yields Landauer's principle relating entropy production and logical irreversibility in computing machines [2, 15, 19] as we will see in Chap. 6. Equation (2.34) extends this principle and suggests that we can determine the average dissipation of an arbitrary NESS, even ignoring any physical detail of the system.

2.5 Discrete Systems

We first study the bound to the entropy production provided by the KLD in discrete nonequilibrium stationary processes, namely in Markov chains satisfying detailed balance condition and in Hidden Markov models.

2.5.1 Markov Chains Obeying Local Detailed Balance

We first analyze how the bound (2.34) is expressed for Markovian time series that obey the detailed balance condition by deriving analytical expressions for both entropy production and the KLD rate.

Among all the stochastic processes, *Markov* processes are the most important ones in physics, chemistry and biology [27]. Let us consider a stochastic process of a random (discrete or continuous) variable X . Such a process is said to be *Markovian* if the probability to observe any sequence of n data x_1^n at times $t_1^n = (t_1, t_2, \dots, t_n)$ satisfies the following property,

$$\rho(x_{n+1}, t_{n+1} | x_1, t_1, \dots, x_n, t_n) = \bar{\rho}(x_{n+1}, t_{n+1} | x_n, t_n), \quad (2.35)$$

where the bar $|$ denotes conditioned probability. Therefore, in a Markov process, the probability to observe a value of the process at a given time t_{n+1} only depends on the state of the process one step before, at time t_n . In order to know the probability to observe a sequence x_1^n , we only need to know $p(x_1, t_1)$ and the transition probability $p(x_2, t_2 | x_1, t_1)$ for successive times t_1, t_2 and any value of x_1, x_2 , since the following property derives from (2.35)

$$\rho(x_1, t_1; \dots; x_n, t_n) = p(x_1, t_1) \cdot p(x_2, t_2 | x_1, t_1) \cdots p(x_n, t_n | x_{n-1}, t_{n-1}). \quad (2.36)$$

Nuclei decay, the voltage in an RC circuit, or the motion of a molecular motor in a microtubule can be described by Markov processes. Concerning microscopic physics, the position of a Brownian particle described by the overdamped Langevin equation can be considered as a Markov process [27]. The probability a random *discrete* variable X to take the value x_i at time t , $p_i = p_i(t)$ obeys the *Master equation*,

$$\dot{p}_i = \sum_j k_{j \rightarrow i} p_j - k_{i \rightarrow j} p_i, \quad (2.37)$$

where $k_{i \rightarrow j}$ is the rate from state i to state j , i.e., the number of times that the transition $i \rightarrow j$ occurs per unit of time. Therefore, $k_{i \rightarrow j} \geq 0$ for all i, j . The Master equation (2.37) can be seen as a gain-loss equation for state i . The term $k_{j \rightarrow i} p_j$ accounts for the net incoming probability from any state j to state i and the term $-k_{i \rightarrow j} p_i$ accounts for the losses from state i to any other state j . The net change of p_i due to an exchange with state j is defined as the *current*

$$J_{j \rightarrow i} = k_{j \rightarrow i} p_j - k_{i \rightarrow j} p_i, \quad (2.38)$$

which allows one to write the master equation as a balance equation

$$\dot{p}_i = \sum_j J_{j \rightarrow i}. \quad (2.39)$$

A Markov process reaches a stationary state when the probability to be at any state i does not change in time, and reaches a stationary value, p_i^{ss} . Equivalently, in the stationary state, $\dot{p}_i = 0$, which implies by (2.37) the balance condition $\sum_j J_{j \rightarrow i} = 0$ for all i . If the following (stronger) condition called *detailed balance condition* is satisfied for every pair of states i and j ,

$$\frac{k_{i \rightarrow j}}{k_{j \rightarrow i}} = \frac{p_j^{\text{ss}}}{p_i^{\text{ss}}}, \quad (2.40)$$

the system also reaches a stationary state. Notice however that detailed balance condition on the transition rates $k_{i \rightarrow j}$ *does not* ensure that the stationary state is an equilibrium state. For a physical system that is in contact with a thermal bath at temperature T during the process, and where the potential energy of the state i is V_i , detailed balance condition is written as

$$\frac{k_{i \rightarrow j}}{k_{j \rightarrow i}} = \exp\left(-\frac{V_j - V_i}{kT}\right), \quad (2.41)$$

and it has to be satisfied for every i, j in order to reach a stationary equilibrium state. Equation (2.41) is compatible with stationary probabilities that are weighed by the Boltzmann factor, $p_i^{\text{ss}} = p_i^{\text{eq}} \propto \exp(-\beta V_i / kT)$. In fact, local detailed balance condition is derived by imposing equilibrium stationary probabilities in detailed balance condition (2.40).

When both the state x_1^n and the times t_1^n are discrete, the Markov process is called a *Markov chain* [27]. A simple example of a three-state Markov chain is shown in Fig. 2.5. A physical system jumps at discrete times $t = 0, \Delta t, 2\Delta t, \dots$ between three states labeled by an index $s = 1, 2, 3$. The transition rates $k_{i \rightarrow j}$ and the balance equation for the probability to stay in one of the three states are illustrated in Fig. 2.5.

Let us now consider a Markov chain X_i where the random variable can only take discrete values and we can only sample the value of the random variable at a finite frequency. For a Markov chain, the probability distribution to observe a sequence x_1^m , $p(x_1^m)$, factorizes as $p(x_1^m) = p(x_1)p(x_2|x_1) \cdots p(x_m|x_{m-1})$, which also holds if we reverse the arguments, i.e., for $p(x_m^1)$. Substituting these expressions into equation (2.33), we get

$$d^X = \sum_{x_1, x_2} p(x_1, x_2) \ln \frac{p(x_2|x_1)}{p(x_1|x_2)} = D_2^X - D_1^X = D_2^X, \quad (2.42)$$

since $D_1^X = 0$ when comparing a trajectory and its reverse. Therefore, d^X only depends on transition probabilities if X is a random Markovian process. This expression was also derived for Markov chains in [9].

We now relate d^X in Eq. (2.42) with the entropy production when the system reaches a NESS, because it is in contact with several thermal baths. In this situation, the local detailed balance condition is satisfied. We call $V(x_i)$ is the energy of the

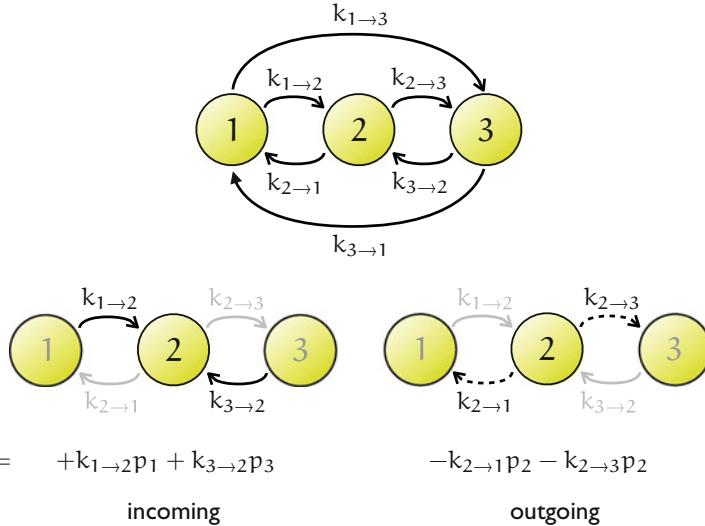


Fig. 2.5 Example of a Markov chain. A system can jump randomly between three states labeled by 1, 2 and 3. When the system is at state i , its state in the next step is $j \neq i$. The transition rates are indicated in the top figure. In the bottom figure, the incoming and outgoing contributions for the time derivative of p_2 are illustrated (solid lines for incoming and dashed for outgoing). In the bottom line the Master equation for the probability to be in state 2 is written

state x_i , and T_{x_1, x_2} is the temperature of the bath that activates the transitions $x_1 \rightarrow x_2$ and $x_2 \rightarrow x_1$. The local detailed balance condition reads in this case

$$\frac{p(x_2|x_1)}{p(x_1|x_2)} = \exp\left(-\frac{V(x_2) - V(x_1)}{k T_{x_1, x_2}}\right). \quad (2.43)$$

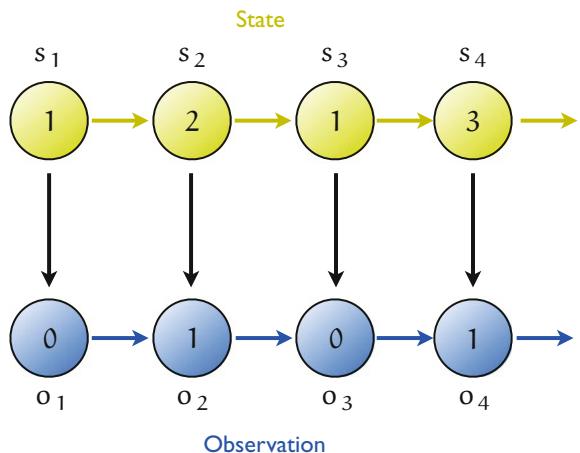
Inserting (2.43) into (2.42),

$$\begin{aligned} d^X &= \sum_{x_1, x_2} p(x_1, x_2) \frac{V(x_1) - V(x_2)}{k T_{x_1, x_2}} \\ &= \sum_{x_1, x_2} p(x_1, x_2) \frac{Q_{x_1, x_2}}{k T_{x_1, x_2}} = \frac{\langle \dot{S}_{\text{prod}} \rangle}{k}, \end{aligned} \quad (2.44)$$

where $Q_{x_1, x_2} = V(x_1) - V(x_2)$ is the heat dissipated to the corresponding thermal bath in the jump $x_1 \rightarrow x_2$, and \dot{S}_{prod} is the total entropy production per data. Therefore, Eq. (2.34) is reproduced, with equality, in the case of a physical system obeying local detailed balance, if we have access to all the variables describing the system. The same conclusion is reached if we induce the NESS by means of non-conservative constant forces.

Equation (2.42) can be explored further by means of the current from the state x_1 to the state x_2 as the net probability flow from x_1 to x_2 , $J_{x_1 \rightarrow x_2} = p(x_1, x_2) - p(x_2, x_1)$.

Fig. 2.6 Example of a Hidden Markov chain. The state of a system s changes in discrete time steps according to the Markovian process described in Fig. 2.5. At every time, the observed state o is obtained according to the following rule: If $s_i = 1$, then $o_i = 0$, else $o_i = 1$. The sequence of observations o_1, o_2, o_3, \dots does not form a Markov chain while the (hidden) state does



If the system is not far from equilibrium the current tends to zero, and the following condition is satisfied $J_{x_1 \rightarrow x_2} \ll p(x_1, x_2)$, yielding

$$\frac{\langle \dot{S}_{\text{prod}} \rangle}{k} = d^X = D_2^X \simeq \sum_{x_1, x_2} \frac{(J_{x_1 \rightarrow x_2})^2}{2p(x_1, x_2)}. \quad (2.45)$$

This expression is well known from linear irreversible thermodynamics [cf. Eq. (2.4)]. Equation (2.45) implies that the time asymmetry of a Markovian process not far from equilibrium is revealed by the currents or probability flows that can be observed. In other words, a Markovian process without flows is time reversible. This is not the case for non-Markovian time series, where irreversibility can show up even in the absence of currents, as shown in the next section.

2.5.2 Hidden Markov Processes

In many experimental situations, a physical process is Markovian at a micro- or mesoscopic level of description, but the observed time series only contain a subset of the relevant observables, being non-Markovian in general. This is the case in biological systems, where one can only register the behavior of some mechanical and maybe a few chemical variables, while most of the relevant chemical variables cannot be monitored. These kind of non-Markovian time series obtained from an underlying Markov process are called *Hidden Markov processes* [23]. If the time and the state of the system are both discrete, the process is called a *Hidden Markov chain*. A simple example of a Hidden Markov chain is shown in Fig. 2.6.

We now show how to calculate the KLD rate between a specific case of hidden Markov chains semi-analytically. We focus on a simple case where an underlying

Markov process is described by two observables X and Y ; however we only observe X whose evolution is described by a hidden Markov chain. The KLD rate for the observable X is

$$\begin{aligned} d^X &= \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{x_1^m} p(x_1^m) \ln \frac{p(x_1^m)}{p(x_m^1)} \\ &= \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{x_1^m} p(x_1^m) \ln \frac{\sum_{y_1^m} p(x_1^m, y_1^m)}{\sum_{y_1^m} p(x_m^1, y_m^1)}. \end{aligned} \quad (2.46)$$

where we have expressed the marginal distribution in X by summing the joint distribution in X, Y to all the possible values that Y can take, $p(x_1^n) = \sum_{y_1^n} p(x_1^n, y_1^n)$. The chain rule ensures that the KLD for the random variable X is smaller than the KLD calculated with full information given by x and Y , $d^X \leq d^{X,Y}$. To compute d^X analytically, it is convenient to write d^X as a difference between two terms,

$$d^X = h_r^X - h^X, \quad (2.47)$$

where

$$\begin{aligned} h^X &= - \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{x_1^m} p(x_1^m) \ln p(x_1^m), \\ &= - \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{x_1^m} p(x_1^m) \ln \sum_{y_1^m} p(x_1^m, y_1^m), \end{aligned} \quad (2.48)$$

is called *Shannon entropy rate*. On the other hand,

$$\begin{aligned} h_r^X &= - \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{x_1^m} p(x_1^m) \ln p(x_m^1), \\ &= - \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{x_1^m} p(x_1^m) \ln \sum_{y_m^1} p(x_m^1, y_m^1), \end{aligned} \quad (2.49)$$

is called the *cross entropy rate*. Since the underlying process is Markovian, the probability distribution in X, Y factorizes (2.36) $p(x_1^m, y_1^m) = p(x_1, y_1)p(x_2, y_2|x_1, y_1) \cdots p(x_m, y_m|x_{m-1}, y_{m-1})$ and both Shannon and cross entropy can be expressed in terms of the trace of a product of random transition matrices \mathbf{T} [12, 14]. These are square $M \times M$ random matrices, where M is the number of values that the variable y can take on, and their entries are given by

$$\mathbf{T}(x_1, x_2)_{y_1 y_2} = p(x_2, y_2|x_1, y_1). \quad (2.50)$$

There are a total number of N^2 transition matrices, where N is the number of values that x can take on. Note the different role played by each variable in this formalism:

x_i are parameters defining the matrix (making \mathbf{T} a random matrix), whereas y_i are subindices of the matrix elements. The Shannon and cross entropy can be expressed in terms of these matrices,

$$h^X = - \lim_{m \rightarrow \infty} \frac{1}{m} \left\langle \ln \text{Tr} \left[\prod_{i=1}^{m-1} \mathbf{T}(x_i, x_{i+1}) \right] \right\rangle, \quad (2.51)$$

$$h_r^X = - \lim_{m \rightarrow \infty} \frac{1}{m} \left\langle \ln \text{Tr} \left[\prod_{i=1}^{m-1} \mathbf{T}(x_{m-i+1}, x_{m-i}) \right] \right\rangle, \quad (2.52)$$

where $\langle \cdot \rangle$ denotes the average over the random process X_i , which is weighted by $p(x_1^m)$. For sufficiently large m , Eqs. (2.51) and (2.52) are self-averaging [12], meaning that we do not need to calculate the average but just compute the trace for a single stationary trajectory. For any sufficiently long time series $\mathbf{x} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)$ with n large, the following expressions converge to $-h$ and $-h_r$ almost surely [12]²

$$\hat{\lambda}^{\mathbf{x}} = \frac{1}{n} \ln \left\| \prod_{i=1}^{n-1} \mathbf{T}(\hat{x}_i, \hat{x}_{i+1}) \right\| \simeq -h^X, \quad (2.53)$$

$$\hat{\lambda}^{\tilde{\mathbf{x}}} = \frac{1}{n} \ln \left\| \prod_{i=1}^{n-1} \mathbf{T}(\hat{x}_{n-i+1}, \hat{x}_{n-i}) \right\| \simeq -h_r^X, \quad (2.54)$$

where $\|\cdot\|$ is any matrix norm that satisfies $\|\mathbf{A} \cdot \mathbf{B}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$. In particular, the trace satisfies this condition for positive matrices. In the context of random matrix theory, $\hat{\lambda}^{\mathbf{x}}$ and $\hat{\lambda}^{\tilde{\mathbf{x}}}$ are known as *maximum Lyapunov characteristic exponents* [5] and measure the asymptotic rate of growth of a random vector when being multiplied by a random sequence of matrices. In practice, we can estimate d^X semi-analytically as

$$\hat{d}^X = \hat{\lambda}^{\mathbf{x}} - \hat{\lambda}^{\tilde{\mathbf{x}}}. \quad (2.55)$$

Here $\hat{\lambda}^{\mathbf{x}}$ and $\hat{\lambda}^{\tilde{\mathbf{x}}}$ are estimated using (2.53) and (2.54) with a single time series \mathbf{x} of size n , following a numerical technique introduced in Ref. [5] to calculate Lyapunov characteristic exponents:

1. We generate a random stationary time series $\mathbf{x} = \{\hat{x}_1^n\}$ and compute the matrices \mathbf{T} analytically.
2. A random unitary vector is multiplied by those matrices in the order given by (2.53) and normalized every l data, keeping track of the normalization factor.
3. The product of these factors divided by n yields $\hat{\lambda}^{\mathbf{x}}$.
4. For $\hat{\lambda}^{\tilde{\mathbf{x}}}$, the same procedure is repeated but using the reversed time series $\tilde{\mathbf{x}} = \{\hat{x}_n^1\}$.
5. The KLD is estimated using Eq. (2.55).

² A sequence of a random variable X , given by X_1, X_2, \dots , is said to converge almost surely to x when the probability that the sequence satisfies $\lim_{n \rightarrow \infty} X_n = x$ is equal to 1.

The technique is semi-analytical since the transition probabilities are known analytically but a single random stationary time series \mathbf{x} is necessary to estimate d^X with the multiplication of n transition matrices that are chosen according to \mathbf{x} .

Let us recall that the estimator $\hat{d}^{\mathbf{x}}$ cannot be applied to empirical time series unless we know the Markov model generating the data. Consequently, it is not useful in practical situations. However, we will use it to test the performance of the estimators of the KLD introduced in the next chapters. On the other hand, one can also analytically estimate of Eqs. (2.51) and (2.52) by using the replica trick, in an analogous way as done in Ref. [7]. The calculation is cumbersome and is explained in Appendix B.3.

References

1. D. Andrieux, P. Gaspard, S. Ciliberto, N. Garnier, S. Joubaud, A. Petrosyan, Phys. Rev. Lett. **98**, 150601 (2007)
2. D. Andrieux, P. Gaspard, Proc. Nat. Acad. Sci. **105**, 9516–9521 (2008)
3. D. Andrieux, P. Gaspard, S. Ciliberto, N. Garnier, S. Joubaud, A. Petrosyan, J. Stat. Mech: Theory Exp. **2008**, P01002 (2008)
4. T. M. Cover, J. A. Thomas, *Elements of Information Theory* (Wiley-interscience, New York, 2006)
5. A. Crisanti, G. Paladin, A. Vulpiani, *Products of Random Matrices in Statistical Physics* (Springer Series in Solid State Sciences) vol. 104 (Springer-Verlag, Berlin, 1993) p. 166
6. G. Crooks, Phys. Rev. E **60**, 2721 (1999)
7. M. De Oliveira, A. Petri, Phys. Rev. E **53**, 2960 (1996)
8. M. Esposito, J.M.R. Parrondo, In preparation (2013)
9. P. Gaspard, J. Stat. Phys. **117**, 599–615 (2004)
10. A. Gomez-Marin, J.M.R. Parrondo, C. Van den Broeck, Phys. Rev. E **78**, 011107 (2008)
11. A. Gomez-Marin, J.M.R. Parrondo, C. Van den Broeck, EPL-Europhys. Lett. **82**, 50002–54000 (2008)
12. T. Holliday, A. Goldsmith, P. Glynn, IEEE Trans. Inf. Theory **52**, 3509–3532 (2006)
13. J. Horowitz, C. Jarzynski, Phys. Rev. E **79**, 21106 (2009)
14. P. Jacquet, G. Seroussi, W. Szpankowski, Theoret. Comput. Sci. **395**, 203–219 (2008)
15. R. Kawai, J.M.R. Parrondo, C.V. den Broeck, Phys. Rev. Lett. **98**, 80602 (2007)
16. D. Kondepudi, I. Prigogine, *From Heat Engines to Dissipative Structures* (Wiley, New York, 1998)
17. S. Kullback, R.A. Leibler, Ann. Math. Stat. **22**, 79–86 (1951)
18. L. Lacasa, A. Nuñez, É. Roldán, J.M.R. Parrondo, B. Luque, Eur. Phys. J. B-Condensed Matter and Complex Systems **85**, 1–11 (2012)
19. R. Landauer, IBM J. Res. Dev. **5**, 183–191 (1961)
20. C. Maes, Séminaire Poincaré **2**, 29–62 (2003)
21. J.M.R. Parrondo, B. Cisneros, Appl. Phys. A Mater. Sci. Process. **75**, 179–191 (2002)
22. J.M.R. Parrondo, C. Van den Broeck, R. Kawai, New J. Phys. **11**, 073008 (2009)
23. L.R. Rabiner, Proc. IEEE **77**, 257–286 (1989)
24. É. Roldán, J.M.R. Parrondo, Phys. Rev. Lett. **105**, 150607 (2010)
25. É. Roldán, J.M.R. Parrondo, Phys. Rev. E **85**, 031129 (2012)
26. U. Seifert, Phys. Rev. Lett. **95**, 40602 (2005)
27. N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, vol. 1 (North holland, 1992)

Chapter 3

Estimating the Kullback–Leibler Divergence

We now investigate how the KLD rate can be estimated from a single empirical stationary trajectory, obtained from a stochastic stationary process whose dynamics is unknown. We assume that the empirical stationary trajectory contains n data of one or several random variables denoted by the letter X . In general, we are given a time series of n data obtained from sampling the random variable X in n successive samplings,

$$\hat{x}_1, \hat{x}_2, \dots, \hat{x}_i, \hat{x}_{i+1}, \dots, \hat{x}_{n-1}, \hat{x}_n, \quad (3.1)$$

where we call \hat{x}_i to the value of the i -th data of an empirical trajectory. We denote such a trajectory as (3.1) by $\mathbf{x} = \{\hat{x}_i\}_{i=1}^n$.

Several recipes to estimate the KLD for stationary time series of both discrete [7, 32–34, 44] and continuous [1, 2, 31, 32, 38, 39] random variables were introduced before our work. However, a full comprehensive analysis for series of both discrete and continuous random variables has not been done until now. In this chapter, we introduce a refinement of the existing methods for estimating the KLD for discrete time series and two new techniques to estimate the KLD for continuous stationary time series. Our estimators are valid in very general situations and only require the data of a single stationary trajectory. The performance of our estimation techniques is analyzed using data from both simulation and experiments in Chaps. 4 and 5, where our estimators are applied to estimate the entropy production in the NESS.

This chapter is organized as follows: In Sect. 3.1 we introduce two estimators for the KLD rate for random discrete stationary processes that are a refinement of two techniques previously introduced in the literature. We extend our study to continuous time series in Sect. 3.2, where we introduce a new technique to estimate the KLD that makes use of autoregressive models. In Sect. 3.3 we show a different approach to estimate the KLD of both continuous and discrete time series using the *horizontal visibility algorithm*.

3.1 Discrete Case

For discrete stationary time series, there are two different strategies to estimate the KLD in the literature: *plug-in* estimators [7, 33, 34], based on empirical counting of sequences of data, and estimators based on compression algorithms [44]. In this section, we introduce a refinement of the these two techniques and analyze their performance is further considered in Chap. 4.

3.1.1 Plug-in Estimators

The simplest approach to estimate the KLD rate is known as the *plug-in* method [38], which consists of an empirical estimation of the probabilities of sequences or *blocks* [34] of m data, $p(x_1^m)$, appearing in Eq. (2.32) [33]. The probability to observe the sequence x_1^m , $p(x_1^m)$, is estimated empirically from simply counting the number of times that x_1^m appears in a single stationary trajectory $\mathbf{x} = (\hat{x}_1, \dots, \hat{x}_n)$ of size n . For a time series of n data, the total number of sequences of m data contained in the series is $n - (m - 1)$. The empirical probability distribution is formally defined as

$$\hat{p}^{\mathbf{x}}(x_1^m) = \frac{1}{n - (m - 1)} \sum_{p=1}^{n-(m-1)} \delta_{\hat{x}_p, x_1} \cdots \delta_{\hat{x}_{p+(m-1)}, x_m}, \quad (3.2)$$

that is, the probability to find a sequence of m data *anywhere* within the sequence \mathbf{x} . In other words, the empirical probability (3.2) is obtained by scanning the series \mathbf{x} in blocks of consecutive m data using a sliding window protocol. An estimation of D_m^X , $\hat{D}_m^{\mathbf{x}}$, is obtained by plugging in the empirical probability distribution into Eq. (2.32):

$$\hat{D}_m^{\mathbf{x}} = D[\hat{p}^{\mathbf{x}}(x_1^m) || \hat{p}^{\mathbf{x}}(x_m^1)] = \sum_{x_1, \dots, x_m} \hat{p}^{\mathbf{x}}(x_1^m) \ln \frac{\hat{p}^{\mathbf{x}}(x_1^m)}{\hat{p}^{\mathbf{x}}(x_m^1)}. \quad (3.3)$$

Note that the probabilities in Eq. (3.3) include the superscript \mathbf{x} to emphasize that they are obtained empirically from a single stationary time series \mathbf{x} and therefore depend on each particular realization. The simplest way to estimate d^X would be by taking $\frac{\hat{D}_m^{\mathbf{x}}}{m}$ for m as large as possible. However, this naive approach is not efficient. The empirical probability $\hat{p}^{\mathbf{x}}(x_1^m)$ —and therefore $\hat{D}_m^{\mathbf{x}}$ —is less accurate as m increases, because the number of possible substrings x_1^m increases exponentially and the statistics shortly becomes poor. It is convenient to find alternative expressions with a fast convergence. It turns out that the slope of $\hat{D}_m^{\mathbf{x}}$ as a function of m ,

$$\hat{D}_m^{\mathbf{x}} = \hat{D}_m^{\mathbf{x}} - \hat{D}_{m-1}^{\mathbf{x}}, \quad (3.4)$$

also converges to the KLD rate but faster than $\frac{\hat{D}_m^x}{m}$. Our plug-in estimator is constructed as the limit

$$\hat{d}^x = \lim_{m \rightarrow \infty} \hat{d}_m^x. \quad (3.5)$$

For a Markovian time series, as shown in Eq. (2.42), the limit is reached for $m = 2$, and using distributions of three or more data we only get redundant information,

$$\hat{d}^x = \hat{d}_2^x = \hat{d}_m^x, \quad (3.6)$$

for any $m > 2$. Therefore, $\hat{d}^x = \hat{d}_2^x$ is an excellent estimator of the KLD, d^X . If \mathbf{x} is a k -th *order Markov chain* (i.e., it is Markovian when considering blocks of k data $\{\hat{x}_1^k\}$), then the limit is reached for $m = k$ [33], i.e.,

$$\hat{d}^x = \hat{d}_k^x = \hat{d}_{k+1}^x = \hat{d}_{k+2}^x = \dots \quad (3.7)$$

In this case, the KLD of blocks of $m < k$ data does not capture the full irreversibility of the process, $\hat{d}_1^x \leq \hat{d}_2^x \leq \dots \leq \hat{d}_k^x = \hat{d}^x$. The convergence of (3.5) is then expected to be fast if a time series can be approximated by a k -th order Markov chain.

If the trajectory \mathbf{x} is sampled from a general non-Markovian process, one needs further information to extrapolate \hat{d}_m^x for $m \rightarrow \infty$, specially when only moderate values of m can be reached. In the examples discussed below, we have found that convergence is well described by the following ansatz, proposed by Schürmann and Grassberger [34] to estimate Shannon entropy rate

$$\hat{d}_m^x \simeq \hat{d}_\infty^x - c \frac{\ln m}{m^\gamma}. \quad (3.8)$$

Here c and γ are parameters that, together with \hat{d}_∞^x , can be obtained by fitting the empirical values of \hat{d}_m^x as a function of m . The fitting parameter \hat{d}_∞^x gives an estimation of the limit (3.5).

3.1.1.1 Insufficient Statistics

This estimation method is efficient as long as there is sufficient statistics in the data, that is, if for every series x_1^m that occurs in the trajectory, its reverse x_m^1 is observed at least once. On the other hand, if we find empirically $\hat{p}^x(x_1^m) \neq 0$ while $\hat{p}^x(x_m^1) = 0$ for at least one case, the argument of the logarithm in Eq. (3.3) diverges, yielding $\hat{d}_m^x = \infty$. We can avoid this divergence by restricting the sum in \hat{D}_m^x to sequences x_1^m whose reverse x_m^1 occur in the time series:

$$\hat{D}_m^x \rightarrow \hat{D}_m^{x*} = \sum_{(x_1^m)^*} \hat{p}^x(x_1^m) \ln \frac{\hat{p}^x(x_1^m)}{\hat{p}^x(x_m^1)}, \quad (3.9)$$

where $(x_1^m)^* = \{x_1^m \mid \hat{p}^{\mathbf{x}}(x_1^m) \neq 0 \text{ and } \hat{p}^{\mathbf{x}}(x_m^1) \neq 0\}$. The KLD calculated with this restriction is a lower bound to its real value, $\hat{D}_m^{\mathbf{x}*} < \hat{D}_m^{\mathbf{x}}$.

A different strategy to avoid the divergence is to artificially bias the empirical probabilities such that all of them become positive. Instead of the observed empirical frequencies, we can use the following biased frequencies [7]

$$\hat{p}^{\mathbf{x}}(x_1^m) = \frac{n^{\mathbf{x}}(x_1^m) + \gamma}{\sum_{x_1^m} [n^{\mathbf{x}}(x_1^m) + \gamma]}. \quad (3.10)$$

Here $n^{\mathbf{x}}(x_1^m)$ is the number of observations of x_1^m in \mathbf{x} and γ is the bias, which is a small number that prevents any of the probabilities to be zero, assigning a probability of order γ/n to sequences that are not observed. The denominator in Eq. (3.10) ensures normalization of $\hat{p}^{\mathbf{x}}(x_1^m)$.

3.1.2 Compression-Based Estimator

Ziv and Merhav introduced in Ref. [44] an estimator of the KLD rate between two probability distributions based on compression algorithms. It consists on slicing or *parsing* stationary discrete time series into smaller parts according to a specific algorithm. The slicing produces a sequence of numbers (often called a *dictionary*) that contains the same data than the original series, but it is divided into subsequences, called *phrases*. These algorithms are called *compression algorithms* because a series of n discrete numbers is always parsed in $m < n$ phrases, therefore the description of the series has been compressed to fewer data.

The estimator by Ziv and Merhav [44] is defined in terms of two concepts which are now described, the compression length of a sequence and the cross parsing length between two different sequences. Given a series $\mathbf{x} = x_1^n$, its *compression length* $c(x_1^n)$ is defined as the number of distinct phrases in which it is parsed using the Lempel-Ziv (LZ) algorithm [43]. The LZ algorithm parses a series sequentially, such that each phrase that is added to the dictionary is the shortest distinct phrase that is not already in the dictionary. For example, let us consider the series $\mathbf{x} = x_1^{11} = (0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0)$. The LZ sequential parsing for this example is as follows: We start with an empty dictionary $\text{Dict} = \{ \}$ and we first store the first element of the series $x_1 = 0$ in the dictionary since it is empty, hence $\text{Dict} = \{0\}$. Then we read the next number, $x_2 = 1$, which is not already in the dictionary, so x_2 is added to the dictionary, $\text{Dict} = \{0|1\}$. The next number in x_1^{11} is $x_3 = 1$, which is already in the dictionary. Then we append to x_3 the next number of the sequence, $x_3^4 = (1, 1)$. This phrase is not in the dictionary and therefore it is parsed, $\text{Dict} = \{0|1|(1, 1)\}$. By doing this for all the series x_1^{11} , we obtain the following dictionary of phrases $\text{Dict} = \{0|1|(1, 1)|(1, 0)|(0, 0)|(1, 1, 0)\}$. The compression length is the number of phrases that the dictionary contains once the series \mathbf{x} is completely parsed, $c(x_1^{11}) = 6$ in this example. We illustrate this simple example in

$$\mathbf{x} = x_1^{11} = (0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0)$$

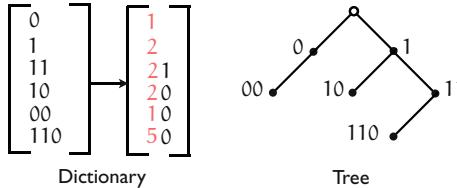


Fig. 3.1 Lempel-Ziv compression algorithm and parsing length. We show how the LZ sequential parsing algorithm is applied to the sequence $\mathbf{x} = x_1^{11} = (0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0)$. The bottom left table is the dictionary of phrases obtained with the LZ algorithm. Every phrase can be encoded with two numbers: (i) A prefix number that indicates in which position on the table is the phrase that equals all the elements of the new word except the last number and (ii) a new number containing the last digit of the new word. Using this procedure, the original sequence can be described by fewer numbers than the number of data of the original sequence. The dictionary can also be built as a binary tree as shown in the bottom right figure. From the parent node, new phrases are added to the tree by adding a node depending on the last value of the new word. A node to the left is added if the last number of the word is 0, and a node to the right otherwise. The compression length of the sequence is the number of phrases in the dictionary, or equivalently, the number of nodes in the tree

Fig. 3.1 where we also show how the dictionary can also be built using a binary tree data structure [4].

The compression length of a stationary time series is related to its Shannon entropy rate [12] in the limit of infinitely long sequences:

$$\lim_{n \rightarrow \infty} \frac{c(x_1^n) \ln c(x_1^n)}{n} = h^X. \quad (3.11)$$

However, since $d^X = h_r^X - h^X$ (2.47), we also require an estimator for h_r^X in order to determine d^X . This is given in terms of another quantity called *cross parsing length*. The cross parsing of a series x_1^n with respect to another sequence z_1^n is obtained by parsing x_1^n looking for the longest phrase that appears anywhere in z_1^n . As an example, let us consider the cross parsing of $\mathbf{x} = x_1^{11} = (0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0)$ with respect to another sequence $\mathbf{z} = z_1^{11} = (1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0)$. The first number in \mathbf{x} is $x_1 = 0$, which is in \mathbf{z} . Therefore we append to x_1 the next number in \mathbf{x} , $x_1^2 = (0, 1)$. This sequence is also somewhere in \mathbf{z} , more precisely it is equal to z_3^4, z_5^6 and z_8^9 , so we append the next item in \mathbf{x} , $x_1^3 = (0, 1, 1)$. Again this sequence is somewhere in \mathbf{z} , $x_1^3 = z_8^{10}$, and it is added to the dictionary, $\text{Dict} = \{(0, 1, 1)\}$ because x_1^4 is not found anywhere in z_1^{11} . We repeat this procedure again starting from x_4 and the resulting dictionary is: $\text{Dict} = \{(0, 1, 1)|(1, 1, 0)|(0, 0, 1, 1, 0)\}$. The cross parsing length is the number of parsed sequences, which in this example is equal to $c_r(x_1^{11}|z_1^{11}) = 3$. The cross-parsing protocol for this example is shown in Fig. 3.2.

Fig. 3.2 Cross parsing of two sequences. We highlight in *green* the phrases of the sequence \mathbf{x} that are parsed according to the sequence \mathbf{z}

$$\mathbf{x} = \mathbf{x}_1^{11} = (0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0)$$

$$\mathbf{z} = \mathbf{z}_1^{11} = (1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0)$$

$$\mathbf{x} = \mathbf{x}_1^{11} = (0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0)$$

$$\mathbf{z} = \mathbf{z}_1^{11} = (1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0)$$

$$\mathbf{x} = \mathbf{x}_1^{11} = (0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0)$$

$$\mathbf{z} = \mathbf{z}_1^{11} = (1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0)$$

In Ref. [44] it is proved that the following quantity tends to the KLD rate between the probability distributions that generated the sequences $\mathbf{x} = \mathbf{x}_1^n$ and $\mathbf{z} = \mathbf{z}_1^n$, which we call p^X and q^Z respectively,

$$\lim_{n \rightarrow \infty} \frac{1}{n} [c_r(x_1^n | z_1^n) \ln n - c(x_1^n) \ln c(x_1^n)] = d(p^X || q^Z). \quad (3.12)$$

We can estimate d^X by using as inputs in the left-hand side of the above equation a stationary time series and its time reverse, that is, using $z_1^n = x_n^1$. The Ziv-Merhav estimator of d^X when using a time series \mathbf{x} of n data is introduced as follows

$$\hat{d}_{ZM}^{\mathbf{x}} = \frac{1}{n} [c_r(x_1^n | x_n^1) \ln n - c(x_1^n) \ln c(x_1^n)], \quad (3.13)$$

which converges to d^X when $n \rightarrow \infty$, although the convergence is slow [44]. This estimator has been used as a measure of distinguishability in several fields such as authorship attribution [10] or biometric identification [11].

When the KLD rate between the probability distributions under consideration is small ($d^X \ll 1$), the estimation given by (3.12) can be even negative [10]. In this case the estimator is said to be *biased* [17]. The estimator gives negative values in some cases because it mixes two types of parsing: the sequential parsing of the trajectory and the cross parsing, which is not sequential. We propose the following correction, which helps to solve this issue and improves the performance of the estimator. We first evaluate (3.13) between different segments of the same trajectory. More precisely, we split \mathbf{x} into two equal parts and apply the original estimator (3.12)

$$\tilde{d}_{ZM}^{\mathbf{x}} = \frac{c_r(x_{n/2}^n | x_1^{n/2}) \ln \frac{n}{2} - c(x_{n/2}^n) \ln c(x_{n/2}^n)}{n/2}. \quad (3.14)$$

If the time series is stationary, the two fragments, $x_1^{n/2}$ and $x_{n/2}^n$, are statistically equivalent and $\tilde{d}_{ZM}^{\mathbf{x}}$ should vanish. However it is usually negative for finite n and exhibits a slow convergence to zero for large n [10]. Then, we define our estimator as

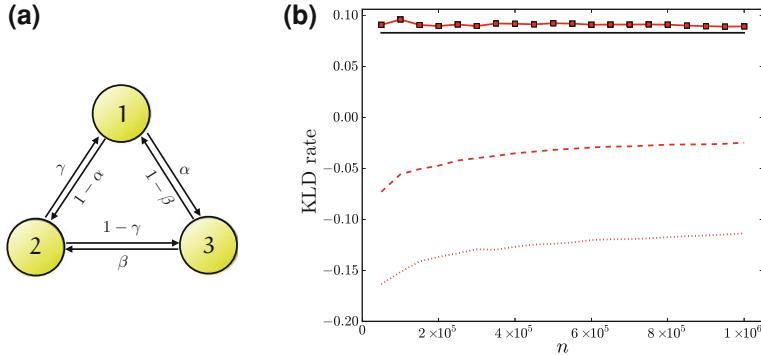


Fig. 3.3 Validation of our KLD compression-based estimator. **a** Sketch of the 3-state toy model used to check the accuracy of our compression estimator (3.15) and **b** comparison between different compression estimators and the analytical value of d^X . The analytical value of d^X for a model with $\alpha = 0.5$, $\beta = 0.7$, $\gamma = 0.6$ ($d^X = d_2^X = 0.08278$) is indicated by the solid black line in the plot. We show the value of the compression estimators obtained from a single stationary time series x_1^n as a function of the length n : the Ziv-Merhav estimator \tilde{d}_{ZM}^X (red dashed line), the bias \tilde{d}_{ZM}^X (red dotted line) and our estimator \hat{d}_c^X (red squares)

$$\hat{d}_c^X = \tilde{d}_{ZM}^X - \tilde{d}_{ZM}^X, \quad (3.15)$$

which still converges to d^X when $n \rightarrow \infty$ but it yields much better results for finite n , as we now discuss with a simple example.

We consider the simplest Markov chain described by the three-state model illustrated in Fig. 3.3a. Trajectories of the model are lists of numbers, 1, 2 or 3, representing the three states of the system. The dynamics is Markovian with transition probabilities given by $p_{1 \rightarrow 2} = 1 - p_{2 \rightarrow 1} = \alpha$, $p_{2 \rightarrow 3} = 1 - p_{3 \rightarrow 2} = \beta$ and $p_{3 \rightarrow 1} = 1 - p_{1 \rightarrow 3} = \gamma$. We call X the random variable describing the state of the system and \mathbf{x} a particular stationary time series, e.g. $\mathbf{x} = (1, 3, 2, 1, 2, \dots)$. This time series is reversible only when the three transition probabilities satisfy the Kolmogorov condition [20], $\alpha\beta\gamma = (1-\alpha)(1-\beta)(1-\gamma)$. In Fig. 3.3b, we compare the value of different compression estimators with the analytical value of d^X as a function of the length of the empirical trajectory n . To study how the KLD compression estimator scales with the number of data we first simulate the model in Fig. 3.3a, generating a single stationary trajectory. The estimators are measured for subsequences containing data from the first to the m -th data of the simulated series, and m is changed in the range shown in the figure. Since the trajectories described by the state of the system are Markovian, d^X only depends on transition probabilities (3.6), $d^X = d_2^X$. We see that Ziv-Merhav estimator \tilde{d}_{ZM}^X fails to estimate d^X accurately when it is small ($d^X \simeq 0.083$) and in some cases gives a negative value. The proposed estimator \hat{d}_c^X , on the other hand, is significantly closer to the analytical result, although slightly overestimates its true value.

We notice that the implementation of the compression estimator is not straightforward. The construction of the LZ dictionary for a sequence requires to look-up in the dictionary every time a new phrase is added. For a series of n data, if the dictionary is constructed as a table like in the bottom left panel of Fig. 3.1, the search becomes slower the more phrases are stored in the dictionary. Typically, the algorithmic cost of this procedure scales like $O(n^2)$, n being the number of data of the parsed time series. To avoid this problem, Benedetto et al. [5] introduced a technique similar to the Ziv-Merhav estimator but using the length of compressed trajectories using commercial zippers instead of the LZ compression algorithm. Their approach speeds-up the estimation of the KLD but they do not provide a strong theoretical argument checking the validity of their estimator. We do not use commercial algorithms but the Ziv and Merhav compression protocol described in [44]. To speed-up the estimation of the KLD, we make use of data structure trees. As shown in the bottom right panel of Fig. 3.1, the construction of the dictionary as a tree has the advantage that searching is faster because the information is branched. Typically, the time to search a phrase after n phrases have been added to the tree scales like $O(\log n)$, which is by far faster than using table lookups of order $O(n^2)$. We now specify step-by-step how we compute our KLD estimator using information trees:

1. A B-tree¹ (binary tree with multiple branches) of the series x_1^n is built according to the LZ sequential parsing.
2. The compression length $c(x_1^n)$ is obtained as the number of nodes of the B-tree.
3. A second B-tree containing all the subsequences in x_n^1 is constructed.
4. The longest sequences in x_1^n that appear in x_n^1 are searched in the B-tree of the reverse sequence.
5. The cross parsing length $c_r(x_1^n|x_n^1)$ is obtained as the number of times in which the search in step 4 is interrupted.
6. The KLD is estimated using the Ziv-Merhav formula (3.13).

Notice that this procedure is done twice as shown in (3.15).

The main advantage of the estimation based on compression technique is that the problem with insufficient statistics arising in the plug-in estimation is avoided using this technique. The compression based estimator is also applied in Chap. 4, where its performance is compared with the plug-in technique.

3.2 Continuous Case

For real-valued stationary trajectories, the KLD rate adopts an analogous expression to the one discussed for the discrete case in Eqs. (2.32) and (2.33). The m -th order KLD for a stationary process described by a continuous random variable X compares the probability densities of sequences of data $\rho(x_1^m)$ and $\rho(x_m^1)$, and is analogous to [cf. (2.32)],

¹ A B-tree is the generalization of a binary tree for non-binary discrete data. Each node can have $m > 2$ children.

$$D_m^X = D[\rho(x_1^m) || \rho(x_m^1)] = \int_{x_1^m} dx_1^m \rho(x_1^m) \ln \frac{\rho(x_1^m)}{\rho(x_m^1)}, \quad (3.16)$$

where $dx_1^m = dx_1 dx_2 \cdots dx_m$, and the integral above is done over the real values that $x_1 \cdots x_m$ can take. The distribution $\rho(x_m^1)$ has to be considered as a distribution $\rho_{\tilde{X}}$ for x_m^1 such that $\rho_{\tilde{X}}(x_1^m) = \rho(x_1^m)$. Therefore, the m -th order KLD is formally expressed as

$$D[\rho_X(x_1^m) || \rho_{\tilde{X}}(x_1^m)] = \int_{x_1^m} dx_1^m \rho_X(x_1^m) \ln \frac{\rho_X(x_1^m)}{\rho_{\tilde{X}}(x_1^m)}. \quad (3.17)$$

The KLD rate per data d^X is defined as in the discrete case [cf. Eq. (2.33)]

$$d^X = \lim_{m \rightarrow \infty} \frac{D_m^X}{m}. \quad (3.18)$$

We are interested in stationary time series of a continuous random variable X that is obtained by sampling X for n times at a regular sampling rate, every Δt . In this situation, it is also interesting to study the KLD per unit of time \dot{D} , defined as

$$\dot{d}^X = \frac{1}{\Delta t} \lim_{m \rightarrow \infty} \frac{D_m^X}{m}. \quad (3.19)$$

Notice that the KLD rate per unit of time yields a lower bound for the average entropy rate per unit of time $\langle \dot{S} \rangle / k \geq \dot{d}^X$ instead of the entropy production per data.

In this section, we first review some of the first attempts in literature to estimate the KLD between random continuous processes. We also introduce a new technique that is applied to estimate the average entropy production rate in the next chapters to both simulations and experimental data.

3.2.1 Symbolization and Nearest-Neighbours

The most naive approach to measure the KLD for continuous random variables is to partition the support of the series x_1^n and x_n^1 into bins of finite size [22, 38]. The m -th order KLD for continuous variables (3.16) can be estimated by plugging in the probabilities of the series to lie within a specific bin

$$\hat{D}_m^X = \sum_{j=1}^K \hat{\rho}_j^X \ln \frac{\hat{\rho}_j^X}{\hat{\rho}_j^{\tilde{X}}}, \quad (3.20)$$

where $\hat{\rho}_j^x = \int_j dx_1^m \hat{\rho}^x(x_1^m)$ is the empirical probability distribution of sequences of m data integrated over the m -dimensional bin j . Equivalently, $\hat{\rho}_j^{\tilde{x}} = \int_j dx_1^m \hat{\rho}^{\tilde{x}}(x_m^1)$ is the same probability distribution but for the time-reversed series. By using this procedure, a continuous series is *symbolized* into a new series that contains the information of the bins of the partition visited by the series. This is a coarse-grained description of the information and therefore only a lower bound only of the actual KLD can be obtained using this method, as discussed in Sect. 2.3.

A better estimation of the KLD can be obtained by binning the continuous trajectories using data-dependent partitions [38]. The most extended technique is called the *nearest-neighbour* method, which is clarified and applied to calculate the mutual information in [22]. In the nearest-neighbour method, the probabilities in (3.20) are replaced by the probabilities to find data within a m -th dimensional ball of radius σ centered in any data x_1^m in the m -dimensional space. The difference with the classical symbolization method lies in the fact that the bins in the nearest-neighbour method are centered in the data and therefore it is easier to have sufficient statistics to estimate the KLD. A comparison between the two techniques for continuous random variables was done in [38, 39].

More recently, another estimation of the KLD rate for continuous data has been introduced in [1, 2]. The concept of *pattern entropy* is defined as the Shannon entropy measured with the probability to observe sequences of a given time τ that lie within a distance σ with respect to a specific pattern or trajectory of time τ . This definition is extended to the cross entropy, and therefore to the KLD. This estimator is used in [1, 2] to check the relation between the dissipation and the KLD between the trajectory distributions in forward and reverse experiments (2.27) by dragging a Brownian particle with an optical trap moving at constant velocity. Their result however is not extended to the stationary states where the control parameter does not change and to the case where only partial information of the system can be sampled.

3.2.2 Autoregressive Models

We describe a novel KLD rate estimator valid for real-valued (continuous) stationary time series that is based on the use of autoregressive models. Our estimator first applies the *same* transformation $\Sigma : \mathbf{x} \rightarrow \boldsymbol{\sigma} = \Sigma(\mathbf{x})$ to the time series $\mathbf{x} = x_1^n$ and the time-reversed series $\tilde{\mathbf{x}} = x_n^1$. The transformation yields two new time series

$$x_1^n \rightarrow \sigma_1^n = \Sigma(x_1^n), \quad (3.21)$$

$$x_n^1 \rightarrow \tilde{\sigma}_1^n = \Sigma(x_n^1). \quad (3.22)$$

We seek a transformation that satisfies the following properties: (i) it is a one-to-one transformation in the trajectory space and (ii) the correlation time of the time series is significantly reduced. These two conditions imply, respectively: (i) The \mathbf{x} and $\boldsymbol{\sigma}$ trajectories contain the same information and therefore the KLD rate calculated using

the statistics of both yields the same value, $\hat{d}^x = \hat{d}^\sigma$, (ii) the KLD rate calculated in the new time series is approximatively the one-time KLD, $\hat{d}^\sigma = \hat{D}_1^\sigma$. If these two conditions are met, the KLD can be measured with the one-time KLD between the new time series,

$$\hat{d}^x = \hat{d}^\sigma \simeq \hat{D}_1^\sigma. \quad (3.23)$$

The main advantage of using such a technique is the possibility to measure the KLD with the one-time divergence (D_1) in the new series, which is easier to calculate than m -block divergences.

A transformation that fulfills the properties outlined above is the residual of an autoregressive (AR) model [6]. To estimate the KLD rate using such a transformation, we first fit the data x_1^n to an autoregressive model of order k and lag l , $AR(k, l)$, which is defined by the following linear model of k parameters A_1, A_2, \dots, A_k

$$x_{1+kl} = \sum_{j=1}^k A_j x_{1+(k-j)l} = A_1 x_{1+(k-1)l} + \cdots + A_k x_1. \quad (3.24)$$

Given a sequence x_1^n that is sampled every l data $x_1^n \rightarrow y_1^{n/l}$, the $AR(k, l)$ model expresses the value of the new sequence $y_1^{n/l}$ at any time as a linear combination of its value at the previous k data. For lag $l = 1$, the $AR(k, l)$ model reduces to the standard $AR(k)$ model described in [6].

Once the parameters of the $AR(k, l)$ model are obtained from the fit, we apply the *residual* transformation Σ to the original trajectory x_1^n . The residual function measures the deviation from the data to the $AR(k, l)$ model. If we call a_j the value of j -th coefficient of the fit of the data to the auto regressive model, we apply the residual function to sequences x_1^{k+1}

$$\begin{aligned} \Sigma(x_1^{(k+1)l}) &\equiv x_{(k+1)l} - \sum_{j=1}^k a_j x_{1+(k-j)l} \\ &= x_{1+kl} - [a_1 x_{1+(k-1)l} + \cdots + a_k x_1]. \end{aligned} \quad (3.25)$$

By applying this function to the trajectory x_1^n we get a sequence of n numbers denoted by σ_1^n . The i -th component of the resulting series is obtained by applying the residual function to the substring $x_i^{i+(k+1)l}$. As an example, the first two numbers of the sequence σ_1^n are

$$\begin{aligned} \sigma_1 &= \Sigma(x_1^{(k+1)l}) = x_{1+kl} - [a_1 x_{1+(k-1)l} + \cdots + a_k x_1], \\ \sigma_2 &= \Sigma(x_2^{(k+1)l+1}) = x_{2+kl} - [a_1 x_{2+(k-1)l} + \cdots + a_k x_2]. \end{aligned}$$

We apply the same function to the time reversed series x_n^1 , obtaining a different series of numbers $\tilde{\sigma}_1^n$. Once we have transformed the original series into the new

ones $x_1^n \rightarrow \sigma_1^n$ and $x_n^1 \rightarrow \tilde{\sigma}_1^n$, we compute the one-data KLD from the PDF of the stationary trajectories σ_1^n and $\tilde{\sigma}_1^n$, following Eq. (3.23).

The KLD in the r.h.s. measures the distinguishability between the probability density to observe σ_1^n in the series and to observe $\tilde{\sigma}_1^n$ in the time-reversed series. The simplest estimation to this number is to measure the one-time KLD rate for σ -sequences,

$$\hat{D}_1^\sigma = \int d\sigma' \hat{\rho}^\sigma(\sigma') \ln \frac{\hat{\rho}^\sigma(\sigma')}{\hat{\rho}^{\tilde{\sigma}}(\sigma')} . \quad (3.26)$$

Equivalently, if the data sampling is done at a regular frequency $f_{\text{act}} = 1/\Delta t$, one can estimate the KLD rate per unit of time,

$$\dot{\hat{D}}_1^\sigma = \frac{1}{\Delta t} \int d\sigma' \hat{\rho}^\sigma(\sigma') \ln \frac{\hat{\rho}^\sigma(\sigma')}{\hat{\rho}^{\tilde{\sigma}}(\sigma')} . \quad (3.27)$$

These estimators contain less information than the one concerning the probability of sequences of $n > 1$ symbols, since they only take into account one-time correlations, therefore the chain rule implies that $\hat{D}_1^\sigma \leq \hat{d}^\sigma$. The equality is saturated only when the time series σ_1^n and $\tilde{\sigma}_1^n$ are completely uncorrelated, and the irreversibility is captured in the one-time statistics. In general, the estimators given by (3.26) and (3.27) bound from below the KLD rate (per data or per time, respectively) of the original sequences, and the bound gets tighter the more the correlation time is reduced after the transformation.

In practice, we have to estimate Eqs. (3.26) and (3.27) from the data of a single stationary trajectory. We use the simplest approach which consists on partitioning the data σ_1^n and $\tilde{\sigma}_1^n$ into M bins that are equally spaced by a quantity $\Delta\sigma$ and counting the number of times that the value of the series lies within the bounds of the bin. We estimate the probability density in the i -th bin by $\hat{\rho}_i^\sigma = n_i/(n\Delta\sigma)$, being n_i the number of times that σ_1^n lies in bin i . Similarly, $\hat{\rho}_i^{\tilde{\sigma}} = n'_i/(n\Delta\sigma)$, being n'_i the number of times that $\tilde{\sigma}_1^n$ lies in bin i . The estimator of \hat{D}_1^σ reads

$$\dot{\hat{D}}_1^\sigma = \frac{\hat{D}_1^\sigma}{\Delta t} = \frac{1}{\Delta t} \sum_{i=1}^M \Delta\sigma \hat{\rho}_i^\sigma \ln \frac{\hat{\rho}_i^\sigma}{\hat{\rho}_i^{\tilde{\sigma}}} . \quad (3.28)$$

We will use this formula to estimate the KLD rate of a continuous system in Chap. 5. We will denote by \hat{d}_{AR}^x the estimator of the KLD rate *per unit of time* for x -sequences using auto regressive models, $\hat{d}_{\text{AR}}^x = \dot{\hat{D}}_1^\sigma$, and therefore

$$\hat{d}_{\text{AR}}^x \equiv \frac{1}{\Delta t} \sum_{i=1}^M \Delta\sigma \hat{\rho}_i^\sigma \ln \frac{\hat{\rho}_i^\sigma}{\hat{\rho}_i^{\tilde{\sigma}}} . \quad (3.29)$$

In Chap. 5 we use this estimator of the KLD rate of stationary continuous time series to distinguish between active and passive biological processes and to estimate the average energy consumption rate of the biological process in the NESS.

3.3 Visibility Technique

We now introduce a novel technique to estimate the KLD of both continuous and discrete time series that is described in our work [26]. The technique makes use of the *horizontal visibility algorithm*, which is a method that maps a time series to a directed network according to a geometric criterion. The KLD between forward and backward trajectories is estimated by the KLD between the *in* and *out* degree distributions of the associated graph.

In this section, we define the estimator based on the visibility algorithm and also show that the estimator correctly distinguishes between reversible and irreversible stationary time series of continuous random variables, including analytical and numerical studies of its performance for: (i) reversible stochastic processes (uncorrelated and Gaussian linearly correlated), (ii) reversible (conservative) and irreversible (dissipative) chaotic maps, and (iii) dissipative chaotic maps in the presence of noise.

The visibility method can also be applied to discrete data, despite in general it underperforms the plug-in method, as we show in Chap. 4, where the visibility technique is used to measure the time irreversibility of a discrete flashing ratchet in an asymmetric potential.

3.3.1 The Method

3.3.1.1 Complex Networks

A network is a collection of links and nodes. Any system where its different parts are coupled each other can be described by a network [13]. In physics, ordered networks such as the structure of solids [3] or random networks like the Ising model with random coupling matrix [35] have been studied at first. More recently, Watts and Strogatz [40] proposed a new kind of networks whose properties lie between those of the ordered and random networks and were called *small-world* networks. Such a network has the features of both ordered and random networks. First, one can reach distant nodes after a few number of jumps between connected nodes, like in a random graph. Second, there are clusters within the network like in ordered graphs. During the last years, small-world networks have been found to reproduce many situations

in the real world, such as metabolic networks, the world wide web, food webs,² epidemic spreading or social ties [30].

First quantitative analysis on complex networks were done in the context of random graph theory by Erdős and Renyi [15, 16]. For the scope of this work, only two simple concepts of graph theory have to be taken into consideration. A key concept in complex networks is the *degree* of a node, k , which is defined as the number of links that connect to a specific node. The degree distribution, $p(k)$, is the distribution of the degree measured over all the nodes of the network, or the probability of an arbitrary node to have degree k (i.e., k links). In some cases, links have a specific direction. For example in a food web [19], predators eat the prey, which can be represented by a node (predator) connected with a link that is directed to another node (prey). A network with directed links is called a directed network, and the degree at every node can be expressed as the sum between nodes that enter into the node k_{in} plus nodes that exit the node k_{out} , yielding $k = k_{\text{in}} + k_{\text{out}}$ for every node in the network. An ingoing and outgoing degree distributions, $p(k_{\text{in}})$ and $p(k_{\text{out}})$, can be defined in a network by measuring the average ingoing and outgoing degree over all the nodes in the network.

3.3.1.2 The Horizontal Visibility Graph

Visibility algorithms map time series to networks according to specific geometric criteria [23, 28]. The general purpose of such methods is to accurately map the information stored in a time series into an alternative mathematical structure. This mapping enables to employ the powerful tools of graph theory to characterize time series from a different viewpoint, bridging the gap between nonlinear time series analysis, dynamical systems, and graph theory [14, 24, 27, 29, 42].

We now focus on a specific subclass of the visibility algorithms called *horizontal visibility algorithm*, firstly proposed in [28]. The algorithm assigns to each data x_i of the series x_1^n to a node in the *horizontal visibility graph* (HVG). Then, two nodes i and j are connected in the graph if one can draw a *horizontal* line in the time series joining x_i and x_j that does not intersect any intermediate data height (see Fig. 3.4). Hence, the nodes i and j are connected if the following geometrical criterion

$$x_i, x_j > x_k, \quad \forall k \mid i < k < j , \quad (3.30)$$

is fulfilled within the time series. In [25, 28], stochastic and chaotic series are characterized using the horizontal visibility algorithm, and the first steps for a mathematically sound characterization of horizontal visibility graphs have been established

² Food webs are networks that represent feeding connections in an ecological community. Nodes represent different species and two nodes are connected if one of the representing species eats the other one.

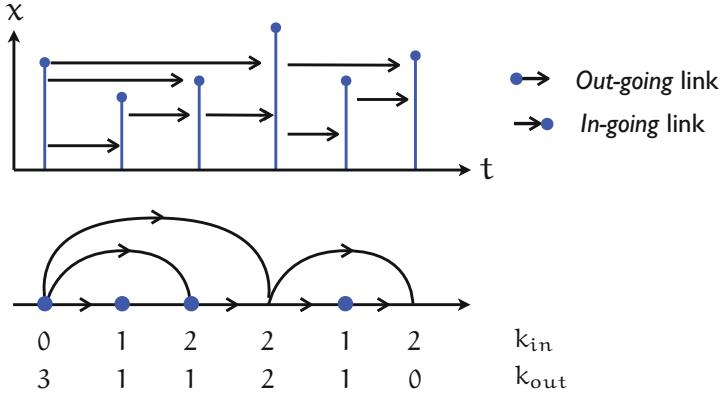


Fig. 3.4 Graphical illustration of the directed horizontal visibility algorithm. In the *top panel* we plot a sample time series $x(t)$ of a continuous random variable X . A bar of height $x(t_i)$ is plotted for every data of the series. Arrows in the *top panel* indicate the data seen by the series at every time t_i according to the horizontal visibility criterion. The associated directed horizontal visibility graph is plotted *below*. In this graph, a node is associated to every data of the series. Each node has an ingoing degree k_{in} , which accounts for the number of links with *past* nodes, and an outgoing degree k_{out} , which accounts for the number of links with *future* nodes. The asymmetry of the resulting graph can be captured in a first approximation through the invariance of the outgoing (or ingoing) degree series under time reversal

in [18]. The method can also be used to characterize not only time series but generic nonlinear discrete dynamical systems, sharing similarities with the theory of symbolic dynamics [29].

3.3.1.3 Directed HVg

So far in the literature the family of visibility graphs are undirected, as visibility does not have a predefined temporal arrow. However, as already suggested in [23], such a directionality can be made explicit by making use of directed networks [30]. We address such directed version, defining a *Directed* Horizontal Visibility graph (DHVg) as a HVg, where the degree $k(t_i)$ of the node corresponding to the data $x_i = x(t_i)$ with $t_i = i \Delta t$, is now splitted in an *ingoing* degree $k_{\text{in}}(t_i)$, and an *outgoing* degree, such that $k(t_i) = k_{\text{in}}(t_i) + k_{\text{out}}(t_i)$. In this case, the ingoing degree $k_{\text{in}}(t_i)$ is defined as the number of links of node t_i with other *past* nodes associated with data in the series, or equivalently, the number of data that see x_i using the horizontal visibility algorithm. Conversely, the outgoing degree $k_{\text{out}}(t_i)$, is defined as the number of links with *future* nodes, or the number of data that can be seen by x_i . The construction of the DHVg associated to a time series is illustrated in Fig. 3.4.

We define the *in* and *out* (or ingoing and outgoing) degree distributions of a DHVg as the probability distributions of k_{in} and k_{out} of the graph which we call $p_{\text{in}}(k) \equiv p(k_{\text{in}} = k)$ and $p_{\text{out}}(k) \equiv p(k_{\text{out}} = k)$, respectively.

3.3.1.4 Estimating the KLD Using Directed Horizontal Visibility Graphs

We propose a new estimator of the KLD valid for both discrete (2.33) and continuous (3.18) random variables. We claim that the time irreversibility can be captured by the KLD between the *in* and *out* degree distributions, $p_{\text{in}}(k)$ and $p_{\text{out}}(k)$. If needed, higher order measures can be used, such as the KLD between the *in* and *out* degree-degree distributions, $p_{\text{in}}(k, k')$ and $p_{\text{out}}(k, k')$. These are defined as the *in* and *out* joint degree distributions of a node and its first neighbors, describing the probability of an arbitrary node whose neighbor has degree k' to have degree k [30].

As we will show in several examples, the information of the outgoing degree distribution k_{out} is sufficient to distinguish between reversible and irreversible stochastic stationary series which are real-valued but discrete in time x_1^n . We compare the outgoing degree distribution in the actual series $P_{k_{\text{out}}}(k|x_1^n) = p_{\text{out}}(k)$ with the corresponding probability in the time-reversed series, which is equal to the probability distribution of the ingoing degree in the actual process $P_{k_{\text{out}}}(k|x_n^1) = p_{\text{in}}(k)$. The KLD between these two distributions is

$$D[p_{\text{out}}(k)||p_{\text{in}}(k)] = \sum_k p_{\text{out}}(k) \ln \frac{p_{\text{out}}(k)}{p_{\text{in}}(k)}, \quad (3.31)$$

which vanishes if and only if the outgoing and ingoing degree probability distributions of a time series are identical, $p_{\text{out}}(k) = p_{\text{in}}(k)$, and it is positive otherwise. We define a new estimator of the KLD of a random process X from the empirical in and out degree distributions measured over the DHVg contracted from a time series \mathbf{x} . The empirical degree distributions associated to a time series \mathbf{x} are denoted by $p_{\text{out}}^{\mathbf{x}}(k)$ and $p_{\text{in}}^{\mathbf{x}}(k)$. We define the *visibility estimator* of the KLD, \hat{d}_{vis} , as

$$\hat{d}_{\text{vis}}^{\mathbf{x}} = D[p_{\text{out}}^{\mathbf{x}}(k)||p_{\text{in}}^{\mathbf{x}}(k)] = \sum_k p_{\text{out}}^{\mathbf{x}}(k) \ln \frac{p_{\text{out}}^{\mathbf{x}}(k)}{p_{\text{in}}^{\mathbf{x}}(k)}. \quad (3.32)$$

The visibility estimator has the advantage that it maps a real-valued time series into two probability distributions, $p_{\text{in}}(k)$ and $p_{\text{out}}(k)$, that have discrete support. In this case, the visibility estimation of the KLD (3.31) is much easier to calculate than the KLD for the original real-valued time series, as shown in the preceding Chapters. The first attempts to estimate the KLD for continuous data, which were discussed in Sect. 3.2.1, generally proceed by first making a (somewhat *ad hoc*) local symbolization of the series, coarse-graining each of the series data into a symbol (typically, an integer) from an ordered set. Then, they subsequently perform a statistical analysis of word occurrences (where a word of length n is simply a concatenation of n symbols) from the forward and backwards symbolized series. The visibility estimator can also be considered as a symbolization since the series is mapped to a discrete series $\{k_{\text{out}}(t_i)\}_{i=1,\dots,n}$ and $\{k_{\text{in}}(t_i)\}_{i=1,\dots,n}$. Consequently, the estimator bounds from below the actual KLD of the time series,

$$\hat{d}_{\text{vis}}^{\mathbf{x}} \leq \hat{d}^{\mathbf{x}}. \quad (3.33)$$

However, at odds with other methods, here the symbolization process (i) lacks *ad hoc* parameters (such as number of symbols in the set or partition definition), and (ii) it takes into account *global* information: each coarse-graining $x_i \rightarrow (k_{\text{in}}(t_i), k_{\text{out}}(t_i))$ is performed using information from the whole series, according to the mapping criterion (3.30). Hence, this symbolization naturally takes into account multiple scales, which is desirable if we want to tackle complex signals [8, 9].

3.3.2 Validation of the Method

We now explore, analytically and numerically, the efficiency of the visibility estimator of the KLD to distinguish between reversible and irreversible stochastic stationary time series. We first analyze reversible time series generated from linear stochastic processes, which yield $\hat{d}_{\text{vis}}^{\mathbf{x}} = 0$. This analysis is extended to chaotic signals, where our method distinguishes between dissipative ($\hat{d}_{\text{vis}}^{\mathbf{x}} > 0$) and conservative ($\hat{d}_{\text{vis}}^{\mathbf{x}} = 0$) chaos, and we show the robustness of the KLD of chaotic signals polluted with noise. In Chap. 4, the method is applied to a physical system in which the time irreversibility can be controlled by tuning the value of a parameter of the system. For that system, we find that the method can, not only distinguish, but also quantify the degree of irreversibility.

3.3.2.1 Random Processes

For simplicity, we first study random uncorrelated processes. A time series of such a process is an *i.i.d.* (independent and identically distributed) sequence of numbers. As a case study, we generate a series of real-valued random $n = 10^6$ numbers distributed according to a uniform distribution that ranges from 0 to 1, $U[0, 1]$. The first data of the series are shown in the top panel of Fig. 3.5. The *in* and *out* degree distributions associated to the series of $n = 10^6$ data are shown in the bottom panel of Fig. 3.5. The distributions are, besides finite size effects, equivalent. The deviation is measured through their KLD (see Table 3.1), yielding $\hat{d}_{\text{vis}}^{\mathbf{x}} \sim 10^{-6}$. We will show that this number is small enough to be considered as a computational zero, $\hat{d}_{\text{vis}}^{\mathbf{x}} \simeq 0$, which indicates the reversible character of the series.

To support our claims, we analytically show that the *in* and *out* degree distributions for *any* uncorrelated random process are indeed identical in the limit of infinite size series. In particular, we show that the *in* and *out* degree distributions of the DHVg of an infinite sequence of *i.i.d* random variables extracted from a continuous probability density $\rho(x)$ are

$$P_{\text{in}}(k) = P_{\text{out}}(k) = \left(\frac{1}{2}\right)^k, \quad k = 1, 2, 3, \dots \quad (3.34)$$

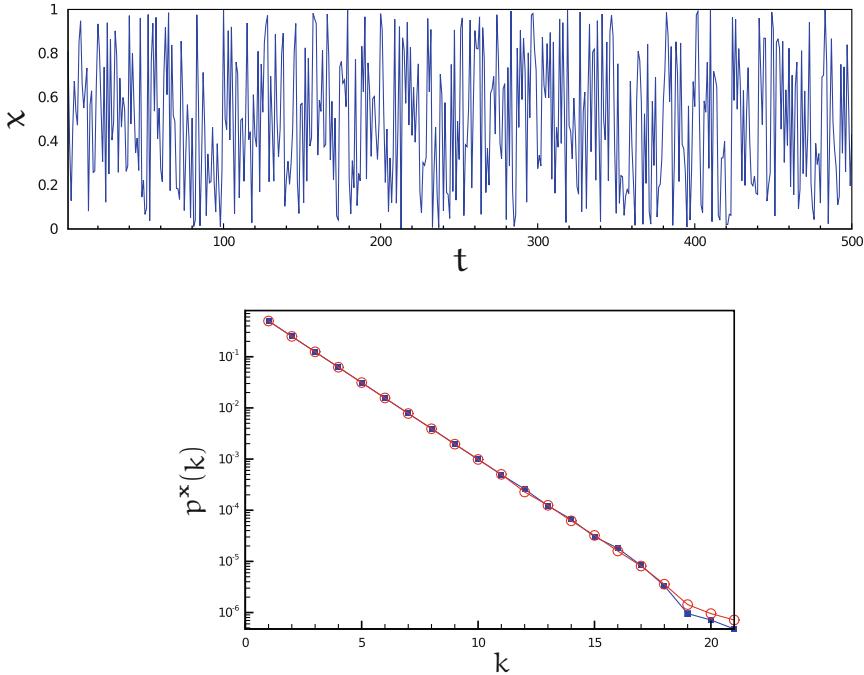


Fig. 3.5 Ingoing and outgoing degree distributions for a random uncorrelated time series. We generate a series of $n = 10^6$ data according to a uniform distribution $U[0, 1]$. *Top* A sample of 500 data of the series. *Bottom* The *in* (blue filled squares) and *out* (red open circles) empirical degree distributions of the DHVg associated to the random series of 10^6 data points

The proof of this result is shown in Appendix C.1. We notice that the *in* and *out* degree distributions are identical, and therefore $d_{\text{vis}} = 0$. The random uncorrelated series are reversible at the level of description of the *in* and *out* degree distributions of the DHVg. The result is valid for any underlying probability density $\rho(x)$, it holds not only for Gaussian or uniformly distributed random series, but for any series of an *i.i.d.* random process.

As a further validation, we also study linearly correlated stochastic processes as additional examples of reversible dynamics [41]. We use the *minimal subtraction procedure* described in [25] to generate such correlated series. This method is a modification of the standard Fourier filtering method, which consists in filtering a series of uncorrelated random numbers in Fourier space.

We study both short and long-time correlated random series. We generate time series of $n = 10^6$ data of an *Ornstein-Uhlenbeck process* with an exponentially-decaying autocorrelation $C(j) = \langle x_i x_{i+j} \rangle \sim \exp(-j/\tau)$ and also using a power-law decaying autocorrelation $C(j) \sim j^{-\gamma}$. In our numerical examples, we set $\tau = 1.0$ and $\gamma = 2.0$ in data units. In Table 3.1, we show that \hat{d}_{vis}^x is very close to zero for

Table 3.1 Values of the KLD visibility estimator for different stochastic processes

Series description	$\hat{d}_{\text{vis}}^{\mathbf{x}}$	$\hat{d}_{\text{vis,dd}}^{\mathbf{x}}$
<i>Reversible stochastic processes</i>		
$U[0, 1]$ uncorrelated	$3.88 \cdot 10^{-6}$	$2.85 \cdot 10^{-4}$
Ornstein-Uhlenbeck ($\tau = 1.0$)	$7.82 \cdot 10^{-6}$	$1.52 \cdot 10^{-4}$
Long-range (power law) correlated stationary process ($\gamma = 2.0$)	$1.28 \cdot 10^{-5}$	$2.0 \cdot 10^{-4}$
<i>Dissipative chaos</i>		
Logistic map ($\mu = 4$)	0.377	2.978
α map ($\alpha = 3$)	0.455	3.005
α map ($\alpha = 4$)	0.522	3.518
Henon map ($a = 1.4, b = 0.3$)	0.178	1.707
Lozi map	0.114	1.265
Kaplan-Yorke map	0.164	0.390
<i>Conservative chaos</i>		
Arnold cat map	$1.77 \cdot 10^{-5}$	$4.05 \cdot 10^{-4}$

Empirical values of the irreversibility measure associated to the degree distribution $\hat{d}_{\text{vis}}^{\mathbf{x}}$ and the degree-degree distribution $\hat{d}_{\text{vis,dd}}^{\mathbf{x}}$ respectively, for the visibility graphs associated to series of 10^6 data generated from reversible and irreversible stochastic processes

these series. The deviation of the KLD estimator from zero is studied in Fig. 3.6. We sample series of size $n' < n$ from an original series of $n = 10^6$ data and measure the KLD estimator for every series of size n' . In Fig. 3.6 we plot the value of \hat{d}_{vis} as a function of the number of data of the subsequence, showing that the KLD estimator scales with the number of data, n , like $1/n$, and therefore it tends to 0 when n is large. Consequently, we also predict reversibility ($d_{\text{vis}} = 0$) for both short (Ornstein-Uhlenbeck) and long-time (power law decaying, data not shown) correlated random series.

3.3.2.2 Chaos

Chaos is defined as aperiodic long-term behavior in a deterministic system that exhibits *sensitive dependence* on initial conditions [37]. We remark that chaotic series are not random but deterministic, and they do not converge to stable points or periodic orbits. Sensitive dependence means that the separation between two trajectories with similar initial conditions increases exponentially with time. *Conservative* chaotic systems do conserve the distance between two trajectories whereas *dissipative* chaotic systems do not. We now apply the visibility estimator of the KLD to examples of both conservative and dissipative chaotic maps found in [36], showing that the KLD distinguishes between these two types of chaotic systems.

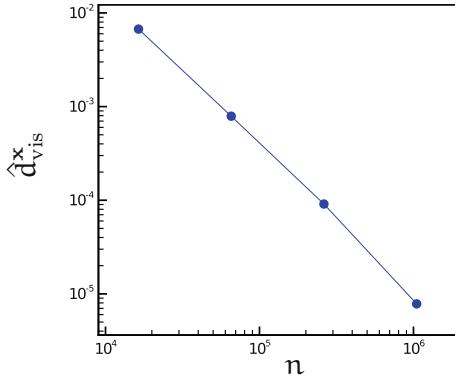


Fig. 3.6 Scaling of the visibility estimator of the KLD for an Ornstein-Uhlenbeck process. Log-log plot of \hat{d}_{vis}^x as a function of the number of data n of a subsequence obtained by sampling n data of a time series of 10^6 data. The original time series is generated according to an Ornstein-Uhlenbeck process of correlation $C(t) = \langle x_i x_{i+j} \rangle \sim \exp(-j/\tau)$, with $\tau = 1.0$. The blue dots are the result of an ensemble average over several realizations

We first study the Logistic map defined by

$$x_{i+1} = \mu x_i (1 - x_i), \quad (3.35)$$

which models the logistic equation for population growth [37]. This model exhibits chaotic behavior for $\mu \geq \mu_\infty = 3.569946\dots$. We focus on the Logistic map at $\mu = 4$, which produces chaotic time series as shown in the top panel of Fig. 3.7, where a sample of 500 data of a series of $n = 10^6$ data is shown. The *in* and *out* degree distributions of the associated DHVg of the series are clearly distinct as shown in the bottom left panel of Fig. 3.7 [cf. Fig. 3.5]. In the bottom right panel of Fig. 3.7 we show the finite size scaling of the estimator \hat{d}_{vis}^x , which converges to an asymptotic value that clearly deviates from zero, $\hat{d}_{vis}^x = 0.377$ (see Table 3.1). We also prove analytically in Appendix C.2 that the *in* and *out* degree distributions are different for the logistic map and therefore $d_{vis} > 0$.

For completeness, we consider other examples of dissipative chaotic systems analyzed in [36]:

1. *The α -map:* $x_{t+1} = 1 - |2x_t - 1|^\alpha$, which reduces to the Logistic and tent maps in their fully chaotic region for $\alpha = 2$ and $\alpha = 1$ respectively. We analyze this map for $\alpha = 3, 4$.
2. *The 2D Hénon map:* $x_{t+1} = 1 + y_t - ax_t^2$, $y_{t+1} = bx_t$, in the fully chaotic region ($a = 1.4$, $b = 0.3$).
3. *The Lozi map:* a piecewise-linear variant of the Hénon map given by $x_{t+1} = 1 + y_n - a|x_t|$, $y_{t+1} = bx_t$ in the chaotic regime ($a = 1.7$ and $b = 0.5$).
4. *The Kaplan-Yorke map:* $x_{t+1} = 2x_t \bmod(1)$, $y_{t+1} = \lambda y_t + \cos(4\pi x_t) \bmod(1)$.

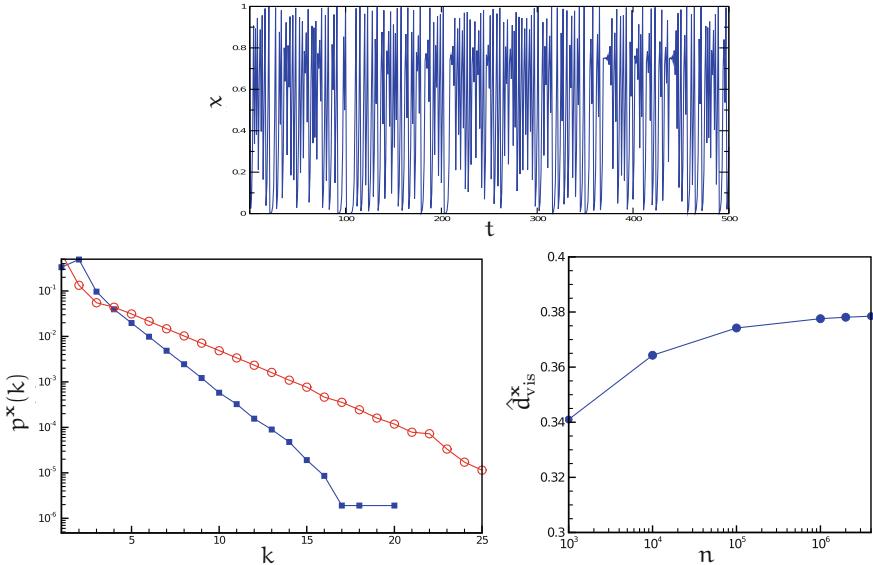


Fig. 3.7 KLD visibility estimator for a logistic map in the chaotic region. *Top* A sample of 500 data of a chaotic time series of $n = 10^6$ data extracted from a fully chaotic Logistic map $x_{t+1} = 4x_t(1 - x_t)$. *Bottom, left* The *in* (blue filled squares) and *out* (red open circles) empirical degree distributions of the DHVg associated to the series. *Bottom, right* Log-log plot of \hat{d}_{vis}^x of the graph associated to a fully chaotic Logistic map $x_{t+1} = 4x_t(1 - x_t)$, as a function of the series size n (dots are the result of an ensemble average over different realizations)

We generate stationary time series using these maps and sample data once the system is in the corresponding attractor. In the case of the 2D Hénon map, the Lozi map and the Kaplan-Yorke map, we only sample the value of one of the two variables, x . In Table 3.1 we show the value of \hat{d}_{vis}^x for these series. In every case, we find an asymptotic positive value, in agreement with the conjecture that dissipative chaos is indeed time irreversible.

Finally, we also consider the *Arnold cat map*: $x_{t+1} = x_t + y_t \bmod(1)$; $y_{t+1} = x_t + 2y_t \bmod(1)$. Unlike the previously discussed maps, this is an example of a *conservative* (measure-preserving) chaotic system with integer Kaplan-Yorke dimension [36].³ We study the time irreversibility of the Arnold cat map by measuring the visibility estimator of the KLD from the data of a series of $n = 10^6$ data obtained using this map. In the top panel of Fig. 3.8 we show a sample of 500 data of the series generated by the Arnold cat map. The *in* and *out* degree distributions of the DHVg are indistinguishable as shown in the bottom left panel of Fig. 3.8. In the bottom right

³ The map has two Lyapunov exponents which coincide in magnitude. $\lambda_1 = \ln(3 + \sqrt{5})/2 = 0.9624$ and $\lambda_2 = \ln(3 - \sqrt{5})/2 = -0.9624$. This implies that the amount of information created in the forward process (λ_1) is equal to the amount of information created in the backwards process ($-\lambda_2$), therefore the process is time reversible.

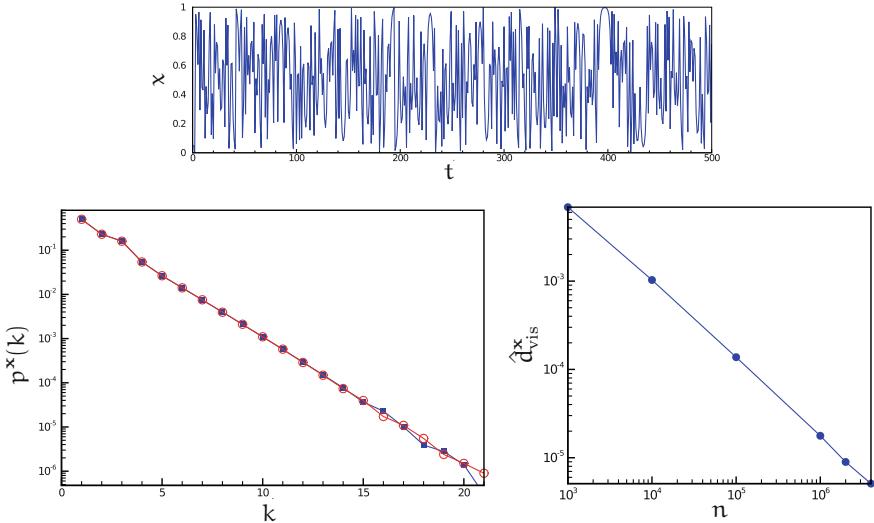


Fig. 3.8 KLD visibility estimator for the Arnold cat map. *Top* A sample chaotic time series (500 data points) extracted from the (chaotic and conservative) Arnold cat map. *Bottom, left* The *in* (blue filled squares) and *out* (red open circles) degree distributions of the DHVg associated to the chaotic series of 10^6 data points. *Bottom, right* Log-log plot of \hat{d}_{vis}^x of the graph associated to the Arnold cat map as a function of the series size n . Dots are the result of an ensemble average over different realizations

panel of Fig. 3.8 we show that d_{vis}^x asymptotically tends to zero with series size (see Table 3.1). This correctly suggests that albeit chaotic, the map is statistically time reversible at the level of description provided by the DHVg.

In summary, our numerical study shows that dissipative chaos is time irreversible, since the KLD visibility estimator tends to a positive value $d_{\text{vis}}^x > 0$ for all the examples shown in Table 3.1. On the other hand, the statistics of the DHVg of the Arnold cat map suggest that conservative chaos is time reversible $d_{\text{vis}}^x \rightarrow 0$.

3.3.2.3 Robustness Against External Noise

Standard time series analysis methods evidence problems when noise is present in chaotic series. Even a small amount of noise can destroy the fractal structure of a chaotic attractor and mislead the calculation of chaos indicators such as Lyapunov exponents [21]. In order to check if our method is robust, we add a random white noise ξ_i to a signal extracted from a fully chaotic Logistic map with $\mu = 4.0$,

$$x_{i+1} = x_i + 4x_i(1 - x_i) + \sigma \xi_i, \quad (3.36)$$

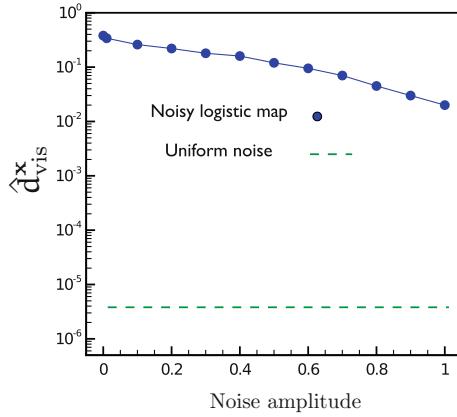


Fig. 3.9 Robustness of the KLD visibility estimator against noise. Semi-log plot of \hat{d}_{vis}^x of the graph associated to series of 10^6 data extracted from a fully chaotic Logistic map $x_{t+1} = 4x_t(1 - x_t)$ polluted with *white* uniform noise $U[-0.5, 0.5]$ according to Eq. (3.36), as a function of the noise amplitude. The corresponding KLD value of a uniform series is plotted for comparison, which is five orders of magnitude smaller even when the chaotic signal is polluted with an amount of noise of the same amplitude

where ξ_i is a random number extracted from the uniform distribution $U[-0.5, 0.5]$ and σ is the noise intensity. We vary the amplitude of the noise σ from 0 to 1 and study how the KLD visibility estimator is robust against changes in the value of σ . In Fig. 3.9 we plot \hat{d}_{vis}^x as a function of the amplitude of the noise. The value of \hat{d}_{vis} for the noise is also plotted for comparison. The KLD of the signal polluted with noise is significantly greater than zero, since it exceeds the one associated to the noise in four orders of magnitude, even when the amplitude of the noise reaches the 100 % of the signal amplitude ($\sigma = 1$). Therefore our method correctly predicts that the signal is irreversible even when adding noise.

3.3.2.4 Table of Results

We now summarize the results obtained with the visibility estimator of the KLD in a table. In the first column of Table 3.1 we show the value of \hat{d}_{vis}^x for the examples of random and chaotic processes discussed in this chapter. We also include the value of the KLD obtained using the *in* and *out* degree-degree distributions, defined as

$$\hat{d}_{vis,dd}^x = D[p_{out}^x(k, k') || p_{in}^x(k, k')] = \sum_{k,k'} p_{out}^x(k, k') \ln \frac{p_{out}^x(k, k')}{p_{in}^x(k, k')} . \quad (3.37)$$

In every case we find $\hat{d}_{vis,dd}^x \geq \hat{d}_{vis}^x$ which is consequence of the chain rule: $D[P_{out}(k, k') || P_{in}(k, k')] \geq D[P_{out}(k) || P_{in}(k)]$. Note that that the method correctly distinguishes between reversible and irreversible processes, as KLD vanishes for

the former and it is positive for the latter. A further analysis is done in Chap. 4 on time series produced by a physical system show that the KLD estimators bases on visibility algorithm can, not only distinguish between reversible and irreversible physical dynamics, but also bound from below the entropy production in the NESS.

References

1. D. Andrieux, P. Gaspard, S. Ciliberto, N. Garnier, S. Joubaud, A. Petrosyan, Phys. Rev. Lett. **98**, 150601 (2007)
2. D. Andrieux, P. Gaspard, S. Ciliberto, N. Garnier, S. Joubaud, A. Petrosyan, J. Stat. Mech: Theory Exp. **2008**, P01002 (2008)
3. N.W. Ashcroft, N.D. Mermin, There is no corresponding record for this reference (1979)
4. R. Bayer, E. McCreight, Acta Inf. **1**, 173–189 (1972)
5. D. Benedetto, E. Caglioti, V. Loreto, Phys. Rev. Lett. **88**, 048702 (2002)
6. G.E. Box, G.M. Jenkins, G.C. Reinsel, *Time Series Analysis: Forecasting and Control* (Wiley, New York, 2011)
7. H. Cai, S.R. Kulkarni, S. Verdú, IEEE Trans. Inf. Theory **52**, 3456–3475 (2006)
8. M. Costa, A.L. Goldberger, C.-K. Peng, Phys. Rev. Lett. **95**, 198102 (2005)
9. M.D. Costa, C.-K. Peng, A.L. Goldberger, Cardiovasc. Eng. **8**, 88–93 (2008)
10. D.P. Coutinho, M.A. Figueiredo, *Pattern Recognition and Image Analysis* (Springer, Berlin, 2005), pp. 355–362
11. D.P. Coutinho, A.L. Fred, M.A. Figueiredo, in *20th International Conference on Pattern Recognition (ICPR), 2010*, IEEE (2010), pp. 3858–3861
12. T.M. Cover, J.A. Thomas, *Elements of Information Theory* (Wiley, New York, 2006)
13. S.N. Dorogovtsev, *Lectures on Complex Networks* (Oxford University Press, London, 2010)
14. J. Elsner, T. Jagger, E. Fogarty, Geophys. Res. Lett. **36**, 16702 (2009)
15. P. Erdős, A. Rényi, Publicationes Math. Debrecen **6**, 290–297 (1959)
16. P. Erdős, A. Renyi, Acta Math. Hung. **12**, 261–267 (1961)
17. P. Grassberger, arXiv preprint [physics/0307138](https://arxiv.org/abs/physics/0307138) (2003)
18. G. Gutin, T. Mansour, S. Severini, Phys. A: Stat. Mech. Appl. **390**, 2421–2428 (2011)
19. S. Jonhson, V. Domínguez-García, M.A. Munoz, PLoS ONE **8**, e74025 (2013)
20. J. Kingman, J. Appl. Probab. **6**, 1–18 (1969)
21. E.J. Kostelich, T. Schreiber, Phys. Rev. E **48**, 1752 (1993)
22. A. Kraskov, H. Stögbauer, P. Grassberger, Phys. Rev. E **69**, 066138 (2004)
23. L. Lacasa, B. Luque, F. Ballesteros, J. Luque, J.C. Nuño, Proc. Nat. Acad. Sci. **105**, 4972–4975 (2008)
24. L. Lacasa, B. Luque, J. Luque, J.C. Nuno, Europhys. Lett. (EPL) **86**, 30001 (2009)
25. L. Lacasa, R. Toral, Phys. Rev. E **82**, 036120 (2010)
26. L. Lacasa, A. Nuñez, É. Roldán, J.M.R. Parrondo, B. Luque, Eur. Phys. J. B Condens. Matter Complex Syst. **85**, 1–11 (2012)
27. C. Liu, W.-X. Zhou, W.-K. Yuan, Phys. A: Stat. Mech. Appl. **389**, 2675–2681 (2010)
28. B. Luque, L. Lacasa, F. Ballesteros, J. Luque, Phys. Rev. E **80**, 046103 (2009)
29. B. Luque, L. Lacasa, F.J. Ballesteros, A. Robledo, PLoS ONE **6**, e22411 (2011)
30. M.E. Newman, SIAM Rev. **45**, 167–256 (2003)
31. F. Pérez-Cruz, in *IEEE International Symposium on Information Theory, 2008. ISIT 2008*, IEEE (2008), pp. 1666–1670
32. A. Porporato, J. Rigby, E. Daly, Phys. Rev. Lett. **98**, 094101 (2007)
33. Z. Rached, F. Alajaji, L.L. Campbell, IEEE Trans. Inf. Theory **50**, 917–921 (2004)
34. T. Schürmann, P. Grassberger, Chaos: Interdisc. J. Nonlinear Sci. **6**, 414–427 (1996)
35. D. Sherrington, S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975)
36. J. Sprott, G. Rowlands, Int. J. Bifurcat. Chaos **11**, 1865–1880 (2001)

37. S. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering* (Perseus Books Group, New York, 2001)
38. Q. Wang, S.R. Kulkarni, S. Verdú, IEEE Trans. Inf. Theory **51**, 3064–3074 (2005)
39. Q. Wang, S.R. Kulkarni, S. Verdú, IEEE Trans. Inf. Theory **55**, 2392–2405 (2009)
40. D.J. Watts, S.H. Strogatz, Nature **393**, 440–442 (1998)
41. G. Weiss, J. Appl. Probab. **12**, 831–836 (1975)
42. Y. Yang, J. Wang, H. Yang, J. Mang, Phys. A; Stat. Mech. Appl. **388**, 4431–4437 (2009)
43. J. Ziv, A. Lempel, IEEE Trans. Inf. Theory **24**, 530–536 (1978)
44. J. Ziv, N. Merhav, IEEE Trans. Inf. Theory **39**, 1270–1279 (1993)

Chapter 4

A Case Study: The Flashing Ratchet

To explore the bound to the entropy production, we study the case of a discrete flashing ratchet model, where we can compare the entropy production with the analytical value and the empirical estimations of the KLD. With this model, we can analyze how information losses affect the estimation of the KLD and the tightness of the bound for the entropy production.

Flashing ratchets are particular examples of Brownian motors, which are extensively discussed in Ref. [15]. Brownian motors are microscopic machines that take advantage of the thermal fluctuations of the environment and are able to transport Brownian particles in a certain direction, and therefore to extract useful work from the fluctuating environment. The dynamics of the transported Brownian particle is irreversible and energy is dissipated to the thermal environment along the motion of the particle. Consequently, Brownian motors are excellent case studies for the relationship between irreversibility and dissipation. In the last years, Brownian motors have been used to model the dynamics of molecular biological motors [6] and artificial Brownian motors have been designed to transport colloidal particles, magnetic vortices or even charge or spin in quantum systems [5].

We study one of the simplest models of Brownian motors called *flashing ratchet*, which consists of a Brownian particle moving in a linear asymmetric potential that is switched on and off randomly [1, 13, 14]. We use a simplified discrete version of the flashing ratchet where the dynamics of the particle is described by a Master equation between a discrete number of states. Our simple model allows one to calculate analytically the entropy production of the system and compare it to the value of the KLD rate between forward and time-reversed trajectories.

This chapter is organized as follows: In Sect. 4.1 we introduce the notion of flashing ratchet in the context of Brownian motors. In Sect. 4.2 we describe the discrete flashing ratchet model that we use in Sect. 4.3 to show that we can estimate the average entropy production of the model with the KLD of trajectories containing full information of the system. In Sect. 4.4 we show that sampling partial information of the system we can only bound from below the average entropy production of the system in the NESS. In Sect. 4.5 we show that we can still detect irreversibility even

when the system does not exhibit any measurable flow studying the ratchet at stall. In Sect. 4.6 we discuss the applicability of the visibility estimator of the KLD to estimate the dissipation of our model. Section 4.7 shows the concluding remarks of the Chapter.

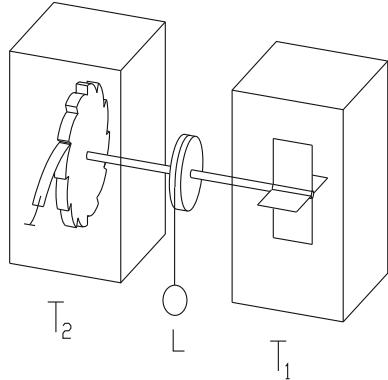
4.1 Ratchets

As we have already discussed, the importance of thermal fluctuations in microscopic systems gives rise to phenomena that are not observed in the macroscopic world. An interesting phenomenon is the possibility to extract work from thermal fluctuations. *Brownian motors* are devices that can rectify thermal fluctuations to transport Brownian particle in a specific spatial direction. The *ratchet effect* refers to the effect of rectifying the Brownian fluctuations to produce useful work, and it was first introduced by Smoluchowski [16]. The first (gedanken) experiment to observe this phenomenon was proposed by Feynman [4]. Feynman's device contains an axle with vanes in one of his ends and a *ratchet* (circular saw with asymmetric saw-teeth) with pawl in the other, as we illustrate in Fig. 4.1. If the vanes and the ratchet are surrounded by gases at different temperatures, T_1 and T_2 respectively, the whole system moves in average in a given direction. This net movement can be used to perform work, for example a small weight can be lifted when the temperatures are different. The systematic extraction of work from thermal fluctuations in Feynman's ratchet is only possible because the system is out of equilibrium [12].

Since the first work by Smoluchowski, a variety of *Brownian motors* have been introduced in literature [15]. These are engines that rectify thermal fluctuations producing directed transport Brownian particles. The essential ingredients of a Brownian motor are: (i) A Brownian particle immersed in a thermal bath, (ii) a spatially-periodic potential and (iii) a mechanism that drives the particle out of equilibrium. *Ratchets* are a paradigmatic example of Brownian motors where an asymmetric potential together with a perturbation that changes the potential randomly or periodically drive the particle out of equilibrium [1]. Several different systems where directed transport is observed have been modeled as thermal ratchets. Two examples are a rotating chemical molecule [7] or molecular biological motors [6]. Furthermore, artificial Brownian motors have been designed inspired in ratchets, e.g. using colloidal particles [10], superconducting vortices [3, 18] or quantum dots [9].

We focus on a specific example, namely, the *flashing ratchet*, originally described in [1]. In a flashing ratchet, a Brownian particle immersed in a thermal bath at temperature T moves in a linear asymmetric potential that is switched on and off randomly. The random switch of the asymmetric potential prevents the particle to reach thermal equilibrium and induces directed transport in a direction that depends on the shape of the potential. The simplest description for a one dimensional ratchet, is given by the Langevin equation for the position of the Brownian particle, x . The equation of motion reads

Fig. 4.1 Feynman's ratchet and pawl motor. Picture taken from [12]



$$\gamma \dot{x}(t) = -\alpha(t) \frac{\partial V(x(t))}{\partial x} + \xi(t), \quad (4.1)$$

where $\xi(t)$ is a Gaussian white noise with zero mean and correlation $\langle \xi(t)\xi(t') \rangle = 2kT\gamma\delta(t-t')$ and $\alpha(t)$ is a dichotomous noise that takes two possible values 0 and 1 with the same probability. When $\alpha = 0$, the particle is free and when $\alpha = 1$ the particle feels the periodic asymmetric potential described by

$$V(x) = \begin{cases} \frac{V_{\max}}{a} \frac{x}{L} & \text{if } 0 \leq \frac{x}{L} \leq a, \\ -\frac{V_{\max}}{1-a} \left(\frac{x}{L} - a \right) + V_{\max} & \text{if } a \leq \frac{x}{L} \leq 1, \end{cases} \quad (4.2)$$

where V_{\max} is the maximum height of the potential, L is the period of the potential, and a is a parameter that controls the asymmetry of the potential (only if $\alpha = 1/2$ the potential is symmetric). Equation (4.2), together with the periodic boundary condition $V(x \pm L) = V(x)$, describe a seesaw periodic asymmetric potential where a net particle current can be observed in the negative direction if $a > 1/2$ and in the positive direction if $a < 1/2$. In Fig. 4.2 we show the potential of the flashing ratchet and the probability distribution of the position of the particle when the potential is switched on and off. As it can be seen in the figure, if the potential is initially switched on, the particle first equilibrates with the thermal bath and the distribution of the position is peaked around the minimum of the potential. When the potential is switched off, the particle diffuses freely and its distribution is Gaussian. Because of the asymmetry of the potential, it is more likely that the particle jumps to the right in the potential of Fig. 4.2. A net current to the right is induced in this case by switching this potential randomly or periodically [13].

A discrete version of the flashing ratchet can be considered if we coarse-grain the spatial dimension in bins of equal spacing Δx . In this case, the position of the Brownian particle can be labeled by a discrete variable indicating in which bin the particle is located. By knowing the position in this coarse grained description and

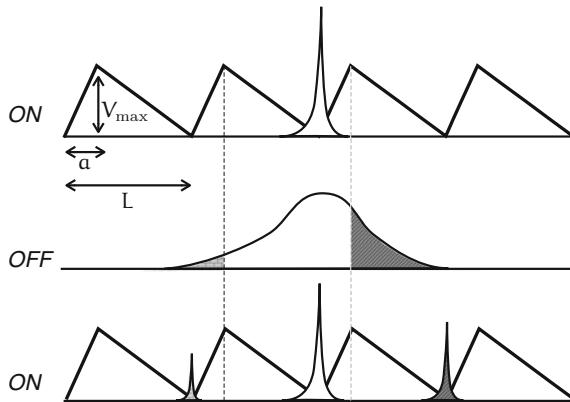


Fig. 4.2 Flashing ratchet. *Top* Seesaw asymmetric potential with parameters a , V_{\max} and L indicated in the figure and probability distribution of the position of the Brownian particle shaded in white. *Middle* Distribution of the position of the particle after the potential is switched off. Dark shaded areas represent regions where the particle jumps to the next or the previous period of the potential if the potential is switched on. *Bottom* Equilibrium distribution of the position of the particle after the potential is switched on. As a result, a net current to the right is produced

the state of the potential $\alpha = 0, 1$ one can describe the dynamics of the system using a Master equation for a discrete set of states as it is shown in Sect. 4.2.

4.2 The Model

Our model is sketched in Fig. 4.3. The model consists of a Brownian particle moving on a one dimensional lattice. The particle is immersed in a thermal bath at temperature T and moves in a periodic, linear, asymmetric potential of maximum height $2V$, which is switched on and off at a constant rate r . Our model corresponds to the generic flashing ratchet described in Sect. 4.1 with $a = 2/3$, $V_{\max} = 2V$, and spatial binning of $\Delta x = L/3$. With this discretization, the state of the Brownian particle can be specified by two random observables: the position of the particle X (0, 1 or 2) and the state of the potential Y (ON , $Y = 1$ or OFF , $Y = 0$).

The particle evolves in continuous time according to a Master equation. The dynamics is described in terms of transition rates of spatial jumps and switching between six different states: $(0, 0)$, $(1, 0)$, $(2, 0)$, $(0, 1)$, $(1, 1)$ and $(2, 1)$. For each possible transition except switches, i.e. $(x_1, y_1) \rightarrow (x_2, y_2)$ with $y_1 = y_2 = y$, we define a transition rate $k_{(x_1, y) \rightarrow (x_2, y)}$ obeying detailed balance,

$$k_{(x_1, y) \rightarrow (x_2, y)} = \exp \left[-\frac{V_y(x_2) - V_y(x_1)}{2kT} \right]. \quad (4.3)$$

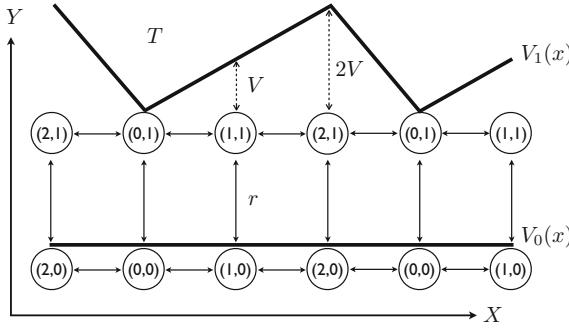


Fig. 4.3 Illustration of our discrete ratchet model. A Brownian particle is immersed in a thermal bath at temperature T and moves in one dimension in an asymmetric linear potential $V_1(x)$ of height $2V$ and $a = L/3$ with periodic boundary conditions. The potential is switched on and off at a rate r , where $V_0(x) = 0$ represents a flat potential, and the switching probability does not depend on the position of the particle. The state of the particle is represented by two random variables (X, Y) indicated in the figure, where $X = \{0, 1, 2\}$ stands for the position of the particle whereas $Y = \{0, 1\}$ for the state of the potential. Using this description, the system can be in six different states, $(0, 0)$, $(1, 0)$, $(2, 0)$, $(0, 1)$, $(1, 1)$ and $(2, 1)$

When the potential is on ($y = 1$), the value of the potential energy $V_1(x)$ is given in Fig. 4.3, which corresponds to (4.2) with $V_{\max} = 2V$ and $a = L/3$. When the potential is off, $V_0(x) = 0$ for all x , and $k_{(x_1,0) \rightarrow (x_2,0)} = 1$ for $x_1 \neq x_2$. The switching rate does not depend on the position of the particle: $k_{(x,y_1) \rightarrow (x,y_2)} = r$ for any value of x and $y_1 \neq y_2$, and consequently violates detailed balance, driving the system out of equilibrium.

We simplify the analysis by mapping the dynamics onto a discrete-time process, a Markov chain. To this end, we record in a time series $(\mathbf{x}, \mathbf{y}) = \{x_1^n, y_1^n\}$ a list of the n first visited states, discarding any information about the time when jumps and switches occur. The resulting Markov chain is defined by the transition probabilities

$$p[(x_2, y_2)|(x_1, y_1)] = \frac{k_{(x_1, y_1) \rightarrow (x_2, y_2)}}{\sum_{x_2, y_2} k_{(x_1, y_1) \rightarrow (x_2, y_2)}}. \quad (4.4)$$

Since we discard any information about the transition times, we focus only on the dissipation and KLD rates per jump or *per data*. For finite switching rate r and since $a > 1/2$, the ratchet rectifies the thermal fluctuations inducing a current to the left in Fig. 4.3 [1, 13]. The system obeys a local detailed balance condition, therefore the trajectory of the particle is a Markov chain obeying local detailed balance condition analogous to the processes discussed in Sect. 2.5.1. The nonequilibrium nature of the switching can be interpreted in two alternative ways: one can imagine that it is activated by a thermal bath at infinite temperature or by an external agent [13]. In either of the two interpretations, switching does not induce any entropy production (the bath needs an infinite amount of energy to change its entropy and the external agent does not produce any entropy change). Therefore, entropy is only produced

when heat is dissipated to the bath at temperature T , which only occurs when the potential is on. For the time series containing the information of both position and state of the system (which we call *full* information), the average entropy production (or dissipation) per data in the time series, $\langle \dot{S}_{\text{prod}} \rangle$, is then [cf. (2.44)]

$$\langle \dot{S}_{\text{prod}} \rangle = \sum_{y=0,1} \sum_{x_1,x_2=0,1,2} p[(x_1, y); (x_2, y)] \frac{V_y(x_1) - V_y(x_2)}{T}. \quad (4.5)$$

Moreover, since the trajectories containing full information are Markov chains obeying detailed balance, the entropy production rate is equal to the KLD rate, $\langle \dot{S}_{\text{prod}} \rangle/k = d^{X,Y} = d_2^{X,Y}$. We now analyze how we can estimate the KLD rate using single stationary trajectories of this model, and how close is this estimation to the entropy production depending on the number of degrees of freedom of the system that are sampled in the time series.

4.3 Full Information

Firstly, we investigate the estimation of the KLD rate when using full information of the system (the position of the particle X and the state of the potential Y), and how close is this KLD rate to the actual entropy production of the process. In Fig. 4.4 we compare the actual dissipation and several empirical estimations of $d^{X,Y}$ for different values of the height of the potential, V . For each value of V , we simulate a single stationary time series of $n = 10^6$ data that contains full information, and calculate the plug-in estimators $\hat{d}_2^{X,Y}$, $\hat{d}_3^{X,Y}$, as well as the compression-based estimator of the KLD rate, $\hat{d}_c^{X,Y}$. We notice that time series containing full information are constructed as follows: $(x, y) = x$ if $y = 0$ and $(x, y) = x + 3$ if $y = 1$.

Since trajectories containing full information are Markovian, the plug-in estimator immediately converges to the dissipation $\hat{d}_2^{X,Y} = \hat{d}_3^{X,Y} = d^{X,Y} = \langle \dot{S}_{\text{prod}} \rangle/k$ if there is enough statistics, which happens when V is below or of order kT . If $V \gg kT$, it is very unlikely to observe uphill jumps such as $(0, 1) \rightarrow (1, 1)$, $(0, 1) \rightarrow (2, 1)$, or $(1, 1) \rightarrow (2, 1)$ in a single stationary trajectory. A time series of n data captures the statistics of jumps with probability well above $1/n$, which amounts to say energy jumps below $kT \ln n$, ($kT \ln 10^6 \approx 14kT$ for the trajectory used in the figures).

If, for instance, the transition $(0, 1) \rightarrow (1, 1)$ is missing in the trajectory, there is no way of estimating $p[(0, 1); (1, 1)]$ which contributes to two terms in $\hat{d}_2^{X,Y}$ [see Eq. (3.3) for $m = 2$]. One of these two terms accounts for jumps $(0, 1) \rightarrow (1, 1)$, which are very unlikely and their contribution to the total dissipation rate is negligible, and the other term accounts for jumps $(1, 1) \rightarrow (0, 1)$, whose probability is larger and therefore contribute more significantly to the entropy production.

In Fig. 4.4, $\hat{d}_2^{X,Y}$ (blue circles) and $\hat{d}_3^{X,Y}$ (green diamonds) have been calculated restricting the average to sequences (of two or three data respectively) whose reverse are also observed in the time series, as given by Eq. (3.9). The sudden drops in $\hat{d}_2^{X,Y}$

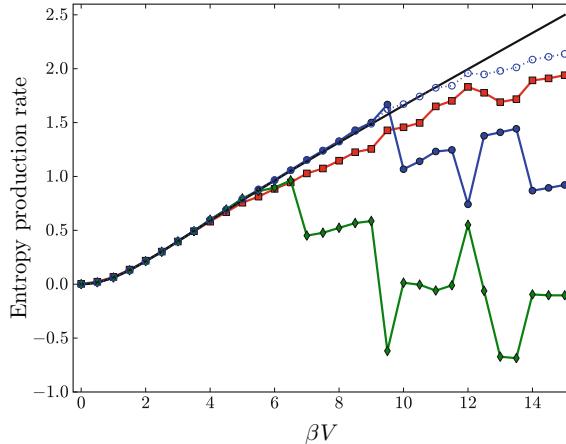


Fig. 4.4 Dissipation and irreversibility in the flashing ratchet using full information. Analytical value of the average dissipation per data in units of kT (black line) as a function of βV in the flashing ratchet ($r = 1$) and different estimators of $d^{X,Y}$. For each value of βV , estimators are obtained from a single stationary time series of $n = 10^6$ data containing full information of the system (position, X , and state of the potential, Y): Plug-in estimators: $\hat{d}_2^{x,y}$ (blue circles), $\hat{d}_3^{x,y}$ (green diamonds), and $\tilde{d}_2^{x,y}$ using biased probabilities with $\gamma = 1$ (blue open circles). Compression estimator: $\hat{d}_c^{x,y}$ (red squares)

and $\hat{d}_3^{x,y}$ are a consequence of lack of statistics in the trajectory. For the specific time series used in Fig. 4.4, the lack of statistics begins at $\beta V \simeq 10$ for $\hat{d}_2^{x,y}$ and arises earlier for $\hat{d}_3^{x,y}$ because the three-data sampling space is bigger and it is easier that some transitions $(x_1, y_1) \rightarrow (x_2, y_2) \rightarrow (x_3, y_3)$ do not appear while their reverse do.

A more efficient way of dealing with the missing sequences is incorporating a small bias to the empirical probabilities, as described in Eq. (3.10). This is equivalent to assigning a probability of order $1/n$ to those transitions that are not observed in a time series of n data. Figure 4.4 shows the value of $\tilde{d}_2^{x,y}$ using a biased two-data probabilities (3.10) with $\gamma = 1$ (blue open circles), which is able to extend the accuracy of the estimation even when there is lack of statistics.

Although in the case of Markovian series with a finite number of states the most convenient strategy is to use the plug-in estimator, we include for comparison the value of the compression estimator $\hat{d}_c^{x,y}$ defined in (3.15) (red squares), which gives accurate values of the dissipation for weak potentials. Furthermore, the compression estimator is better than some plug-in estimators even for strong potentials, since it does not exhibit sudden jumps due to lack of statistics.

4.4 Partial Information

We now analyze the performance of our estimators when there is not access to the full description of the system. We now assume that *only* the position of the ratchet, X , is observable. Accordingly, we simulate trajectories containing full information, and we remove the information of the state afterwards, $(\mathbf{x}, \mathbf{y}) \rightarrow \mathbf{x}$. The resulting time series $\mathbf{x} = \{x_i^n\}$ are not Markovian and hence the limit (3.5) is not reached for small values of m . In this case, we proceed by obtaining m -data KLD rates, $\hat{d}_m^{\mathbf{x}}$, for m as large as possible and fit the resulting values to the ansatz (3.8).

We have generated trajectories of $n = 10^7$ data for values of V that range from 0 to $2kT$. Once we remove the information of the state of the potential from these time series, we are able to estimate $\hat{d}_m^{\mathbf{x}}$ up to $m = 9$ with no lack of statistics. Figure 4.5 shows the value of the plug-in estimators $\hat{d}_m^{\mathbf{x}}$ for $m = 2, 3, 5, 7, 9$ and the extrapolation $\hat{d}_{\infty}^{\mathbf{x}}$ (orange pentagons connected by a dashed line to guide the eye) resulting from the fit to the ansatz (3.8). For each value of βV , we fit $\hat{d}_m^{\mathbf{x}}$ as a function of m for $m = 2, 3, \dots, 9$ to Eq. (3.8) using the curve fitting tool available in MATLAB, which provides a robust least-squares fit with bisquare weights as described in [11]. The fit itself for a particular value of the potential, $\beta V = 1$, is shown in the inset of Fig. 4.5. Our ansatz reproduces the dependence of $\hat{d}_m^{\mathbf{x}}$ with m but the final estimator $\hat{d}_{\infty}^{\mathbf{x}}$ still bounds significantly from below the actual dissipation (black solid line in Fig. 4.5).

Despite plug-in estimators bound from below the entropy production, they are able to distinguish between equilibrium and NESS, even with partial information. In equilibrium ($V = 0$), the trajectories are reversible and all the estimators vanish, $\hat{d}_m^{\mathbf{x}} = 0$ for $m = 2, \dots, 9$, whereas for the NESS ($V > 0$) they detect the irreversibility of the process yielding $\hat{d}_m^{\mathbf{x}} > 0$ for all m . This is illustrated in Fig. 4.6, where we plot the dependence of the plug-in estimators with the size of the trajectory. For $\beta V = 0$, $\hat{d}_2^{\mathbf{x}}, \hat{d}_3^{\mathbf{x}}$ and $\hat{d}_5^{\mathbf{x}}$ tend to zero when increasing the number of data whereas they saturate to a positive value in the NESS ($\beta V = 1$).

There are two possible origins for the discrepancy between $\hat{d}_{\infty}^{\mathbf{x}}$ and the dissipation: either (i) our fit underestimates the actual KLD rate d^X of the trajectory; or (ii) the bound (2.34) is not tight when using only the information of the position. To address this question we need to calculate the actual value of d^X . Since the position of the ratchet \mathbf{x} is a hidden Markov chain, we can calculate its KLD rate d^X semi-analytically, using the Lyapunov exponents (2.53, 2.54) introduced in Sect. 2.5.2.

In Fig. 4.7 we show the value of the semi-analytical calculation of d^X using the norm of transition matrices, Eq. (2.55), which is not significantly different to the empirical estimation $\hat{d}_{\infty}^{\mathbf{x}}$. We therefore conclude that $\hat{d}_{\infty}^{\mathbf{x}}$ is a good estimation of d^X , but still d^X only yields a lower bound to dissipation whose accuracy is in principle hard to determine. This is an expected result, since the position of a particle in a flashing ratchet does not obey the Gallavotti-Cohen theorem [8].

Summarizing, although $\hat{d}_{\infty}^{\mathbf{x}}$ turns out to be a good estimator of d^X , using only information of the position we only get a lower bound to the dissipation. We also

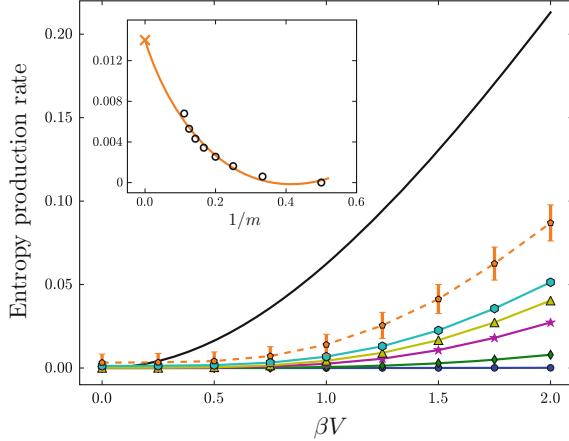
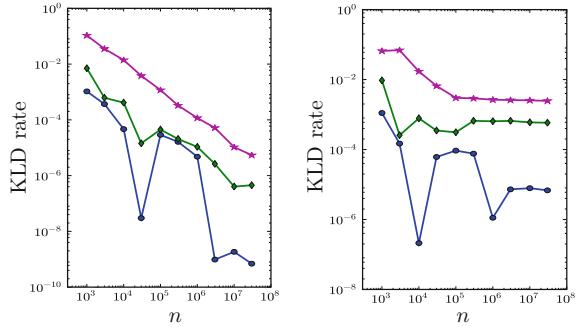


Fig. 4.5 Dissipation and irreversibility in the flashing ratchet using partial information. Average dissipation per data in units of kT (black line) and plug-in estimators of d^X using partial information given by the position (X) for a discrete flashing ratchet with $r = 1$. For each value of βV , we calculate estimators from a single stationary time series of $n = 10^7$ data containing partial information: \hat{d}_2^x (blue circles), \hat{d}_3^x (green diamonds), \hat{d}_5^x (purple stars), \hat{d}_7^x (yellow triangles), \hat{d}_9^x (cyan hexagons) and the result from the fit \hat{d}_∞^x (orange pentagons with error bars and connected by a dashed line). Inset \hat{d}_m^x as a function of $1/m$ for $m = 1, \dots, 9$ for $\beta V = 1$ (open black circles) and the fit to the ansatz (orange line). The y-intercept of the fit is indicated by an orange cross and it is equal to \hat{d}_∞^x .

Fig. 4.6 Scaling of plug-in estimators of d^X , \hat{d}_m^x , with the size of the time series n , for a flashing ratchet ($r = 1$), for $\beta V = 0$ (left) and $\beta V = 1$ (right) \hat{d}_2^x (blue circles), \hat{d}_3^x (green diamonds) and \hat{d}_5^x (purple stars). We simulate a single stationary trajectory \mathbf{x} of 10^7 data and calculate the estimators for subsequences containing the first n data of \mathbf{x}



show in Fig. 4.7 the value of \hat{d}_c^x , which is well below the plug-in estimator \hat{d}_∞^x . The compression estimator \hat{d}_c^x lies between \hat{d}_7^x and \hat{d}_9^x (not shown in the plot), indicating that it is only able to capture up to 8-data correlations. For completeness, we include the calculation of d^X based on the replica trick (see Appendix B.3). The estimation using replica trick yields a tight bound for $V < kT$, but departs from d^X for larger values of V . This deviation is caused by the estimation of the limits in Eqs. (B.34, B.37), where we take $\alpha \rightarrow 0$ when α is defined only for integer values, one of the standard drawbacks of the replica trick [2].

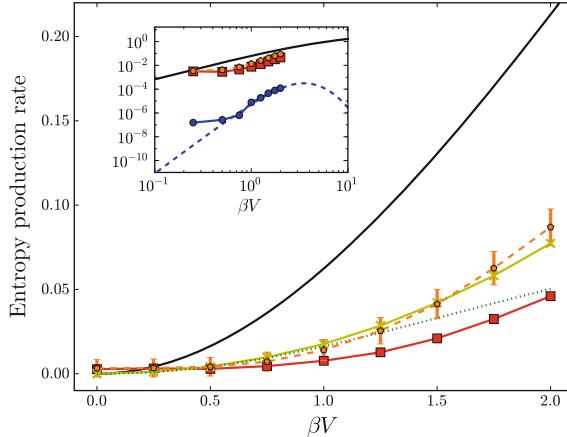


Fig. 4.7 Average dissipation per data in units of kT (black line) and different estimators of d^X for a flashing ratchet described with partial information ($r = 1, n = 10^7$ data) as a function of βV : \hat{d}_∞^X (orange dashed pentagons) \hat{d}_c^X (red squares), replica estimation of d^X (green dotted line) and semi-analytical value of d^X (yellow crosses). Inset Dependence of the average dissipation (black line), \hat{d}_2^X (analytical values in blue dashed line), \hat{d}_c^X and \hat{d}_∞^X on βV in the vicinity of $\beta V = 0$

Although our estimators give low values of the dissipation when using partial information, they still capture the asymptotic behavior close to equilibrium, i.e. for βV small. Entropy production decreases as V^2 when $V \rightarrow 0$, so do plug-in estimators $\hat{d}_3^X, \dots, \hat{d}_9^X, \hat{d}_\infty^X$, and the compression estimator \hat{d}_c^X . Some of them are plotted in the inset of Fig. 4.7. On the other hand, $\hat{d}_2^X \propto j^2 \propto V^6$, since the current is $j \propto V^3$ in this case [see Eq. (2.45)]. Recall that, close to equilibrium, calculating \hat{d}_2^X is equivalent to estimating the entropy production using currents and standard linear irreversible thermodynamics, $d_2^X \propto j^2$, as shown in Eq. (2.45). It is then remarkable that the estimators involving the statistics of three or more data are able to reproduce qualitatively the behavior of the dissipation in cases where linear thermodynamics fails.

4.5 Detecting Irreversibility in the Absence of Currents

The improvement observed when using the plug-in estimators of higher order than \hat{d}_2^X is more significant in a NESS which does not exhibit observable currents in X . In this case, $j = 0$ and $\hat{d}_2^X = 0$, but using higher order statistics we can still detect the time irreversibility of the trajectory. This happens for example if we add to the flashing ratchet an external force F opposite to the current, i.e., pointing in the positive x -direction. The force modifies the energy landscape and consequently the spatial transition rates $k_{(x_1,y) \rightarrow (x_2,y)}$ by a factor $e^{\beta F L_{(x_1,y);(x_2,y)} / 2}$, where $L_{(x_1,y);(x_2,y)}$ is the

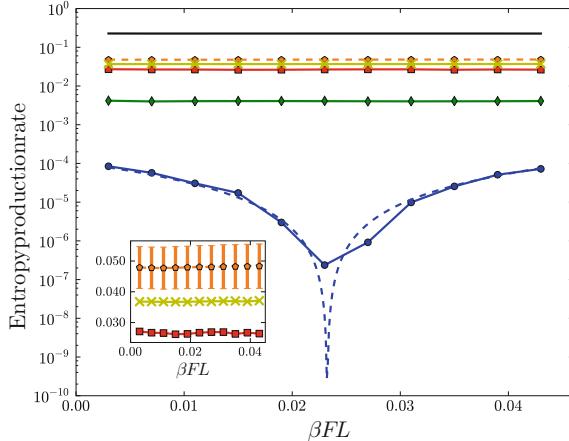
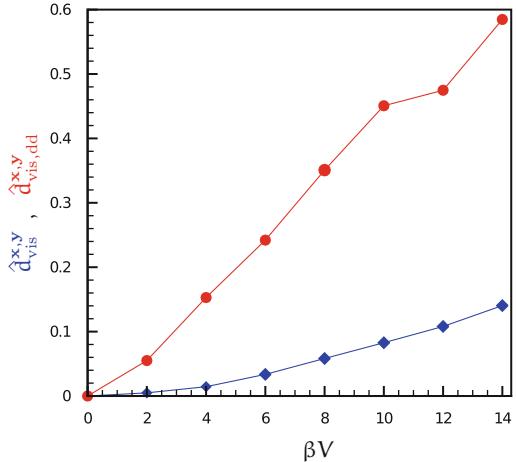


Fig. 4.8 Dissipation and irreversibility in a stalled ratchet. Average dissipation and KLD rate obtained from partial information in a forced flashing ratchet. Average dissipation per data (in units of kT) in the flashing ratchet (with $r = 2$, and $\beta V = 2$) and different estimations of d^X obtained from a single time series of $n = 10^7$ data containing partial information (position) as a function of the external force F : analytical value of the average dissipation (black line), \hat{d}_2^X (blue circles, analytical values in blue dashed line), \hat{d}_3^X (green diamonds), \hat{d}_c^X (red squares), semi-analytical calculation of d^X (yellow crosses) and \hat{d}_∞^X (orange hexagons). The minimum in \hat{d}_2^X corresponds to the stalling force. *Inset* \hat{d}_c^X , semi-analytical value of d^X and \hat{d}_∞^X as a function of the external force

spatial distance that separates the two points (x_1, y) and (x_2, y) . Here $L_{(x_1,y);(x_2,y)}$ is defined positive if the jump $(x_1, y) \rightarrow (x_2, y)$ points in the same direction as the force (i.e. to the right), and negative otherwise. At the *stalling force* F_{stall} , the current is canceled by the force and the system does not move on average when it is described only by X , but still dissipates energy. This situation is analogous to an ATP-consuming molecular motor that stalls under the action of an external force opposite to its direction of motion [17]. At the stall force, the system looks like it is in equilibrium If we only have access to the information of the position: the spatial current vanishes, and so does \hat{d}_2^X , as shown in Fig. 4.8.

However, there is a finite dissipation rate (black line in the figure) and the corresponding irreversibility is captured by the statistics of substrings of 3 data or more. Although d^X is below the real dissipation by an order of magnitude (see the semi-analytical value of d^X , yellow crosses in Fig. 4.8), it does not exhibit any sensible change at stall force. Finally, both \hat{d}_∞^X and \hat{d}_c^X provide estimates of d^X which are correct within one order of magnitude (see the inset of Fig. 4.8).

Fig. 4.9 Visibility estimators of the KLD using full information in the flashing ratchet. Values of $\hat{d}_{\text{vis}}^{\text{x,y}}$ and $\hat{d}_{\text{vis,dd}}^{\text{x,y}}$ for a discrete flashing ratchet ($r = 1$) as a function of V/kT . For each value of V , we generate a stationary time series of $N = 2^{19}$ steps described with full information (position and state of the potential)



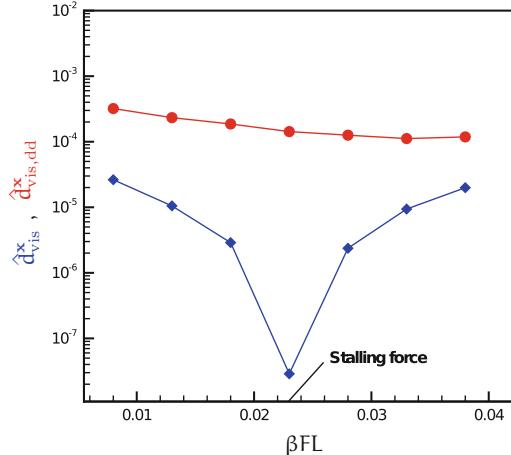
4.6 Estimating Dissipation with the Visibility Technique

We now study how our estimators of the KLD rate based on the horizontal visibility algorithm can be used to estimate the average dissipation rate in the NESS using single stationary trajectories of our discrete ratchet model. Although the visibility technique is originally designed to distinguish between equilibrium and NESS using continuous data, as shown in Sect. 3.3, we can also apply this method to discrete systems. We now go one step further and analyze if our visibility-based estimator of the KLD can also quantify the amount of irreversibility—or estimate the dissipation rate—in the discrete ratchet model. The performance of the KLD estimators using degree and degree-degree distributions are compared.

We first work with trajectories with full information generated with our discrete ratchet model. In Fig. 4.9 we show the values of $\hat{d}_{\text{vis}}^{\text{x,y}}$ and $\hat{d}_{\text{vis,dd}}^{\text{x,y}}$ as a function of V/kT , for stationary time series of 2^{19} data. Note that for $V = 0$, both KLD using degree distributions and degree-degree distributions vanish. On the other hand, both $\hat{d}_{\text{vis}}^{\text{x,y}}$ and $\hat{d}_{\text{vis,dd}}^{\text{x,y}}$ increase with V , providing a lower bound to the dissipation rate, which also increases with V [cf. Fig. 4.4]. Visibility estimators of the KLD increase with V even for large values of the potential where the statistics are poor and the plug-in estimators of the KLD fail when estimating irreversibility as shown in Fig. 4.4. However, the value of the visibility estimators of the KLD using full information ($\hat{d}_{\text{vis}}^{\text{x,y}}$ and $\hat{d}_{\text{vis,dd}}^{\text{x,y}}$) are one order of magnitude below the dissipation and the plug-in estimators of the KLD. The degree distributions capture the irreversibility of the original series but it is difficult to establish a quantitative relationship between $\hat{d}_{\text{vis}}^{\text{x,y}}$ and the KLD between trajectories.

We now illustrate that the KLD based on the degree-degree distribution takes into account more information of the visibility graph structure than the KLD using degree distributions, providing a closer bound to the physical dissipation as it is expected

Fig. 4.10 Irreversibility measures $\hat{d}_{\text{vis}}^{\text{x}}$ and $\hat{d}_{\text{vis,dd}}^{\text{x}}$ in the flashing ratchet ($r = 2, V = 2kT$) as a function of FL/kT . For each value of the force, we make use of a single stationary series of size $N = 10^6$ containing partial information (the state information is removed)



by the chain rule, $\hat{d}_{\text{vis,dd}}^{\text{x}} \geq \hat{d}_{\text{vis}}^{\text{x}}$. The improvement is significant in some situations like the flashing ratchet with a force that opposes to the net current on the system. In Fig. 4.10 we address this situation, using series described with only partial information of the system (position of the particle). We show how $\hat{d}_{\text{vis}}^{\text{x}}$ tends to zero when the force approaches to the stalling value. Therefore, the value of $\hat{d}_{\text{vis}}^{\text{x}}$ is misleading in this case since it predicts reversibility, as do the KLD estimators based on local flows or currents like \hat{d}_2^{x} (see Fig. 4.8). However, $\hat{d}_{\text{vis,dd}}^{\text{x}}$ captures the irreversibility of the time series, and yields a positive value at the stalling force (cf. Fig. 4.8).

4.7 Conclusions

We have shown that it is possible to estimate the entropy production rate in a discrete system by analyzing statistical properties of a time series observed in a NESS. The KLD rate per data between the time series and its time reversed is a lower bound to the entropy production rate of a discrete flashing ratchet model in the NESS.

We have tested three estimators of KLD rate: plug-in estimators, the compression estimator and the estimator based on visibility algorithm. We have tested the performance of our estimators in the discrete flashing ratchet. We have shown that the KLD is a powerful tool to identify nonequilibrium states and to estimate the entropy production of a process, if this entropy production is of order of the Boltzmann constant. We have also shown that the bound given by the KLD can detect a non-zero dissipation even when the data does not exhibit any measurable flows.

Let us summarize our results by presenting a “recipe” to estimate the KLD from an experimental time series recorded from a discrete system in a NESS. If the number of possible states of the system is small enough, the best approach is to calculate

the plug-in estimators \hat{d}_m^x (3.4) and then check the convergence when m increases. The possible lack of statistics can be circumvented using a small artificial bias, as discussed in Sect. 3.1.1. If \hat{d}_m^x saturates for some value m^* , then the time series is an m^* -th order Markov process and $\hat{d}^x = \hat{d}_{m^*}^x$. Otherwise, we can use the ansatz (3.8) and obtain \hat{d}_∞^x which is a good estimate of the KLD rate.

A second and complementary approach is the use of the compression estimator introduced in Sect. 3.1.2. The estimator yields correct results in the examples that we have analyzed, but there is no clue about the corresponding error. Nevertheless, the compression estimator could be the only possible approach if the number of states of the time series is large. In this case, the calculation of empirical probability distributions $\hat{p}(x_1^m)$ would be unfeasible even for short substrings.

When using full information of the discrete ratchet, the estimators based on the visibility technique are sensitive to small changes in the irreversibility of the trajectories and reproduce qualitatively the dependence of the entropy production with parameters of the system. However, the bound to the dissipation is weak even using full information of the system. The second symbolization (after the first symbolization given by the discretization) produced by the visibility algorithm produces, by virtue of the chain rule, that the bound to the dissipation is weaker than using the m -data statistics of the trajectories. We have also found that the degree distribution captures only the information of the current whereas the degree-degree distributions can be used to predict dissipation when the current of the system vanishes.

The extension of our technique to systems with many states (or described by real-valued observables) is relevant in many practical situations, especially to analyze data coming from biological systems. This can be done considering time asymmetric functionals of the data, which reduce the number of observables, and hence the number of states, but keep information about the irreversibility of the series. In the next chapter, we address this issue by considering biological time series obtained from bullfrog's ear hair bundles.

References

1. A. Ajdari, J. Prost, Comptes rendus de l'Académie des sciences. Série 2, Mécanique, Physique, Chimie, Sciences de l'univers, Sciences de la Terre **315**, 1635–1639 (1992)
2. A. Crisanti, G. Paladin, A. Vulpiani, *Products of Random Matrices in Statistical Physics* (Springer Series in Solid State Sciences) (Springer, 1993)
3. L. Dinis, E. González, J.V. Anguita, J.M.R. Parrondo, J. Vicent, New J. Phys. **9**, 366 (2007)
4. R.P. Feynman, R.B. Leighton, M. Sands, E. Hafner, Am. J. Phys. **33**, 750 (1965)
5. P. Hänggi, F. Marchesoni, Rev. Mod. Phys. **81**, 387 (2009)
6. F. Jülicher, A. Ajdari, J. Prost, Rev. Mod. Phys. **69**, 1269 (1997)
7. T.R. Kelly, H. De Silva, R.A. Silva, Nature **401**, 150–152 (1999)
8. D. Lacoste, K. Mallick, Phys. Rev. E **80**, 021923 (2009)
9. H. Linke, W. Sheng, A. Löfgren, H. Xu, P. Omling, P. Lindelof, EPL (Europhysics Letters) **44**, 341 (1998)
10. H. Linke, B. Alemán, L. Melling, M. Taormina, M. Francis, C. Dow-Hygelund, V. Narayanan, R. Taylor, A. Stout, Phys. Rev. Lett. **96**, 154502 (2006)

11. MATLAB R2011b documentation (curve fitting toolbox). <http://www.mathworks.es/help/toolbox/curvefit/>
12. J.M.R. Parrondo, P. Español, Am. J. Phys. **64**, 1125–1129 (1996)
13. J.M.R. Parrondo, B. Cisneros, Appl. Phys. A: Mater. Sci. Process. **75**, 179–191 (2002)
14. J.M.R. Parrondo, L. Dínis, Contemp. Phys. **45**, 147–157 (2004)
15. P. Reimann, Phys. Rep. **361**, 57–265 (2002)
16. M. v. Smoluchowski, Z. Phys. Chem **92**, 9 (1917)
17. K. Svoboda, S.M. Block, Cell **77**, 773–784 (1994)
18. J. Villegas, S. Savel'ev, F. Nori, E. Gonzalez, J. Anguita, R. Garcia, J. Vicent, Science **302**, 1188–1191 (2003)

Chapter 5

Application to Biology: The Ear Hair Bundle

Biological processes are excellent candidates to apply our technique. The majority of the biological processes are nonequilibrium processes where measurable currents or flows are present. Examples of nonequilibrium biological processes are the dynamics of molecular motors, DNA packing in viruses or transport through membranes. In all these cases, the energy of the system is in part used to perform work and the rest is dissipated as heat to the thermal environment. By sampling the appropriate degrees of freedom, such as the position of a molecular motor along a microtubule, one can also measure time irreversibility using the KLD.

Active biological processes are those where energy from ATP hydrolyzation is used to perform work. One example is the active ion transport in a membrane against the concentration gradient. The distinction between active and passive (non ATP-consuming) biological process requires in principle the knowledge of many degrees of freedom of the system. In [14] however, Martin et al. discovered a simple method to discriminate between active and passive processes in bullfrog's ear hair bundles. By measuring only the position of the top of the hair bundle in both spontaneous and forced oscillations, they were able to distinguish between alive and dead hair cells by means of the verification of the fluctuation dissipation relationship.

In this chapter, we show how to discriminate between oscillations produced by active or passive mechanisms in ear hair bundles by using trajectories of the top of the hair bundle obtained *only* from the spontaneous oscillations. We apply our new method to estimate the KLD described in Sect. 3.2.2, consisting of a fit of an auto-regressive (AR) model to the original time series and the computation of the KLD between the residuals obtained when that AR model is applied both to the series and to its time reverse.

We measure our KLD rate estimator in both simulated and experimental trajectories of the bullfrog's ear hair bundle. In simulations, we are able to distinguish between different qualitative behaviors and to detect bifurcations. Using experimental data extracted from bullfrog's sacculus, we are able to distinguish between active and passive oscillations and to predict that the minimum amount of dissipation is given by the ATP rate of a single adaptation motor.

This chapter is organized as follows: In Sect. 5.1 we introduce examples of irreversible processes in biology and describe the active processes, focusing on the sound transduction process done by ear hair bundles. In Sect. 5.2 we apply our technique to simulations of the ear hair bundle. Section 5.3 is devoted to distinguish between active and passive oscillations from experimental data of bullfrog's saccus. In Sect. 5.4 we discuss the main results of this chapter.

5.1 Irreversibility in Biology: Ear Hair Bundles

Biological systems are among the most important examples of microscopic systems in which thermal fluctuations play an important role on the dynamics. Molecular motors, bacteria, viruses or ion channels are few examples of small biological systems that are able to perform mechanical work using the energy released from chemical reactions that occur in the environment. The majority of biological processes are driven by a chemical reaction, namely the hydrolyzation of ATP (adenosine triphosphate), a chemical compound formed by a nucleoside and three phosphate groups. ATP hydrolyzation is an exothermic (energetically favorable) chemical reaction in which ATP loses a phosphate group and transforms into ADP (adenosine diphosphate) and an energy of order $\Delta G_{ATP} \simeq -25 \text{ kT}$ is released [6]. The vast majority of biological processes use the energy of the ATP hydrolyzation to drive a biological system out of equilibrium, producing particle or energy flows. Such processes are called *active*. We now cite a few examples of biological systems depicted in Fig. 5.1 that are involved in active processes:

- **Molecular motors:** Molecular biological machines, formed by protein subunits, can transform chemical energy into mechanical work since they are able to move in a given direction even in the presence of forces opposing the direction of the motion of the motor. *Kinesin* motors, for example, can transport vesicles by moving in a regular 8 nm stepping over a protein microtubule against a maximum force of order 6 pN [22]. Muscle contraction is originated by the motion of *Myosin* motors that are connected to *actin* filaments, which form the muscular tissue [8]. *F1-ATPase* is a motor formed by two protein subunits that is capable to rotate in a specific direction and plays a key role in the synthesis of ATP in photosynthesis [17].
- **Viruses:** Bacteriophages are viruses that can inject DNA inside a bacteria, which produces a replication cycle of the virus and the death of bacteria. It has been observed that bacteriophage ϕ 29 can package their replicated DNA before infecting new cells. This packing is an irreversible process which can be performed even against large opposing forces of the order of magnitude of 50 pN [20].
- **Ion channels:** Ions can be transported through biological membranes by diffusion when the ionic concentration on the two sides of the membrane is different. Moreover, ions can also be transported *against* the concentration gradient using the energy of ATP hydrolyzation, for example in the sodium-potassium pump, which is involved in the transmission of electric signals in nerves [2].

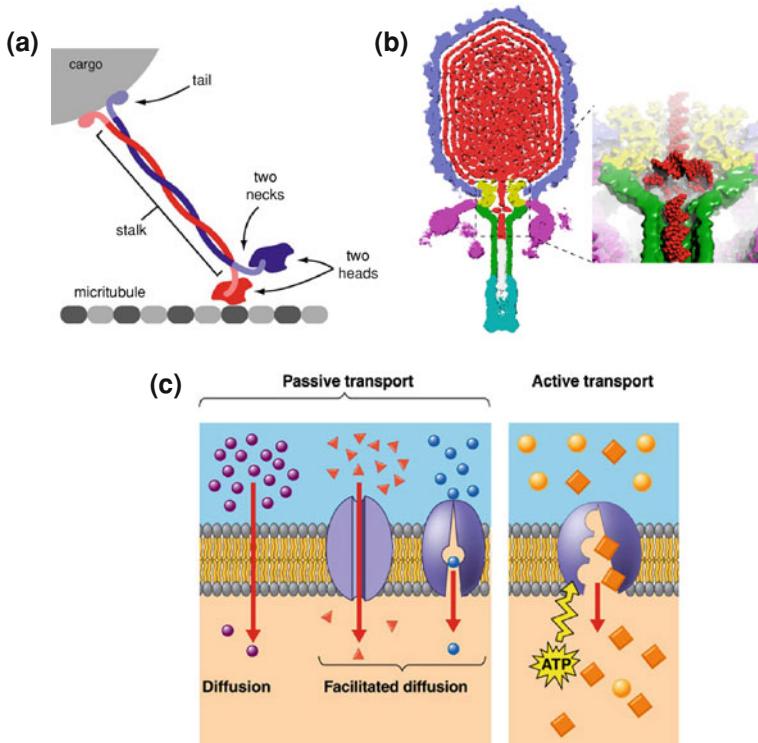


Fig. 5.1 Examples of active processes in biology. **a** A Kinesin motor, formed by two protein subunits, can transport a cargo along a microtubule in 8-nm steps. Picture taken from [10]. **b** Bacteriophage ϕ 29 virus. DNA (in red) is highly packed inside the capsid of the virus. Picture taken from [18]. **c** Passive and active transport of ions through a membrane. Picture taken from [9]

Active processes are also responsible of biological processes involved in hearing [7]. Mechanical stimuli produced by sound are amplified in the inner ear. The active *amplifier* is formed by an ensemble of cells located in the cochlea, the ear hair cells, which use the ATP hydrolysis as energy input and whose structure and function we now discuss.

In vertebrates, the hearing process starts when the sound enters into the external auditory canal and hits the ear drum. When the ear drum is stimulated, the bones located in the middle ear move in a synchronous way. The last bone hits another membrane connected to the cochlea, which is located in the inner ear and filled by an aqueous solution with a high concentration of potassium called *endolymph* (see Fig. 5.2a). Inside the cochlea, the spiral organ of Corti is responsible of the transduction of the sound. Within the spiral organ of Corti, *hair cells* play a key role in the transduction of the sound since they transform the mechanical impulse into an electric signal that travels to the auditory nerves.

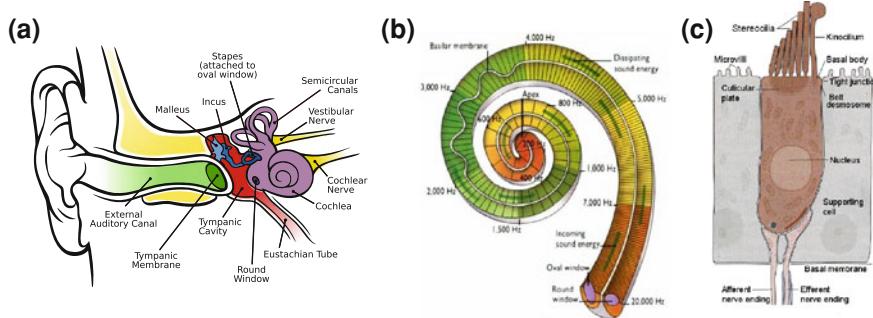


Fig. 5.2 Auditory system in vertebrates. **a** Human auditory system. Eardrum (green) hits the bones located in the middle ear (blue) which stimulate the oval window in the beginning of the cochlea (purple). Picture taken from [3]. **b** Distribution of hair cells of different characteristic frequencies inside the cochlea. Picture taken from [1]. **c** Ear hair cell. The ensemble of stereocilia above the basilar membrane form the hair bundle. Picture taken from [4]

Hair cells have a high frequency selectivity. Different hair cells are optimized to process sounds of different amplitudes and frequencies, and they are distributed in the cochlea according to their characteristic frequency as shown in Fig. 5.2b. The structure of hair cells is shown in Fig. 5.2c. An ensemble of actin filaments called *stereocilia* bend at the base of the apical surface of the hair cells and form the *hair bundles*. Hair bundles are stimulated by mechanical stimuli that enter into the cochlea and the collective oscillations of the stereocilia produce a flow of Ca^{2+} ions through the cell body which is the origin of the electrical impulse transmitted to the auditory nerves.

Ear hair bundles are therefore the final responsible of the transduction of the sound and our object of study. They are formed by an ensemble of 20–300 cylindrical actin stereocilia which are connected by flexible tip links, which are composed essentially by a protein called cadherin (see Fig. 5.3a, b). Each stereocilium has ion channels that can be open or closed and are permeable to Ca^{2+} ions. When the sound is propagated to the cochlea, stereocilia bend at their base and the hair bundle is deflected. This motion is accompanied by a shearing between stereocilia which is thought to open the ion channels. As a result, a current of Ca^{2+} ions towards the cell body is originated. Moreover, the Ca^{2+} concentration regulates the motion of myosin motors that can rectify or amplify the motion of each stereocilium in a mechanism called *adaptation* [7, 11].

From the point of view of biophysics, the dynamics of hair bundles has four main characteristics which can only be explained by the presence of active processes [7]:

- **Amplification** of mechanical stimuli has been observed in bullfrog's sacculus. Amplification of a sinusoidal stimulation of nanometer scale amplitude and frequencies of the order of Hertz has been reported in bullfrog's sacculus hair bundles [12]. The amplification factor can be even of order 10.

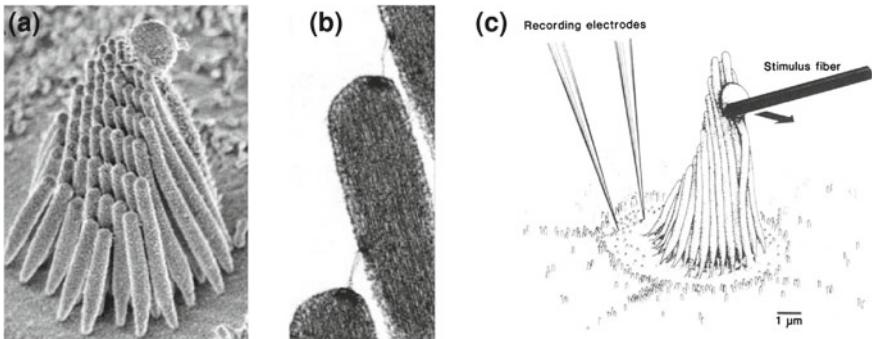


Fig. 5.3 Microscopic structure of the ear hair bundle. **a** Ear hair bundle of bullfrog's sacculus. The distance from bottom to the top is $\sim 6.7 \mu\text{m}$. **b** Magnified picture of two stereocilia connected by a tip-link. **c** Scheme of a ear hair bundle controlled by a stimulus fiber where two electrodes are placed at its base to measure the transduction electric current. Pictures taken from [5, 11]

- **Frequency selectivity:** Different hair cells are adapted to different oscillation frequencies. Frequency tuning is very sharp in ear hair bundles, as it was observed in frogs [13], where hair cells tuned in the 5–50 Hz range.
- **Compressive nonlinearity:** The relation between sensitivity to an external force and the amplitude of the force is non linear. For example, in mammals, a relative increase of 1 million times in the amplitude of the stimulus increases the amplitude of the response of the hair bundle only by a factor 100 [7]. In frogs, the relation between sensitivity and amplitude of the stimulation is linear below amplitudes of 5 nm and above 100 nm but behaves like a power law of power $-2/3$ in the range 5–100 nm [13].
- **Spontaneous otoacoustic emission:** Hair bundles can produce sound in a quiet environment as well as produce sounds at frequencies that are different to the stimulation frequency. Spontaneous oscillations of hair bundles has been observed in frog's sacculus [15], and this motion has been discarded to be caused only by thermal fluctuations but also because of other active mechanisms [14].

These four characteristics of the dynamics suggest that the dynamics of the hair bundle can be described by a dynamical system operating near a *Hopf bifurcation*, which is a mathematical model that shares the aforementioned properties. A bifurcation occurs when a qualitative change on the behavior of a dynamical system is produced when the value of one parameter of the system changes. Specifically, a *Hopf bifurcation* occurs when a stable point of a system transforms into a limit cycle with an unstable point inside the limit cycle [21]. In other words, a system experiences a Hopf bifurcation when its dynamics changes from a damped oscillation (stable point) towards a sustained oscillation of constant amplitude (limit cycle) or vice versa. We notice that the Hopf bifurcation can be subcritical or supercritical, as illustrated in Fig. 5.4. Recently, both linear [14] and nonlinear [16] models with a Hopf bifurcation have been used to describe the dynamics of hair bundles.

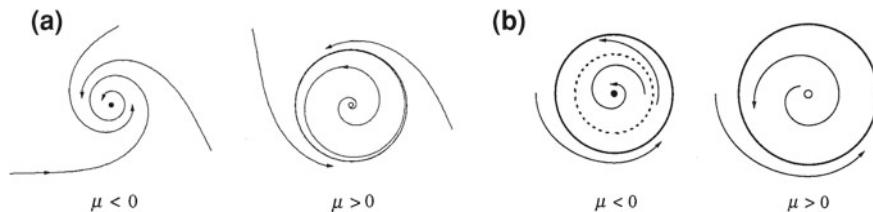


Fig. 5.4 Supercritical and subcritical Hopf bifurcations. **a** Example of a supercritical Hopf bifurcation described by the differential equations for the polar coordinates $\dot{r} = \mu r - r^3$; $\dot{\theta} = \omega + br^2$. The bifurcation occurs when μ changes from a negative value to a positive value. When μ changes from a negative value to a positive value, a stable fixed point turns into an unstable fixed point surrounded by a stable limit cycle. **b** Example of a subcritical Hopf bifurcation described by $\dot{r} = \mu r + r^3 - r^5$; $\dot{\theta} = \omega + br^2$. When μ changes from a negative to a positive value a unstable limit cycle (dashed circle) shrinks into an unstable fixed point. The system oscillates when it reaches the stable outer limit cycle [21]

The biophysical description of the transduction of the sound at the molecular level can be done by taking into account mechanical properties of the main constituents of the hair bundle. Unlike the vision or the olfaction, sound amplification can be explained without the intervention of many intermediate biochemical reactions. Instead of that, the *gating-spring model* explains the transduction of the sound with the contributions of three mechanical elements: stereocilia, gating-springs and transduction channels [11]. The basilar membrane of hair cells is resistant to stereociliar bending, which is quantified by a stereociliar pivot stiffness, K_{sp} . Gating-springs are mechanical elements that connect ion channels of different stereocilia. When the bundle deflects, gating-springs are stretched and transduction channels open. The resistance of gating-springs to stretch is measured by the gating-spring stiffness, K_{gs} . Tip links are thought to be a portion of each gating-spring. Gating of ion channels is stochastic and produces the shortening of the gating-spring by a few nanometers. This stochastic change of the force between stereocilia enters into the gating-spring model as a nonlinear term that is essential to explain the *gating compliance*, that is, the negative value of the gating-spring stiffness for small displacements of the bundle. Figure 5.5 sketches the gating-spring mechanism in hair bundles.

Apart from pure mechanical components, hair bundles need an active energy source in order to perform work. Motor proteins present inside the stereocilia are responsible of the active process that enhances the amplification of the sound [7]. Moreover, molecular motors play also a key role on the *adaptation* of the hair bundle, that is, the reset of the responsiveness of the hair bundle after strong stimuli. Actin filaments that form stereocilia can be stretched or compressed because of the action of myosin-1c adaptation motors as shown in Fig. 5.5b. The adaptation, and therefore motor activity, is strongly dependent on Ca^{2+} concentration, despite the molecular description of the Ca^{2+} is still not known in detail. The simplest model [11] expresses the active force exerted by the motors, f_a , in terms of the calcium concentration inside the hair cell C with a linear relation

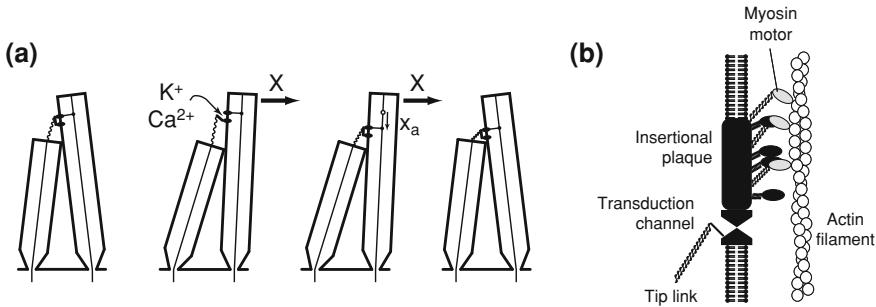


Fig. 5.5 Gating-spring model of the hair bundle. **a** Sketch of two stereocilia linked by a gating spring. When the two stereocilia are deflected, the gating-spring stretches and opens the transduction channel, which originates a current of ions towards the intracellular medium. The ion current activates the adaptation motors that relax the spring to its original extension. **b** Sketch of the structure and location of the adaptation Myosin motors that are responsible of the active processes in the hair bundles. Myosin motors are connected to the outer membrane of the stereocilia and move along actin filaments located inside stereocilia

$$f_a = f_{\max}(1 - SP_o), \quad (5.1)$$

where P_o is the open channel probability, f_{\max} is the maximum force that can be exerted by the motors, and S is the strength of Ca^{2+} in the motor force. S is defined as

$$S \equiv -\frac{C_M}{f_{\max}} \frac{df_a}{dC}, \quad (5.2)$$

where C_M is the maximum Ca^{2+} concentration in the motor site.

The detection of active processes in hair bundles requires in principle knowing many degrees of freedom of the system, such as the position of the top of the hair bundle, the position of the collection of motor proteins or the Ca^{2+} concentration. Ear hair bundles can oscillate spontaneously in two different manners: First, they can oscillate due to the active processes such as motor activity, and secondly, they can perform passive oscillations in the absence of any active process but due to thermal fluctuations of the environment. If one can only measure experimentally the position of the top of the bundle as a function of time, distinguishing between active and passive stochastic oscillations is not straightforward, since in both cases the trajectory of the top of the bundle is stochastic and oscillatory. However, in [14] Martin et al. were able to distinguish between active and passive oscillations in bullfrog's ear hair bundle using only the information of the position of the top of the bundle. By taking into account the fluctuations of the position of top of the bundle in both spontaneous and forced oscillations, they measured the effective temperature that satisfies the fluctuation-dissipation relationship (FDR) (1.41). The temperature that the environment might have in order to sustain in equilibrium the correlations of the position of the top of the bundle (in Fourier space) $\tilde{C}_x(\omega)$ with a response to a weak external force $\tilde{\chi}(\omega)$ is

$$\frac{T_{\text{eff}}(\omega)}{T} = \frac{\omega \tilde{C}_x(\omega)}{2kT \text{Im}[\tilde{\chi}(\omega)]}. \quad (5.3)$$

This effective temperature is a measure of the degree of violation of the FDR. In [14], it was found that the effective temperature of passive (dead) cells equals to the temperature of the environment, $\frac{T_{\text{eff}}(\omega)}{T} \simeq 1$ for all frequencies ω whereas the effective temperature of active (alive) cells does not $\frac{T_{\text{eff}}(\omega)}{T} \neq 1$. Moreover, in the active case, the dependence of the effective temperature on the frequency can be fitted to that of a linear stochastic oscillator model operating near a Hopf bifurcation.

In the next sections, we develop a method to distinguish between active and passive oscillations in ear hair bundles by using the information of the position of the top of the bundle recorded *only* from spontaneous—but not from forced—oscillations. We use the KLD rate of trajectories of the hair bundles from both experiments and simulations to do this distinction.

5.2 Simulations

We first check if we can distinguish between active and passive oscillations by measuring the KLD rate between forward and time-reversed trajectories of the top of the hair bundle obtained from numerical simulations. We simulate the dynamics of the hair bundle by using the nonlinear stochastic oscillator model introduced by Nadrowski et al. in [16]. Nadrowski's model has the main characteristics of the gating-spring model discussed in the previous section and reproduces with high accuracy the experimental values of relevant physical magnitudes of the hair cells, such as the characteristic frequency of the spontaneous oscillations or the sensitivity to an external force. The model is valid for hair cells with characteristic frequencies on the order of magnitude ranging from 1 to 100 Hz. The model is described by three variables: the position of the top of the bundle, X , a second variable that plays the role of the position of the collection of motors, X_a , and the intracellular Ca^{2+} concentration, C . The following set of equations

$$\lambda \frac{dX}{dt} = -K_{\text{gs}}(X - X_a - DP_o) - K_{\text{sp}}X + \eta, \quad (5.4)$$

$$\lambda_a \frac{dX_a}{dt} = K_{\text{gs}}(X - X_a - DP_o) - \gamma N_a f p(C) + \eta_a, \quad (5.5)$$

$$\tau_c \frac{dC}{dt} = -C + C_M P_o + \delta c. \quad (5.6)$$

reproduce the spontaneous oscillations of the hair bundle in the absence of any external force. The spontaneous oscillations are reproduced in the variable X for some specific values of the parameters of the model [16] which we now define. All the ingredients of the gating-spring model discussed in Sect. 5.1 are included in this nonlinear stochastic oscillator model as we now discuss.

The dynamics of X is governed by Eq. (5.4), where λ is the friction coefficient associated to X . K_{gs} and K_{sp} are gating-spring and stereociliar pivots stiffnesses, respectively. $D = d/\gamma$ is the displacement of the bundle's top due to the opening of a channel, which causes an extension of the gating spring by an amount d , γ being a geometrical factor relating d and D . P_o is the probability of the transduction channels to be open, which is assumed to be nonlinear [11]

$$P_o = (1 + Ae^{-(X - X_a)/\delta})^{-1}, \quad (5.7)$$

where A and δ are two constants defined in [16]. The term $-K_{\text{gs}}(X - X_a - DP_o)$ accounts for the deflection of the top of the bundle due to the change of the gating-spring extension. Notice that $K_{\text{gs}}DP_o$ is a term that depends on the open channel probability and is positive, therefore it can model the gating compliance phenomenon. The term $-K_{\text{sp}}X$ accounts for the resistance of the stereociliar pivots to bend at their base. The last term in (5.4) models the thermal noise, that is, $\eta(t)$ is a Gaussian white noise with zero mean and correlation $\langle \eta(t)\eta(t') \rangle = 2kT\lambda\delta(t - t')$.

Equation (5.5) describes the evolution of X_a . λ_a is the slope of the force-velocity curves of the motors, N_a is the total number of motors present in the hair bundle, f is the force exerted by a single motor and $p(C)$ is the probability for an adaptation motor to be bound to actin, which is assumed to be linear, $p(C) \simeq p_0 + p_1C$. The first term in (5.5), $K_{\text{gs}}(X - X_a - DP_o)$, models the change of X_a produced by the change in the gating-spring extension. The second term in (5.5) models the regulation of the motor activity by Ca^{2+} and is analogous to Eq. (5.1) by identifying the maximum force of the motors with $f_{\text{max}} = \gamma N_a f$ and the strength of the Ca^{2+} feedback with $S = -C_M p_1/p_0$. The noise term for X_a does not satisfy fluctuation-dissipation theorem despite it is also Gaussian with zero mean and correlation. The noise intensity is in this case $\langle \eta(t)\eta(t') \rangle = 2kT_a\lambda_a\delta(t - t')$, with $T_a = 1.5 \text{ T}$. The fact that $T_a > T$ which accounts for an active mechanism that serves as energy source to perform work.

Equation (5.6) describes the relaxation dynamics of the intracellular Ca^{2+} concentration, which is assumed to be initially zero. The relaxation time is denoted by τ_c , and C_M is the maximum concentration of calcium at the location of the motors. δc takes into account the fluctuations in the dynamics of Ca^{2+} , δc being a Gaussian white noise of zero mean and correlation $\langle \delta c(t)\delta c(t') \rangle = (1/2)C_M^2 N^{-1} \tau_c \delta(t - t')$, where N is the number of stereocilia and τ_c is the dwell time of the transduction channels. For the values of the parameters that fit the spontaneous oscillations of bullfrog's hair bundles, one can define an effective temperature in the intensity of calcium concentration fluctuations which is of order $T_c/T \simeq 0.13$, suggesting that Ca^{2+} fluctuations are one order of magnitude below thermal fluctuations [16].

For the specific case of bullfrog's hair bundles, the majority of the parameters listed are known from experimental observations (see Table 1 in [16]). There are only two free parameters on the model: the maximal force produced by the motors in the direction of the movement, $f_{\text{max}} = N_a f p_0$, and the feedback strength of the calcium regulation, defined as $S = -C_M p_1/p_0$. In the absence of noise, the

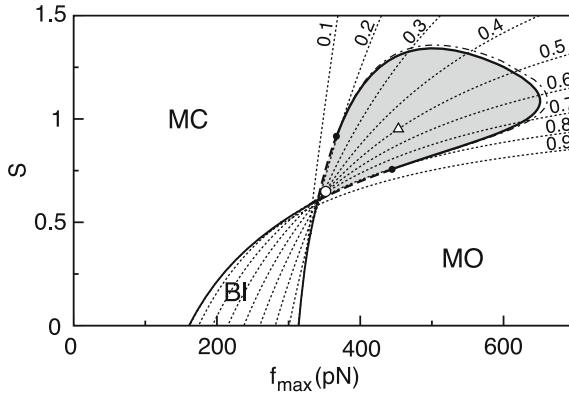


Fig. 5.6 State diagram in the absence of noise in the Nadrowski model as a function of f_{\max} and S . Spontaneous oscillations occur for parameter values in the *dark shaded area*. *Dashed curves* indicate the values of the parameters for which the open channel probability takes a constant value. The *dashed boundary* separating the oscillatory from the monostable (open and closed) regions corresponds to a subcritical Hopf bifurcation whereas the *solid boundary* represents the region where the Hopf bifurcation is supercritical. Picture taken from [16]

oscillator described by this model exhibits different dynamical behaviors depending on the values of f_{\max} and S as shown in Fig. 5.6.

- **Monostable closed (MC):** At low values of the motor force, adaptation motors cannot open the transduction channels and the majority of them are closed. This is also the case when calcium feedback is very high and motors cannot exert forces strong enough to open the channels.
- **Monostable open (MO):** When the force exerted by the motors is very large, the majority of the channels are maintained open.
- **Oscillatory (O):** At intermediate values of Ca^{2+} feedback and motor forces indicated in Fig. 5.6 the system oscillates spontaneously.
- **Bistable (BI):** Also at intermediate values of calcium feedback and motor forces, but in a different region in which calcium feedback is weaker than when oscillations occur, the system is in a bistable region where open and close channels coexist.

In the case of bullfrog's ear hair bundle, the *operating point*, that is, the values of the parameters that reproduce Bullfrog's bundle's spontaneous oscillations, is located at $f_{\max} = 325$ pN and $S = 0.65$ (circle mark in Fig. 5.6), which is very close to the boundary between the oscillatory region and the bistable region. When crossing the boundary between the different regions, the system experiences a Hopf bifurcation. Notice that the bifurcation is subcritical near the operating point (dashed boundary in Fig. 5.6) and it becomes supercritical for large values of f_{\max} (solid boundary in Fig. 5.6). The values of the parameters for which the sensitivity of the system to an external force is maximum is also indicated in Fig. 5.6 with a triangular mark.

We simulate the model described by Eqs. (5.4)–(5.6) for different values of f_{\max} and S using Euler's numerical integration scheme described in Appendix D.1. We first

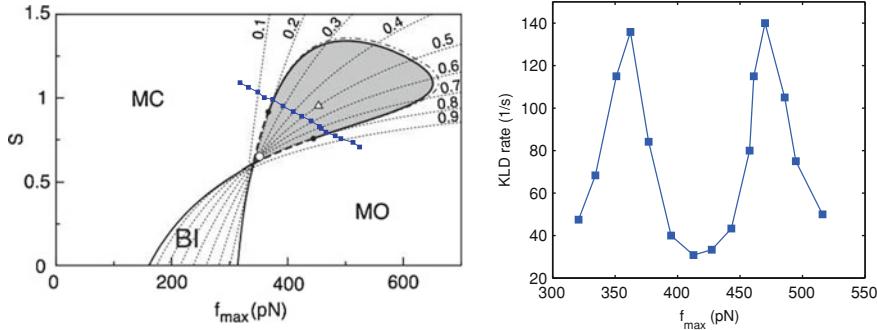


Fig. 5.7 KLD rate of trajectories of the top of the bundle when crossing a supercritical Hopf bifurcation in Nadrowski's model. *Left* Variation of the simulation parameters along a supercritical Hopf bifurcation. *Right* KLD rate of single stationary trajectories of the position of the top of the hair bundle, $X(t)$, with total time of the trajectory $\tau = 1,200$ s and $f_{\text{acq}} = 8.3$ kHz, \hat{d}_{AR}^X , as a function of f_{\max} along the line depicted in the *left panel*

analyze the dynamics of the system when it crosses the bifurcation in the region where it is supercritical. We vary the parameters following the values indicated in the left panel in Fig. 5.7. For each point (f_{\max}, S) in the curve indicated in the figure, we simulate a single trajectory of $n_0 = 10^8$ data points with time step $\Delta t_0 = 0.012$ ms, that is, simulations of total time $\tau_0 = n_0 \Delta t_0 = 1,200$ s. From the simulations, we only sample the position of the top of the bundle, $(X, X_a, C) \rightarrow X$, using a sampling frequency of $f_s = 1/(10\Delta t_0) = 8.3$ kHz (i.e. we sample the trajectories $X(t)$ every 10 data). In the end, we work with time series of $n = 10^7$ data points with time step $\Delta t = 0.12$ ms and total time $\tau = n\Delta t = 1,200$ s. For every trajectory, we measure the KLD rate *per unit of time* with the estimation technique described in Sect. 3.2.2 [Eq. (3.29)]. In the estimation, we first fit the data to an AR(10, 50) model and calculate the KLD using the residual associated to this AR model. This procedure is equivalent to consider blocks of 10 data separated $\tau_{\text{lag}} = l\Delta t = 6$ ms, $\tau_{\text{block}} = k\tau_{\text{lag}} = 60$ ms being the total time of the block, which is much smaller than the period of the oscillations, $\tau_{\text{osc}}, \tau_{\text{block}} \ll \tau_{\text{osc}} = O(1\text{s})$.

We show in the right panel of Fig. 5.7 the value of \hat{d}_{AR}^X (3.29) as a function of f_{\max} along a curve in the state diagram that crosses the supercritical bifurcation shown in the left panel of Fig. 5.7. Our estimator increases when approaching the oscillatory region and has two maxima that occur when the system crosses the supercritical Hopf bifurcation. This result suggests that the time irreversibility in X is maximum in the boundary between the oscillatory and the monostable (open or closed) regions, and therefore KLD can be used to detect the presence of a bifurcation.

An analogous result is obtained when the bifurcation is crossed with a similar curve but in the region where the Hopf bifurcation is subcritical. We simulate trajectories of the same total time and sampling frequency than in the supercritical case. Similarly to what happens in the supercritical Hopf bifurcation, the KLD rate calculated using the position of the top of the bundle \hat{d}_{AR}^X increases when the system approaches the

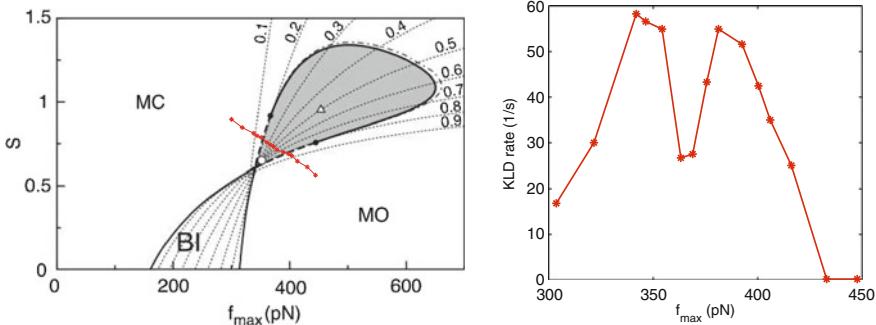


Fig. 5.8 KLD rate of trajectories of the top of the bundle when crossing a subcritical Hopf bifurcation in Nadrowski's model. *Left* Variation of the simulation parameters along a subcritical Hopf bifurcation. *Right* KLD rate of single stationary trajectories of the position of the top of the hair bundle, $X(t)$, with total time of the trajectory $\tau = 1,200$ s and $f_{\text{acq}} = 8.3$ kHz, \hat{d}_{AR}^X , as a function of f_{max} along the line depicted in the *left* panel

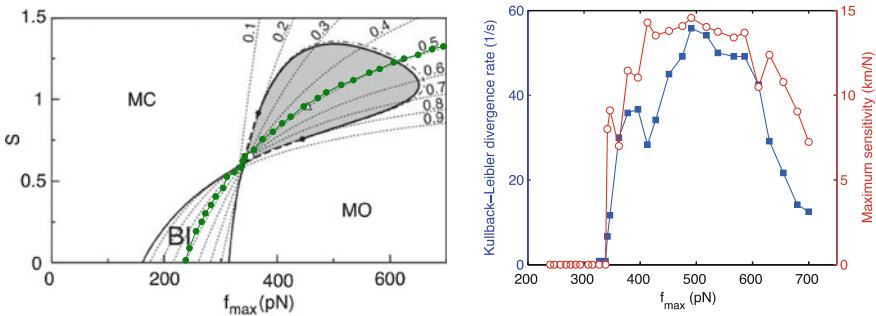


Fig. 5.9 KLD rate of trajectories of the top of the bundle in the bistable to oscillatory transition in Nadrowski's model. *Left* Variation of the simulation parameters along a curve with constant open channel probability $P_o = 0.5$. *Right* KLD rate per unit of time \hat{d}_{AR}^X (blue squares, left axis) and maximum sensitivity (red circles, left axis) as a function of f_{max} for single stationary trajectories of $\tau = 1,200$ s and $f_{\text{acq}} = 8.3$ kHz with the parameters given by the values in the *left* panel

oscillatory region from any monostable region and decreases within the oscillatory region the farther the system is from the bifurcation (see Fig. 5.8). If the bifurcation is crossed twice two maxima are observed in the vicinity of the subcritical bifurcation like in the case of the supercritical bifurcation [cf. Fig. 5.7].

We now analyze the behavior of the KLD when we vary the parameters following a path with constant open channel probability close to $P_o \simeq 0.5$ in the state diagram in the absence of noise (see left panel of Fig. 5.9). For every value of (f_{max}, S) in the state diagram along the curve shown in the left panel of Fig. 5.9, we simulate the hair bundle dynamics using Nadrowski's model and generate trajectories of $\tau = 1,200$ s with simulation time step of $\Delta t_0 = 0.012$ ms. We restrict ourselves to the information given by the position X , and sample the trajectories at $f_s = 8.3$ kHz, obtaining, for

each value of the parameters, a single stationary trajectory $X(t)$ that contains $n = 10^7$ data points each of them sampled every $\Delta t = 0.12$ ms. We measure the KLD rate estimator \hat{d}_{AR}^X as well as the maximum sensitivity of the hair bundle for every point in the parameter space curve. When increasing the value of S , the system crosses the bifurcation going from the bistable to the oscillatory regime. For the values of f_{\max} and S indicated in the left panel of Fig. 5.9, \hat{d}_{AR}^X increases with the force for moderate values of f_{\max} and it exhibits a maximum that occurs near the maximum sensitivity point.

We now focus on the problem of distinguishing between active and passive oscillations of the hair bundle using only the information from the position of top of the bundle. In this case, $X(t)$ is also a partial description of the dynamics, and it is not straightforward to guess from the time traces if they are generated by an active or a passive oscillator. We first notice that, if we neglect calcium dynamics, Nadrowski's model can be expressed in terms of a potential $V(X, X_a)$ that depends on the variables X and X_a ,

$$\lambda \frac{dX}{dt} \simeq -\frac{\partial V(X, X_a)}{\partial X} + \eta, \quad (5.8)$$

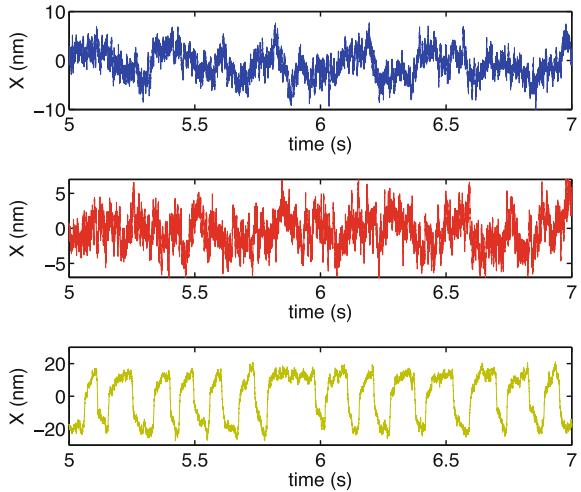
$$\lambda_a \frac{dX_a}{dt} \simeq -\frac{\partial V(X, X_a)}{\partial X_a} - \gamma N_a f p(C) + \eta_a, \quad (5.9)$$

where

$$V(X, X_a) = \frac{K_{\text{gs}}}{2}(X - X_a - DP_o)^2 + K_{\text{sp}}X^2. \quad (5.10)$$

Notice that the above approximation is valid if the dependence of P_o on X and X_a given by Eq. (5.7) is ignored. In this approximation, the majority of the terms in Nadrowski's model are conservative. The non conservative term, $-\gamma N_a f p(C)$, accounts for the active process generating the spontaneous oscillations. As a consequence, one can simulate passive oscillations by setting $f_{\max} = 0$, which makes the active term to vanish. However, the richness of this model allows us to analyze the system in a variety of situations aside from the active and passive cases. We now study how the KLD rate per unit of time scales with time for the different regions of the state diagram of Nadrowski's model and whether or not we can distinguish between them by means of the behavior of the KLD. In Fig. 5.10 we sample the position of the top of the bundle $X(t)$ from trajectories simulated using Nadrowski's model. In all cases, we generate trajectories with $\Delta t_0 = 0.012$ ms and sample the trajectories with an acquisition frequency of $f_{\text{acq}} = 8.3$ kHz. In the top and the middle panel of the figure we show the position of the top of the bundle obtained from trajectories generated in the monostable region where the majority of the ion channels are open or closed (MO and MC). In these regions, the oscillator does not perform spontaneous oscillations similar to the ones observed experimentally [15]. The third situation (bottom panel in Fig. 5.10) corresponds to the operating point of the hair bundle, where spontaneous oscillations are produced by active processes.

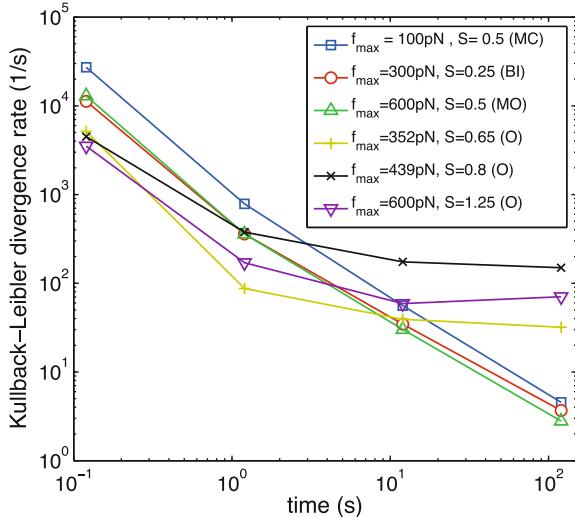
Fig. 5.10 Position of the top of the hair bundle (in nm) as a function of time (in seconds) in simulations of the Nadrowski's model. *Top* $f_{\max} = 100 \text{ pN}$ and $S = 0.5$, corresponding to a monostable closed situation. *Medium* $f_{\max} = 300 \text{ pN}$ and $S = 0.25$ corresponding to a bistable oscillation. *Bottom* $f_{\max} = 325 \text{ pN}$ and $S = 0.65$, corresponding to the operating point of the hair bundle



Statistically speaking, the latter time series is in appearance more irreversible than the other two because of the presence of irreversible patterns in the series. More specifically, in the oscillation shown in the bottom panel of Fig. 5.10, slow excursions in the negative direction of X are interrupted by fast jumps in the positive direction in accordance to the structure of the spontaneous oscillations observed in bullfrog's hair bundles [15].

We analyze how our KLD rate estimator $\hat{d}_{\text{AR}}^{\mathbf{x}}$ scales with time when we consider time series of different total time t , $\{X(t')\}_{t'=0}^t$, in an equivalent analysis to the one performed in Sect. 3.3, where we used the scaling KLD estimator based on visibility algorithm to distinguish between equilibrium and NESS. We denote by $\hat{d}_{\text{AR}}^{\mathbf{x}}(t)$ the KLD rate estimator using the residual of the $AR(k, l)$ model calculated using a single stationary trajectory of the top of the bundle of total time t . We investigate the behavior of $\hat{d}_{\text{AR}}^{\mathbf{x}}(t)$ for different regimes of the oscillator by using the same estimator as before, measuring the residual of an $AR(10, 50)$ model. We generate trajectories of $\tau = 125 \text{ s}$, with $\Delta t_0 = 0.012 \text{ ms}$ and sampled with $f_{\text{acq}} = 8.3 \text{ kHz}$, and measure $\hat{d}_{\text{AR}}^{\mathbf{x}}$ of subsequences of total time $t < \tau$, starting at $t' = 0$ and ending in $t' = t$, $\{X(t')\}_{t'=0}^t$. In Fig. 5.11 we show the value of $\hat{d}_{\text{AR}}^{\mathbf{x}}(t)$ as a function of time t in double logarithmic scale for different values of time $t = 0.125, 1.25, 12.5$ and 125 s . The scaling of $\hat{d}_{\text{AR}}^{\mathbf{x}}(t)$ with time is different depending on which region of the state diagram the oscillator is, as shown in Fig. 5.11. If the oscillator is in the monostable state (both open or closed) or in the bistable state, $\hat{d}_{\text{AR}}^{\mathbf{x}}(t)$ scales like $1/t$ and tends to 0 when t is large (see Fig. 5.11). Conversely, in the oscillatory regime, $\hat{d}_{\text{AR}}^{\mathbf{x}}(t)$ also decreases with time but it tends asymptotically to a constant positive value when time increases. Therefore, only in the regime where spontaneous oscillations are reproduced, the KLD tends to a positive value which suggests that the system is in a NESS.

Fig. 5.11 Scaling of the KLD rate with time of the trajectory in the different regions of Nadrowski's model. KLD rate as a function of time $\hat{d}_{\text{AR}}^x(t)$ using the residual of an $AR(10, 50)$ model for trajectories of the bundle's top $X(t)$ as a function of the total time of the trajectory (in s). For each of the 6 cases shown in the figure, we simulate a single stationary trajectory of total time $\tau = 125$ s which is sampled at a rate of $\Delta t = 0.12$ ms



The value of the plateau of the KLD rate for long times is of the order of 100 kT/s . This result can be related to the average dissipation per unit of time of the system provided the bound (2.34). In this case, $\frac{\langle \dot{W}_{\text{diss}} \rangle}{kT} \geq \hat{d}_{\text{AR}}^x$, and \hat{d}_{AR}^x tends to a value whose order of magnitude is 100 s^{-1} which implies that the average dissipation of the oscillator is at least 100 kT/s . We notice that the ATP consumption rate for a myosin adaptation motor is of order $r_{\text{ATP}} = 6 \text{ s}^{-1}$ (see Table 13.1 in [6]) and therefore the energy consumption rate of the ATP hydrolysis is of order

$$\dot{W}_{\text{diss}} \simeq \Delta G_{\text{ATP}} \times r_{\text{ATP}} = \frac{25kT}{\text{ATP}} \times \frac{6\text{ATP}}{s} = 150 \text{ kT/s.} \quad (5.11)$$

Consequently, our KLD rate estimator predicts that the minimum irreversibility to sustain spontaneous oscillations in Nadrowski's model is of the order of magnitude of the dissipation rate of a single adaptation motor.

We notice that the KLD rate measured using the information of the position can be misleading when distinguishing between active and passive processes. Passive oscillations are reproduced in Nadrowski's model when $f_{\text{max}} = 0$, which is in the MC region, where the KLD vanishes. However, not all the active oscillations $f_{\text{max}} > 0$ yield a positive value of the KLD rate for trajectories of the bundle's top, and in some cases such as the blue, red and green curves in Fig. 5.11 one might think that oscillations are passive since $\hat{d}_{\text{AR}}^x(t) \rightarrow 0$ when t is large. Therefore, only $\hat{d}_{\text{AR}}^x(t)$ tends to a positive value we can ensure that the trajectories generated by the Nadrowski's model are generated by underlying active processes. This case is similar to a discrete flashing ratchet with infinite switch rate, where, even using full information of the system, the KLD rate vanishes whereas the dissipation does not [19].

5.3 Application to Experimental Data

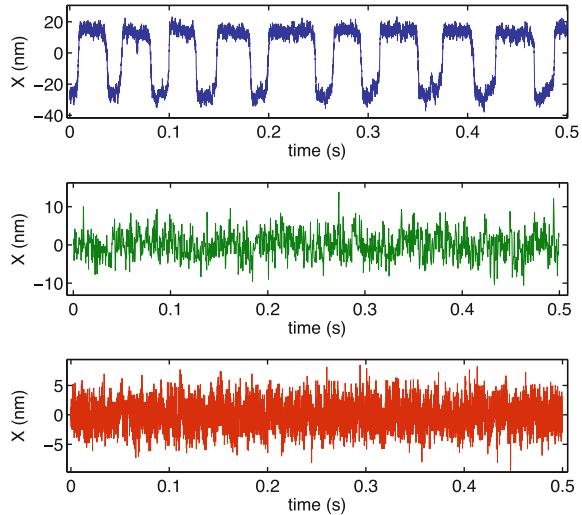
We now study if our technique can be applied to distinguish between active and passive oscillations of hair bundles using experimental data from bullfrog's sacculus. The experimental data was obtained by J. Barral and P. Martin in *Laboratoire de Physico-Chimie Curie (Institut Curie)*.¹ The data was obtained with the same protocol as that described in [15]. The ear is dissected from an American bullfrog (*Rana catesbeiana*) after the animal has been sacrificed (a needle is inserted into the brain from the back of the head). Bullfrog's sacculus is mounted in a two-compartment chamber where cell bodies are immersed in a saline solution and hair cells in a second compartment filled by an artificial endolymph. Experiments are done at a room temperature $T \sim 21^\circ\text{C}$. Video microscopy is used to find active hair bundles that oscillate spontaneously. A flexible glass fiber of $\sim 500\text{ nm}$ diameter is attached to the top of the bundle as sketched in Fig. 5.3c and monitored at a magnification of $1,000\times$ on a photodiode. The displacement monitor provides a linear relationship between the displacement of the tip of the fiber and the voltage of the photodiode which allows to measure the position of the top of the bundle with an accuracy of $\sim 1\text{ nm}$.

Using the experimental procedure described above, we monitored the position of the top of the bundle in spontaneous oscillations that lasted $\sim 100\text{ s}$ with a peak-to-peak amplitude of $\sim 10\text{ nm}$. We use the data of different hair cells that oscillate with different characteristic frequencies of the order of Hz in all cases. Sampling rate was of the order of kHz in all cases. On the other hand, oscillations of dead cells and experimental noise were also recorded to check if we can distinguish them from the oscillations of active (living) cells by means of the KLD rate of the trajectories of the position of the top of the bundle.

First, we measured the position of the top of a hair bundle $X(t)$ of a cell that exhibits spontaneous oscillations of amplitude $\sim 20\text{ nm}$ using an acquisition frequency of 14 kHz. In the top panel of Fig. 5.12 we show a sample of 0.5 s of such trajectory, where the top of the bundle oscillates between two equilibrium positions. We notice that in this case, $X(t)$ resembles to a square wave which seems to be more reversible than the series obtained simulating Nadrowski's model in the oscillatory region (cf. bottom panel in Fig. 5.10). Secondly, we added gentamicin to the endolymph of the chamber where hair cells are exposed. As indicated in [15], gentamicin blocks transduction channels and produces that the oscillations of the bundle become Hookean of amplitude $\sim 5\text{ nm}$ as shown in the middle panel of Fig. 5.12. As a third case study, we use the data of hair cells recorded several minutes after performing spontaneous oscillations. After this time, hair cells die and oscillate in a passive manner because of thermal fluctuations. In this case, the position of the bundle is also stochastic and oscillatory with an amplitude of $\sim 10\text{ nm}$ as shown in the bottom panel of Fig. 5.12.

¹ Full address: Laboratoire de Physico-Chimie Curie, Unité Mixte de Recherche 168, Institut Curie, 26 rue d'Ulm, F-75248 Paris Cedex 05, France.

Fig. 5.12 Displacement of the top of the bundle with respect to its average value (in nm) as a function of time (in s) for different cells. *Top* hair cell that exhibits active oscillations, recorded with an acquisition frequency of 14 kHz. *Middle* same cell after the exposure to gentamicin also sampled at 14 kHz. *Bottom* passive cell displacements, sampled at 2.5 kHz



We now analyze how the KLD rate of a single stationary trajectory of the top of the bundle $X(t)$ —obtained in a single experiment—scales with time for different hair cells. We use the same estimator of the KLD as that used in simulations in Sect. 5.2, that is, \hat{d}_{AR}^X using the residual of an $AR(10, 50)$ model. Apart from the cells whose trajectories are sampled in Fig. 5.12, we also consider cases where we measured the position of the bundle in two other cells that oscillate spontaneously at different frequencies. For completion, we also recorded the experimental noisy signal from the photodiode. Figure 5.13 is a summary of our results where we show the value of \hat{d}_{AR}^X for the different cases as a function of time t (details of every cell are specified in the caption). We observe that for the three cases in which the cell oscillates spontaneously, the KLD rate saturates for long times to a positive value that is around $\sim 100 \text{ s}^{-1}$. The KLD of the rest of the signals, including the passive oscillation, experimental noise and the oscillation of one of the cells in the presence of gentamicin, tend to zero like $1/t$ as shown in Fig. 5.13.

We can discriminate between oscillations generated with active and passive processes using the scaling of \hat{d}_{AR}^X with time since it tends to zero for passive processes and to a positive value for oscillations produced by active processes as we illustrate in Fig. 5.14. We remark that our method correctly predicts that the hair bundle is dissipating work even if guessing whether or not $X(t)$ is irreversible from visual inspection of the shape of the time series is not straightforward (as in the series of the top panel in Fig. 5.12).

In the active case, the value at which the KLD rate saturates for active cells is of order $\hat{d}_{\text{AR}}^X \sim 100 \text{ s}^{-1}$, which corresponds to an average dissipation per unit of time of $\langle \dot{W}_{\text{diss}} \rangle \sim 100 \text{ KT/s}$ as discussed in the previous section. This value bounds from below the average energy consumed from ATP hydrolyzation per unit of time of a

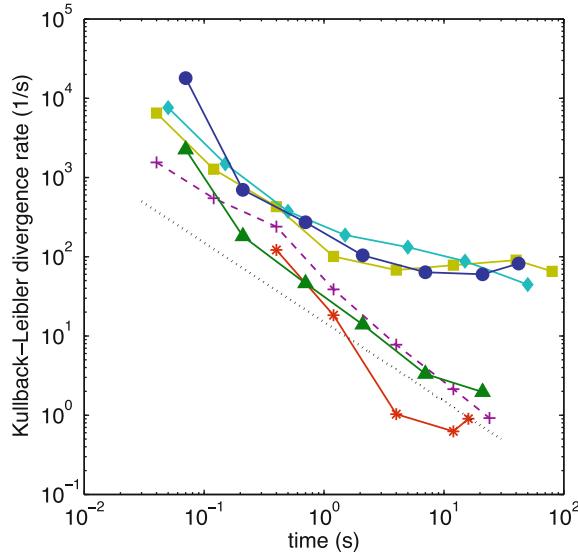


Fig. 5.13 Scaling of the KLD rate estimator $\hat{d}_{\text{AR}}^{\text{X}}$ with time for spontaneous oscillations of the top of the bundle obtained from bullfrog's sacculus. We measure the estimator of the KLD rate defined in Eq. (3.29) for subsequences of time t of a single experiment, $\{X(t')\}_{t'=0}^t$, $\hat{d}_{\text{AR}}^{\text{X}}(t)$, using the residual of an $AR(10, 50)$ model. Yellow squares cell 2 (active), with $\tau = 90$ s and $f_{\text{acq}} = 25$ kHz. Cyan diamonds cell 3 (active), with $\tau = 50$ s and $f_{\text{acq}} = 20$ kHz. Green triangles cell 1 with same total recording time and acquisition frequency in gentamicin endolymph. Red stars passive cell, with $\tau = 30$ s and $f_{\text{acq}} = 2.5$ kHz. The KLD of the experimental noise from the photodiode is also shown in dashed lines (Magenta “+”, with $\tau = 30$ s and $f_{\text{acq}} = 25$ kHz). The dotted black line is added to guide that has a dependence of 1/time to guide the eye

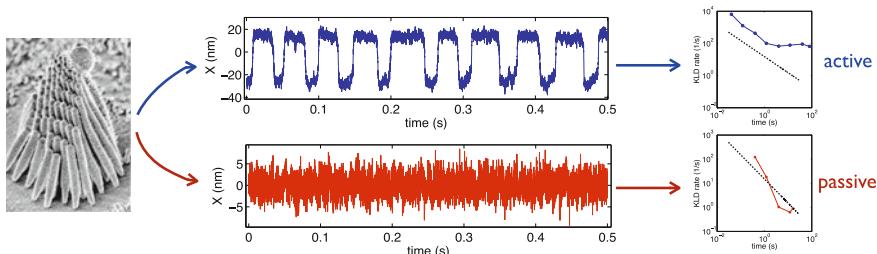


Fig. 5.14 Illustration of the method used to distinguish between active and passive oscillations in the ear hair bundle. The recording of the position of the top of the bundle is a stochastic time series whose KLD scales with time differently when it is oscillating in an active or a passive manner (it tends to a positive value in case of active oscillations and it goes to zero like 1/time for passive oscillations)

single adaptation motor as proved before in Eq. (5.11). Therefore using the KLD rate of trajectories of the top of the bundle we can predict the minimum irreversibility needed to sustain spontaneous oscillations in bullfrog's ear hair bundle.

5.4 Conclusions

In this Chapter, we have applied our new estimator of the KLD rate that uses the residual of an auto regressive (AR) model to measure the time irreversibility of real-valued time series. Biological systems in the stationary state produce stochastic time series whose time irreversibility can be measured with this estimator of the KLD rate and relate its value with the dissipation rate or the ATP consumption rate of the system. In particular, we have applied our technique to distinguish between oscillations produced by active and passive processes in ear hair bundles.

We have first applied our technique to simulations produced by a nonlinear stochastic model of the hair bundle. We have measured the KLD rate of stationary trajectories of the top hair bundle obtained from numerical simulations. Although the position of the top of the bundle is a partial description of the system, our estimator of the KLD rate detects when the system crosses a Hopf bifurcation when varying the parameters of the model. Moreover, we are also able to distinguish between different dynamical behaviors, such as oscillatory or monostable, in terms of the scaling our KLD rate estimator. In simulations, our measure of irreversibility underestimates the dissipation of active oscillations in the monostable region, whereas it coincides with the average ATP consumption rate of a single adaptation motor in active oscillations.

We have also applied the same technique to experimental time series obtained from the recording spontaneous oscillations in hair bundles extracted from bullfrog's sacculus. We have measured the position of the top of hair bundles in spontaneous oscillations in both active (alive) and passive (dead) hair cells. We are able to discriminate, by means of the value of the KLD, between active and passive hair cells of different oscillation frequencies and sampled with different data acquisition frequencies. This distinction can be done by measuring the scaling of our KLD rate estimator measured using subsequences of a single experiment of no more than 100 s, and succeeds to predict irreversibility in time series that look reversible at low frequencies. Our KLD estimator tends to zero in the case of passive oscillations whereas it tends asymptotically to a positive value in the case of spontaneous oscillations generated by active processes. The saturating value of the KLD rate at long times (and therefore our prediction of the minimum irreversibility to sustain the spontaneous oscillations) is very close to the energy consumption rate by a single myosin adaptation motor, which is the biological active process responsible of the spontaneous oscillations.

References

1. <http://nl.bu.edu/wp-content/uploads/2011/10/i10-85-cochlea22.jpg>
2. E. Dunham, I. Glynn, J. Physiol. **156**, 274–293 (1961)
3. http://upload.wikimedia.org/wikipedia/commons/d/d2/Anatomy_of_the_Human_Ear.svg
4. http://michaeldmann.net/pix_4b/hair_cell.gif
5. J. Howard, A. Hudspeth, Proc. Nat. Acad. Sci. **84**, 3064–3068 (1987)
6. J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer Associates, Sunderland, 2001)
7. A. Hudspeth, Neuron **59**, 530–545 (2008)
8. H. Huxley, Science **164**, 1356–1366 (1969)
9. <http://nestbio.blogspot.com.es/2010/11/passive-and-active-transport.html>
10. <http://www.stanford.edu/group/blocklab/ScienceLimping/StructureCartoon.jpg>
11. G.A. Manley, R.R. Fay, A.N. Popper, *Active Processes and Otoacoustic Emissions*, vol. 30 (Springer, New York, 2008)
12. P. Martin, A. Hudspeth, Proc. Nat. Acad. Sci. **96**, 14306–14311 (1999)
13. P. Martin, A. Hudspeth, Proc. Nat. Acad. Sci. **98**, 14386–14391 (2001)
14. P. Martin, A. Hudspeth, F. Jülicher, Proc. Nat. Acad. Sci. **98**, 14380–14385 (2001)
15. P. Martin, D. Bozovic, Y. Choe, A. Hudspeth, J. Neurosci. **23**, 4533–4548 (2003)
16. B. Nadrowski, P. Martin, F. Jülicher, Proc. Nat. Acad. Sci. USA **101**, 12195–12200 (2004)
17. H. Noji, R. Yasuda, M. Yoshida, K. Kinoshita, Nature **386**, 299–302 (1997)
18. <http://biomedicalcomputationreview.org/content/viral-closeup>
19. É. Roldán, J.M.R. Parrondo, Phys. Rev. Lett. **105**, 150607 (2010)
20. D.E. Smith, S.J. Tans, S.B. Smith, S. Grimes, D.L. Anderson, C. Bustamante, Nature **413**, 748–752 (2001)
21. S. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering* (Perseus Books Group, New York, 2001)
22. K. Svoboda, C.F. Schmidt, B.J. Schnapp, S.M. Block et al., Nature **365**, 721–727 (1993)

Part III

Experimental Tests and Applications

of Stochastic Thermodynamics

Chapter 6

Energetics of Symmetry Breaking

A symmetry breaking (SB) involves an abrupt change in the set of microstates that a system can explore. This change has unavoidable thermodynamic implications. According to Boltzmann's microscopic interpretation of entropy, a shrinkage of the set of compatible states implies a decrease of entropy, which eventually needs to be compensated by a dissipation of heat and consequently requires work. Examples are the compression of a gas and the erasure of information. On the other hand, in a spontaneous SB, the available phase space volume changes without the need of work, yielding an apparent decrease of entropy. Here we show that this decrease of entropy is a key ingredient in the Szilard engine and the Landauer's principle and report a direct measurement of the entropy change along SB transitions in a Brownian particle. The SB is induced by a bistable potential created with two optical traps. The experiment confirms the theoretical results based on fluctuation theorems, allows us to reproduce the Szilard engine extracting energy from a single thermal bath, and shows that the signature of a SB in the energetics is measurable, providing new methods to detect the coexistence of metastable states in macromolecules.

The energetics associated to SB transitions and, in general, to the manipulation of metastable states has special relevance in a number of interesting physical situations, some of them realized experimentally in the last years. The original Szilard engine, a celebrated and refined version of the original Maxwell demon, can extract work from a single thermal bath using the information created in a SB [11, 16, 19]. Landauer's principle accounts for the minimum dissipation associated to the erasure of information, which is a manipulation of the two metastable states making up a single bit memory [11] [cf. Eq. (1.16)]. The erasure can be interpreted as the restoration of a broken symmetry and has been reproduced with a Brownian particle in a double well potential created by optical tweezers [2]. In molecules, metastable states correspond to different molecular conformations as well as to kinetic states of special relevance in biophysics [1]. The energetics of processes involving metastable states has become a tool to measure conformational free energies in those contexts. An extended version of the nonequilibrium work theorem relates the probability distribution of the work in a process connecting two metastable states with their

conformational free energies [9, 12]. Maragakis et al. [12] applied this result to numerical simulations of switches between two different conformations of alanine dipeptide and Alemany et al. [1] have used it to obtain conformational free energies in stretching experiments of DNA structures that posses intermediate and/or misfolded kinetic states.

In this chapter we report on an experimental realization of a SB consisting of a continuous transition from a single well to a double well potential affecting a Brownian particle. We reproduce the transition moving apart two optical traps and measure the heat dissipated by the particle to the surrounding water that acts as a thermal reservoir. An electrostatic field allows us to tune the bias towards one or the other trap and explore the relationship between the energetics of the SB and the probability to adopt one of the instances. We finally build a Szilard engine as a SB followed by the restoring of the symmetry. This process completes a cycle that extracts energy from the thermal bath if the electrostatic field along the process is properly chosen.

This chapter is organized as follows: In Sect. 6.1 we discuss the thermodynamic consequences of a symmetry breaking transition and its relationship with the equations relating dissipation and irreversibility at the microscopic scale. We derive a universal formula that expresses the total entropy production in a symmetry breaking (and symmetry restoration) in terms of the probability of a microscopic system to choose a particular instance i in the symmetry breaking. In Sect. 6.3 we report the experimental validation of our theoretical results obtained using an experimental setup formed by a Brownian particle trapped with optical tweezers, which is described in Sect. 6.2. In Sect. 6.4 we show how we can use our setup to design a Szilard engine and propose a new experimental design of an optimal Szilard engine. We also report in Sect. 6.5 results from simulations where the force exerted by the optical traps on the Brownian particle can be measured directly. The average work dissipated in a symmetry breaking is found to be related to the probability of choice according to our formulas. The discussion of the results is done in Sect. 6.6.

6.1 Symmetry Breaking and Symmetry Restoration

When a symmetry is broken, a system “makes a choice” among a set of instances $i = 1, \dots, m$. For a classical infinite system, a symmetry breaking (SB) consists of a sudden change in the set of available states: the whole phase space Γ is partitioned into non overlapping regions Γ_i , corresponding to the different instances $i = 1, \dots, m$. The partition occurs when certain control parameter λ crosses a critical value λ_c above which the system can no longer move spontaneously from one region to another and gets confined within Γ_i with a probability p_i , $\sum_i p_i = 1$. The notion of SB can be extended to finite systems with metastable states. The confinement is not strict in this case: the system can jump form a region Γ_i to another Γ_j . However, if the average residence time in each region is much larger than the time scale of the process under consideration one can talk about an effective SB. In this case, the SB transition is

not localized in a single value of the control parameter λ but it is rather a continuous transition where metastable states develop.

Consider a system with Hamiltonian $H(z; \lambda)$ where z is the microstate of the system ($z \in \Gamma$), and λ a control parameter, and an isothermal process at temperature T involving a SB, where the parameter changes in time as λ_t with $t \in [0, \tau]$. The average work required to complete the process, when the system adopts instance i , is break bounded by

$$\langle W \rangle_i^{(\text{SB})} - \Delta F_i \geq kT \ln p_i, \quad (6.1)$$

where k is the Boltzmann constant and $\Delta F_i = F_{\tau,i} - F_0$ is the change in free energy. The initial free energy is defined as usual, $F_0 = -kT \ln Z(T, \lambda_0)$ where $Z(T, \lambda) = \int_{\Gamma} dz e^{-\beta H(z; \lambda)}$ is the partition function of the system. On the other hand, the final free energy $F_{\tau,i} = -kT \ln Z_i(T, \lambda_\tau)$ is a conformational free energy defined in terms of the partition function restricted to the region Γ_i , i.e., $Z_i(T, \lambda_\tau) = \int_{\Gamma_i} dz e^{-\beta H(z; \lambda_\tau)}$. The bound in Eq. (6.1) is met with equality if the process is quasistatic. Recalling the relationship between the free energy, F , the internal energy U , and the entropy S of a system, $F = U - TS$, and the first law of thermodynamics $\Delta U = W + Q$, where Q is the heat or energy transfer from the thermal reservoir to the system, we derive a bound for the entropy production

$$\langle S_{\text{prod}} \rangle_i^{(\text{SB})} = \Delta S_i - \frac{\langle Q \rangle_i^{(\text{SB})}}{T} \geq kT \ln p_i. \quad (6.2)$$

We notice that the origin of the term $k \ln p_i$ in Eqs. (6.1) and (6.2) can be easily understood. An ideal SB comprises a contraction of the set of available states from Γ to Γ_i without the need of any extra work. This amounts to an increase of free energy $-kT \ln(Z_i/Z)$ which is not compensated by work and heat dissipation. Assuming an instantaneous SB, $p_i = Z_i/Z$, yielding the extra term $kT \ln p_i$ in Eqs. (6.1) and (6.2). In Fig. 6.1 we sketch a symmetry breaking transition where a microscopic system has to choose between three different non-overlapping subsets of phase space and we sketch an intuitive but non rigorous proof of the formula (6.1) in the quasistatic limit relating the dissipative work on the SB with the probability of choice.

Equation (6.1) can be derived using the KPB theorem introduced in Sect. 2.2.3. Let us consider a process where a microscopic system breaks a symmetry and ends in a particular region of the phase space Γ_i as in the example shown in Fig. 6.1. The average work done along this process when the system chooses the instance i is [10]

$$\langle W \rangle_i - \Delta F \geq kT \ln \frac{p_i}{\tilde{p}_i}, \quad (6.3)$$

where $\Delta F = F_\tau - F_0$ is an *equilibrium* free energy difference calculated using the partition function in the whole phase space in the end of the process, $F_\tau = -kT \ln Z(T, \lambda_\tau)$, where $Z(T, \lambda_\tau) = \int_{\Gamma} dz e^{-\beta H(z; \lambda_\tau)}$ is the partition function of the system at the end of the process. In the backward process, the system starts in equilibrium and therefore $\tilde{p}_i = Z_i/Z$, which yields,

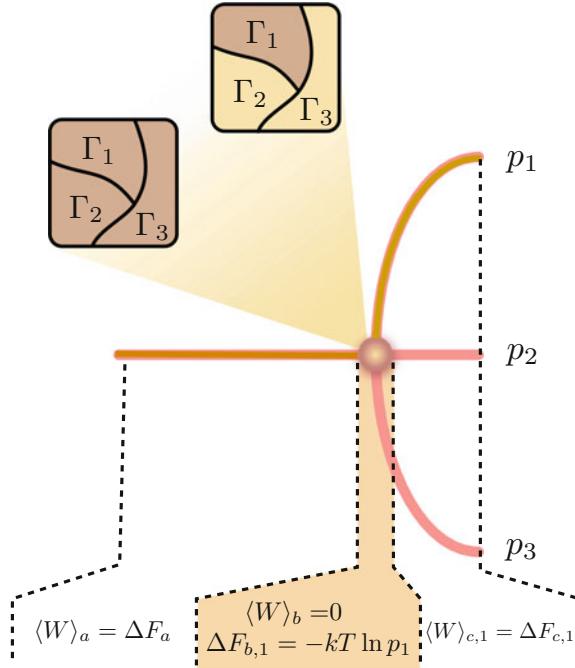


Fig. 6.1 Sketch of a quasistatic symmetry breaking. A microscopic system adopts instance 1 among three possibilities $i = 1, 2, 3$. Before and after the symmetry breaking, the reversible work is given by the increment of free energy: $\langle W \rangle = \Delta F$. At the symmetry breaking, a reduction of the phase space volume from Γ to Γ_1 occurs without the need of any external work: $\langle W \rangle_b = 0$. This reduction of phase space volume induces an increase of free energy $\Delta F_{b,1} = -kT \ln(Z_1/Z) = -kT \ln p_1$, where Z and Z_1 are partition functions calculated over Γ and Γ_1 , respectively, and $p_1 = Z_1/Z$ is the probability that the system ends within region Γ_1 when the symmetry is broken. The total reversible work is $\langle W \rangle_1^{(SB)} = \langle W \rangle_a + \langle W \rangle_b + \langle W \rangle_{c,1} = kT \ln p_1 + \Delta F_1$, where $\Delta F_1 = \Delta F_a + \Delta F_{b,1} + \Delta F_{c,1} = F_{\text{fin},1} - F_{\text{init}}$ is the total free energy change

$$\langle W \rangle_i - (F_\tau - F_0) \geq kT \ln p_i - (F_\tau - F_{\tau,i}). \quad (6.4)$$

By defining $\Delta F_i = F_{\tau,i} - F_0$ as the difference between the *nonequilibrium* free energy when the system chooses i and the equilibrium initial free energy, we obtain Eq. (6.1),

$$\langle W \rangle_i - \Delta F_i \geq kT \ln p_i. \quad (6.5)$$

Equation (6.1) is therefore a particular case of the KPB theorem for a SB transition. However, unlike in KPB theorem, Eq. (6.1) does not depend explicitly on the equilibrium backward probabilities \tilde{p}_i but only on the (forward) probabilities to choose a particular instance in the SB, p_i .

We notice that the relation between the average work and probability of choice can be expressed as an equality using exact results from extended fluctuation theorems

(EFTs), which express symmetry relations of thermodynamic functions for processes where the system is restricted to a subset of the phase space in the beginning and/or end of the process [9, 12]. The following equation [see Eq. E.5 in Sect E.1]

$$\langle W \rangle_i - \Delta F_i = kT \ln p_i + kT D[\rho_i(W) || \tilde{\rho}_i(-W)], \quad (6.6)$$

shows that the correction to the formula when the symmetry breaking is not quasistatic is given by a KLD between the probability density of the work when the system chooses i , $\rho_i(W)$, and the distribution of (minus) the work for trajectories that start in $\tilde{\Gamma}_i$ in the backward process, $\tilde{\rho}_i(-W)$. Since the KLD is positive, the correction term is positive, yielding the bound in (6.1), $\langle W \rangle_i - \Delta F_i \geq kT \ln p_i$. In the quasistatic limit, the forward and backward work distributions coincide and $\langle W \rangle_i - \Delta F_i = kT \ln p_i$.

We remark that despite Eqs. (6.1) and (6.2) suggest that total entropy (or dissipation) are negative along a SB, they are not in contradiction with the Second Law of Thermodynamics, because the final state $\rho_i(x)$ is not in complete equilibrium and the final entropy cannot be considered as a true thermodynamic entropy. The conformational entropies S_i or free energies F_i , however, even though cannot be considered as true thermodynamic potentials (they are not state functions, for instance) are useful tools to analyze the energetics of processes involving SB transitions. An alternative interpretation of the decrease of entropy is that it is compensated by an increase of the meso- or macroscopic uncertainty, quantified by the Shannon entropy of the SB outcome, $H = -\sum_i p_i \ln p_i$. Notice that the average of Eq. (6.1) over p_i yields precisely $-kT H$. We notice that in a recent work [17], it has been shown that the energy cost of a memory erasure, which can be seen as the reverse process of a symmetry breaking, is precisely $\langle W \rangle_{\text{eras}} - \Delta F_{\text{eras}} \geq kT H$.

Similar inequalities hold for a process where a symmetry is restored. To assess the energetics of a symmetry restoration (SR) we have to consider the time reversal of the restoration, which is a SB. Let us call \tilde{p}_i the probability that the system adopts instance i in this SB resulting from the time reversal of the original process. Under time reversal, reversible work and free energy increment change sign. Therefore:

$$\langle W \rangle_i^{(\text{SR})} - \Delta F_i \geq -kT \ln \tilde{p}_i, \quad (6.7)$$

where $\Delta F_i = F_\tau - F_{0,i}$ is the free energy change of the symmetry restoration. Notice that now the initial free energy that depends on the instance i . For the entropy production:

$$\langle S_{\text{prod}} \rangle_i^{(\text{SR})} \geq -k \ln \tilde{p}_i. \quad (6.8)$$

Landauer's principle follows immediately from Eq. (6.7) applied to a one-bit memory consisting of a physical system with two stable states, 0 and 1, each one with the same free energy $F_0 = F_1$. The minimal cost of erasing a bit or, more precisely, to drive bit $i = 0$ or 1 to the state 0 (restore-to-zero operation) is $\langle W \rangle_i^{\text{erasure}} \geq$

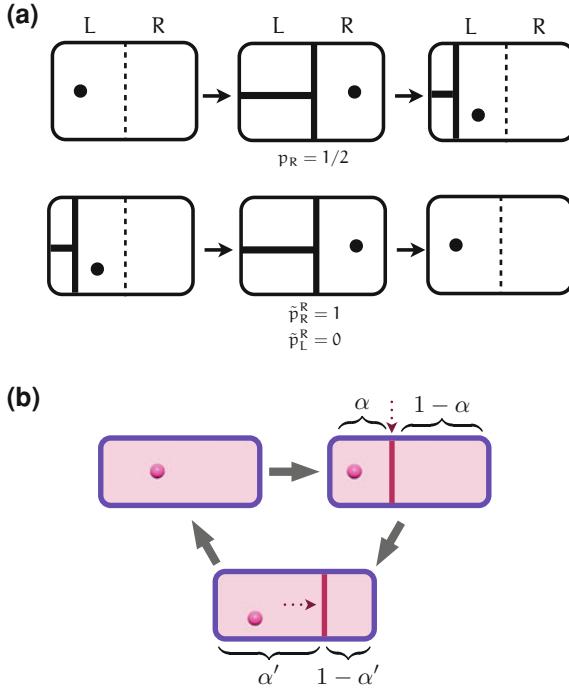


Fig. 6.2 Illustration of the meaning of \tilde{p} . **a** The Szilard engine. The *top figure* sketches the Szilard engine when the particle is in the *right half* of the box when the measurement is done, which happens with probability $p_R = 1/2$. In the corresponding backward process for the outcome $i = R$ in the forward process, the protocol $\tilde{\lambda}_t^R$ is sketched in the *bottom figure*. Under the protocol $\tilde{\lambda}_t^R$, the particle ends always in the *right half* in the time of the measurement in the backward process, $\tilde{p}_R^R = 1$. **b** A piston is introduced at a distance α from the *left wall* of a box containing a single Brownian particle; the piston is shifted to position α' and then removed. When the piston is inserted the system “makes a choice” between the *left*, L , or *right*, R , side of the box with probabilities $p_L = \alpha$ and $p_R = 1 - \alpha$, respectively. The symmetry is restored when the piston is removed at position α' . The process of restoring the symmetry (*piston removal*) reversed in time is also a symmetry breaking with probabilities $\tilde{p}_L = \alpha'$ and $\tilde{p}_R = 1 - \alpha'$

$-kT \ln \tilde{p}_i - \Delta F_i = -kT \ln \tilde{p}_i$ for $i = 0$ or 1 , since in both cases $\Delta F_i = F_i - F_0 = 0$. If the initial bit is unknown, the best we can do is $\tilde{p}_i = 1/2$ and $\langle W \rangle_i^{\text{erasure}} \geq kT \ln 2$.

The energetics of the Szilard engine [11, 17] can be as well easily reproduced from Eqs. (6.1) and (6.7). In Szilard’s setup, a system undergoes a SB and chooses between two instances $i = L$ or R corresponding to the left and right halves of the box, with probability p_L and p_R , respectively. Then we measure the instance that has been chosen and restore the broken symmetry driving the system back to the original state through some protocol λ_t^i . The time reversal of this protocol, $\tilde{\lambda}_t^i$, is a SB transition with possibly different probabilities \tilde{p}_j^i (see Fig. 6.2a). Notice that probabilities \tilde{p}_j^i are measured at the same (forward) time during the process. We

remark that the superindex i in \tilde{p}_j^i stands for the protocol followed after the symmetry is broken whereas the subindex j for the choice of the system in the reverse process, and therefore $\sum_j \tilde{p}_j^i = 1$.¹ The probabilities \tilde{p}_j^i depend on the protocol λ_t^i after the choice and therefore can be varied independently as shown in a simple example in Fig. 6.2b.

The work necessary to implement the SB is bounded by Eq. (6.1) and the work necessary to restore the symmetry is bounded by Eq. (6.7). Therefore, the total average work that we have to perform to run the whole cycle obeys

$$\begin{aligned} \langle W \rangle &= \sum_i p_i \left[\langle W \rangle_i^{(\text{SB})} + \langle W \rangle_i^{(\text{SR})} \right] \\ &\geq kT \sum_i p_i \ln \frac{p_i}{\tilde{p}_i^i}, \end{aligned} \quad (6.9)$$

and $\langle S_{\text{prod}} \rangle = \langle W \rangle / T$. Notice that the right hand side in (6.9) is not a KLD because \tilde{p}_i^i are not normalized in general, and $\sum_i \tilde{p}_i^i$ can be greater than 1. The right hand side can be even negative, which corresponds to a work extraction by the demon from the thermal bath. Two particular cases of demons are of particular interest:

- *Optimal demons* The protocol after the SB is such that the probabilities in the backward process are $\tilde{p}_i^i = 1$, which are optimal in the sense that the value of the average work done in the whole cycle is minimal [7]. By replacing $\tilde{p}_i^i = 1$ in (6.9), the average work is bounded by $\langle W \rangle \geq -kT H(p_i)$ where $H(p_i) = -\sum_i p_i \ln p_i$ is the Shannon entropy associated to the uncertainty of the state of the particle after the SB. The original Szilard engine is an optimal demon with $p_i = 1/2$ and $\tilde{p}_i^i = 1$, yielding $\langle W \rangle \geq -kT \ln 2$, i.e., the extraction of an energy $-kT \ln 2$ in a cycle.
- *Blind demons* If the demon does not use information from the measurement and performs always the same protocol, i.e., $\lambda_t^i = \lambda_t$, then $\tilde{p}_j^i = \tilde{p}_j$ normalized to unity $\sum_j \tilde{p}_j = 1$ yielding $\langle W \rangle \geq kT D(p_i || \tilde{p}_i) \geq 0$ which coincides with the KPB formula linking dissipation and irreversibility (KLD) measured using coarse grained information [cf. Eq. (2.22)].

A variety of intermediate cases can be implemented to build demons. The key ingredient to extract work from the thermal bath is to find a feedback protocol where p_i and \tilde{p}_i^i are such that the average work $\langle W \rangle$ in Eq. (6.9) is negative; for instance by choosing protocols where $\tilde{p}_i^i > p_i$, a condition that is met in the original Szilard engine. In the example in Fig. 6.2b one can obtain $\langle W \rangle_L < 0$ by choosing $\alpha' > \alpha$. A explicit experimental construction of the Szilard engine using optical tweezers can be found in Sect. 6.4.

¹ \tilde{p}_j^i can be seen as a conditional probability, since it measures the probability to choose the instance j in the backward process $\tilde{\lambda}_t^i$ implemented as the time reversal of λ_t^i . Therefore it is the probability to choose j in the backward process conditioned to the fact that the choice in the forward process was i , $\tilde{p}_j^i = \tilde{p}(j|i)$.

6.2 Experimental Setup

We now discuss the experimental setup that we designed to test the universal relations concerning the energetics of the SB and SR. The experiment was done by Ignacio A. Martínez and Dmitry Petrov in *ICFO*, Institute of Photonic Sciences.²

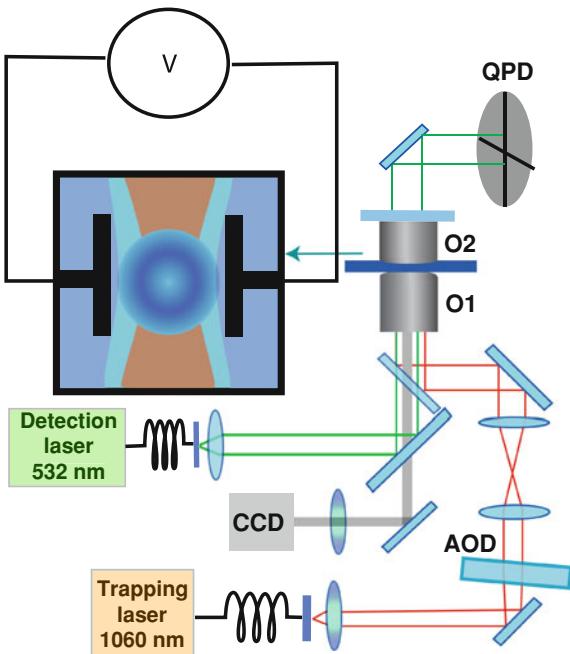
Figure 7.1 shows a schematic of the optical apparatus. An acousto-optical deflector (AOD) (ISOMET LS55 NIR) steers a 1,060 nm optical beam from a laser coupled into a single-mode fiber (ManLight, ML10-CW-P-OEM/TKS-OTS, maximal power 3 W). The AOD modulation voltage is obtained from an arbitrary waveform generator (TaborElectronics WW1071) controlled by a LabView program. The beam deflected by the AOD is expanded and inserted through an oil-immersed objective O1 (Nikon, CFI PL FL 100X NA 1.30) into a custom made fluid chamber. An additional 532 nm optical beam from a laser coupled to a single-mode fiber (OZOptics) is collimated by a ($\times 10$, $NA = 0.10$) microscope objective and passes through the trapping objective. The forward scattered detection beam is collected by a ($\times 10$, $NA = 0.10$) microscope objective O2, and its back focal-plane field distribution is analyzed by a quadrant position detector (QPD) (New Focus 2911) at an acquisition rate of 20 kHz. A 532 nm band pass filter in front of the QPD blocks beams with wavelengths different from the detection beam wavelength. The AOD permits the control of the position of the beam focus, and we used in the time-sharing regime to create the protocols described in the text. We used it also for the calibration procedures and for mapping the force distribution in the optical trap. A fluid chamber with microspheres was placed on a piezoelectric-controlled calibrated stage (PiezosystemJena, Tritor 102) allowing 3D translation (Fig. 6.3).

Polystyrene micro spheres (G. Kisker-Products for Biotechnology, <http://www.kisker-biotech.com/>), polystyrene microparticles PPs-1.0 with a diameter of $1.00 \pm 0.05 \mu\text{m}$) were diluted in distilled de-ionized water to a final concentration of a few spheres per ml. The spheres were inserted into a custom made electrophoretic fluid chamber with two electrodes connected to a computer controlled electric generator and an amplifier. The chamber was previously described in [3, 15].

The individual traps (the fixed trap *F*, and the other one *M* that is moved along the *x*-axis) are not generated simultaneously but with a single optical beam that is rapidly switched between a number of optical focuses by changing the frequency of acoustic waves propagating in the AOD. Multiple optical traps are possible as long as the focused beam is returned to the same location faster than the time it takes for the probe to diffuse away from that location, typically on the order of tens of milliseconds. The time sharing regime of AOD has the greatest additional flexibility because it allows independent control of both the positions of two optical traps (changing the frequency of acoustic waves) and their stiffnesses. In fact, since the probe observes an average intensity, the trapping force at each location may be changed by a proper choice of the timing ratio between two frequency components at

² Full address: ICFO—Institut de Ciències Fotòniques, Optical Tweezers Group, Mediterranean Technology Park, Av. Carl Friedrich Gauss, num. 3, 08860 Castelldefels (Barcelona), Spain.

Fig. 6.3 Schematic of the experimental setup used to test the energetics of the symmetry breaking and symmetry restoration



the AOD input. Time-sharing techniques using AODs have been described previously (i.e., in [20]).

The alternation of trap positions is controlled by timing signals generated by a modulation generator adjusted to give a low frequency square wave with a controllable duty ratio. The analogue voltage is transferred to the modulation input of a RF driver. The drivers' RF output is then transferred to the AOD. The relative optical intensity in each trap depends on the duty ratio of the modulation signal, and the distance between the traps is defined by the amplitude of the square wave. In order to provide an additional flexibility to the optical trapping system a DC modulation signal was added to the square modulation wave. It permits to shift the positions of both traps simultaneously. We used this signal by a calibration of the traps as it is described below. The time-sharing protocol permitted to control the velocity of the M trap and the position of its center following prescribed experimental protocols.

The system needs the calibration of the position detection system. At a given position of the quadrant position detector (QPD), the displacements of the sphere produce a unique and linear QPD response only within a restricted region of the probe displacement, usually several hundred nanometers. As we expect an increase of the amplitude of the displacements on application of an additional noise, the linearity range of the detection system and the force map of the optical trap should be characterized.

First, we calibrated the QPD measuring the ratio between the QPD output signals (in volts) and the real displacements of the probe (in nm) by using the natural thermal

fluctuations of the probe as described in [20]. The trap was calibrated in the absence of the additional force. Using a relatively strong trap where the Brownian fluctuations did not surpass several tens of nanometers, we first obtained the PSD of the position of the particle. The experimental PSD is then fitted to a Lorentzian function [20]. From the fit, we obtain the stiffness κ and the calibration coefficient S_{QPD} (nm/V) relating the QPD output signal with the absolute displacement of the sphere.

The spheres were trapped above the bottom surface of the fluid chamber at the distance about 20 μm controlled by the piezoelectric stage. The volume value of the viscous friction coefficient γ was corrected due to the proximity of the chamber surface using the Faxen's law as described in [5].

We then calibrated the acousto-optical deflector (AOD), measuring the ratio between the amplitude of the AOD input modulation signal V_{AOD} (in volts) and the corresponding sphere displacement (in nm) using the previously measured QPD calibration factor S_{QPD} . We moved the trapped sphere in the range of several tens of nanometers, changing the AOD modulation signal, and then measured the output signal of the QPD. This provided us with the calibration factor S_{AOD} (nm/ V_{AOD}) for the small range of the sphere displacement. We then confirmed with an image analysis that this factor holds for the AOD modulation amplitudes corresponding to the sphere displacements of $\pm 3 \mu\text{m}$.

Figure 6.4a shows the dependence of the QPD output signal on the shift of the sphere achieved by changing the AOD modulation voltage. The linear relationship between the displacements of the optically trapped sphere and the output signal of the position detection system can be seen within the range $\pm 1,000$ nm. The linear range of the position detection can be extended further using a method suggested recently in [13].

For completion, we measured the map of the forces that the optical trap exerted on the sphere. These measurements define a range where the linear relationship between the optical force and the sphere displacement is valid. The force map was characterized by the method proposed in [8]. In that method, a strong, stationary, and previously calibrated trap holds a sphere of a given diameter. The optical trapping potential of this trap is strong enough that external forces can not move the sphere out of the range of the harmonic approximation. The intensity of the second trap, whose optical trapping potential for the given sphere has to be analyzed, is adjusted to be low. Its optical force acts as a perturbation exerted on the probe. When the second trap is scanning near the strong trap, it produces a variable force measured through the displacement of the probe in the strong trap.

To experimentally realize this technique, we used the AOD in a time-sharing regime [14]. Figure 6.4b illustrates the force map of the optical trap used in the experiments. The linear range of the restoring optical force extends up to ± 500 nm.

The calibration curves permitted us to calculate the position of the center of the movable trap and the position of the trapped sphere in the absolute values as presented in Fig. 6.5.

To characterize the stability of the electric charges on a sphere, we applied a 1 Hz sinusoidal voltage to the electrodes of the electrophoretic chamber. The amplitude of

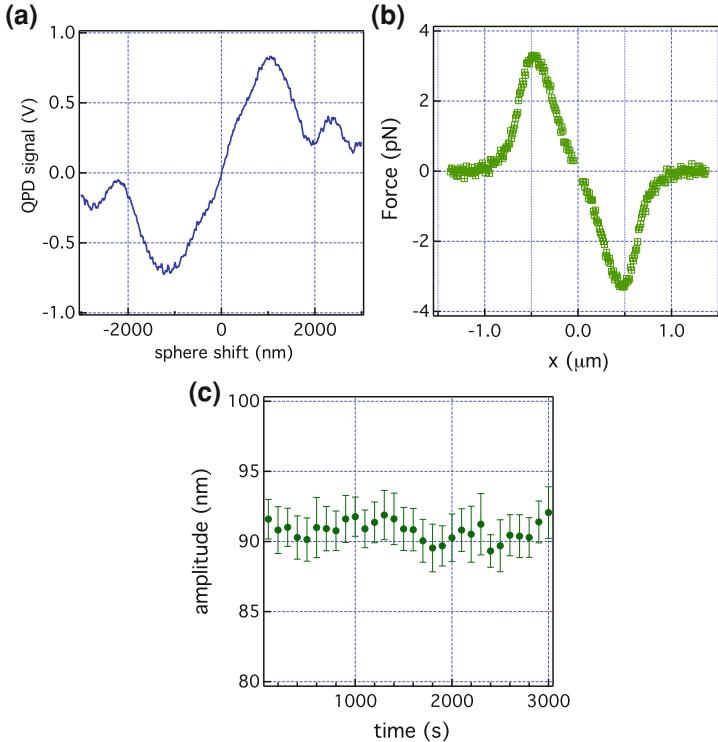


Fig. 6.4 Calibration of our experimental setup. **a** Output signal of the QPD as a function of the position of the micro sphere. The output signal of the QPD depends linearly on the sphere position when the sphere displacements is within the range $\pm 1,000$ nm. The minor deviations from the linear dependence in the linear part of the QPD response are due to the non uniformity in the detection beam intensity as well as because of the fluctuations of the position of the sphere. **b** The force exerted on the $1\mu\text{m}$ optically-trapped sphere versus the radial distance from the trapping beam focus. The incident intensity of the trapping beam at the input pupil of the trapping objective is 5 mW. We confirmed that the force map is not affected by further increase of the strong trap stiffness (*keeping the same intensity of the weak trap*). **c** The amplitude of the sphere displacement in an AC electric voltage 30 V at 1Hz during 3,000 s

the displacement of the optically trapped sphere measured with the position detection system did not change more than 6 % over at least 3,000 s (see Fig. 6.4c).

6.3 Experimental Test

Inequalities (6.1) and (6.2) are universal, i.e., they do not depend on the details of the SB or even on the physical nature of the system under consideration. We have tested both inequalities experimentally using a Brownian particle immersed in water trapped with a two optical tweezers.

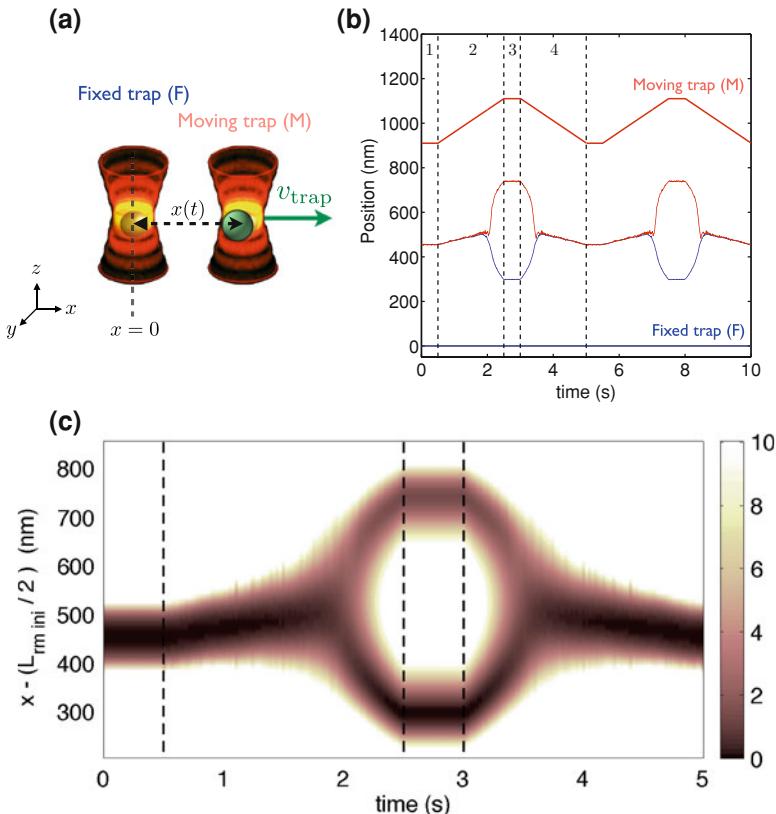


Fig. 6.5 Experimental protocol of symmetry breaking and symmetry restoration. **a** scheme of the experiment a trap is held fixed with its center in $x = 0$ and a second trap is moved at constant velocity v_{trap} along the x -axis. The position of the bead $x(t)$ is monitored with sub-nanometer precision. **b** Position of centers of the two optical tweezers as a function of time during the experimental protocol. The center of the fixed trap F (blue dotted line) stays in $x = 0$ and the center of the moving trap M moves as indicated in red dotted line. The symmetry breaking experiment can be implemented with the cycle formed by the subprocesses $1 \rightarrow 2 \rightarrow 3$. The average position of the bead after implementing the protocol cyclically during $t = 2,400$ is indicated in blue when the bead chooses the F trap and in red when it chooses the M trap. **c** Empirical potential $U(x, t) = -\ln \rho(x, t)$ measured with $\rho(x, t)$ obtained from the statistics of trajectories of the bead during $t = 2,400$ s

In our experiment, we trap a polystyrene spherical bead ($1 \mu\text{m}$ diameter) in water with two optical tweezers created by a single laser beam in a time-sharing regime (see Sect. 6.2 and Fig. 6.5a). One of the traps, labelled F , is held fixed and the other one, labelled M , is moved along the x -axis following the four step protocol depicted in Fig. 6.5b:

1. The two traps are initially at rest during a period of time $\tau_1 = 0.5$ s with their centers separated by a distance $L_{\text{ini}} = 910$ nm.
2. Trap M is moved along the x -axis at constant velocity v_{trap} during a time τ_2 .

3. The two traps are again kept fixed for a time $\tau_3 = 0.5$ s with their centers separated by $L_{\text{fin}} = 1,110$ nm.
4. Trap M is moved back from L_{fin} to its initial position L_{ini} with velocity $-v_{\text{trap}}$ in a time τ_2 .

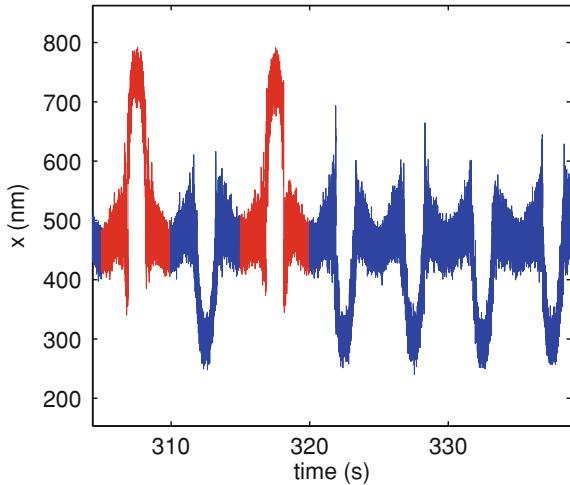
The total duration of the cycle is $\tau = 2\tau_2 + 1$ s. By repeating this protocol cyclically, we can study both the SB (steps 1 – 2 – 3) and the symmetry restoration (steps 3 – 4 – 1). Moreover, a uniform electrostatic field created by electrodes in the chamber allows us to bias the motion of the bead towards the M or F trap.

The protocol can be considered quasistatic when the velocity of the trap is around 100 nm/s or below. At this velocity, heat dissipation due to drag is of order $\gamma v_{\text{trap}}^2 \approx 10^{-22}$ J/s ≈ 0.02 kT/s, where $\gamma = 6\pi R\eta$ is the drag coefficient, $R = 0.5$ μm being the radius of the bead and $\eta = 8.9 \times 10^{-4}$ Pa · s the dynamic viscosity of water at 25 °C. We have implemented two quasistatic protocols with $v_{\text{trap}} = 100$ nm/s ($\tau_2 = 2$ s) and $v_{\text{trap}} = 36.36$ nm/s ($\tau_2 = 5.5$ s). In our experiments, the whole cycle is repeated under the same conditions (bead and electrostatic field) for at least 2,400 s, i.e., 480 times for the fast and 200 times for the slow protocol. At the end of the SB protocol (steps 1 – 2 – 3), Kramers transitions are not observed and one can unambiguously distinguish two different final meso-states for the bead depending on the final state of the bead, M or F (see Fig. 6.5b, c).

We are able to track the position of the bead $x(t)$ with sub-nanometer precision with an acquisition frequency $f_{\text{acq}} = 1$ kHz. In Fig. 6.5b, we plot the average position of the bead of both F and M realizations as a function of time, $\langle x(t) \rangle_{F,M}$, which is calculated as an ensemble average over F and M trajectories, respectively. The trajectories of the bead showed a drift whose characteristic frequency was of the order of Hz ($\ll f_c < f_{\text{acq}}$, f_c being the corner frequency of the bead in the optical trap, which is of the order of 100 Hz). The drift was corrected in all the experimental series by imposing that the average position in every cycle, calculated using the data from $t = 0$ to $\tau_{\text{drift}} = 0.25$ s takes a constant value throughout the experiment. A sample of seven cycles corresponding to the protocol shown in Fig. 1.1 of the main text is shown in Fig. 6.6. In each repetition of the cycle that contains the SB and SR, the trajectory is labeled as $F(M)$ if the position of the bead is below (above) a threshold, which is set to the average of the position in the first $\tau_{\text{drift}} = 0.25$ s. The probability of the bead to choose the F and M trap in the SB is obtained by counting the number of times the trajectory is labeled as F or M according to this criterion.

To measure the thermodynamic quantities associated to the stochastic trajectories of the bead, it is necessary to measure the energy landscape generated by the two traps in the experiment, $U(x, t)$, which depends on both time and the position of the bead. Since our experimental setup does not allow a direct measurement of the force exerted on the bead, we infer the potential energy from position histograms. In the quasistatic regime, $U(x, t)$ can be estimated, up to an additive function of time, from the equilibrium probability density $\rho(x, t)$. Before the SB, $U(x, t) = -kT \ln \rho(x, t) + kT \ln Z_t$, where Z_t is the partition function at time t . However, when the barrier between the two potential wells is high and Kramers transitions are rare or non-existent, it is not possible to obtain the relative depth of each well

Fig. 6.6 Trajectories of the position of the bead as a function of time obtained using the experimental protocol corresponding to Fig. 6.5. Cycles where the particle ends in the F trap after the SB are depicted in blue whereas those where the particle ends in the M trap are plotted in red



from the probability density $\rho(x, t)$, which only provides local information of the potential. To overcome this problem, we estimate $\rho(x, t)$ from the probability of the trajectories of the bead in the experimental protocol to pass through x at time t .

The empirical estimation of $\rho(x, t)$ is done using data points in two time windows $[t - S/2, t + S/2]$ and $[\tau - t - S/2, \tau - t + S/2]$, of width $S = 50$ ms and centered around t and $\tau - t$, respectively, where the two traps are exactly in the same position in the stages 2 and 4 of the process. For times within steps 1 and 3, since the optical potential does not change, we have used data from the whole interval 1 and 3 respectively. The probability density $\rho(x, t)$ is the normalized histogram of those data with a bin size $\Delta x = 10$ nm. From this estimation of $\rho(x, t)$, we define a "trajectory-potential" obtained from the statistics of the trajectories of the bead in the experimental protocol and given by $U(x, t)/kT = -\ln \rho(x, t)$.

To measure the heat transferred along a stochastic trajectory, it is necessary to know the value of $U(x, t)$ at any value of the position x and any time t during the process, as shown in Eq. (E.23) (see Sect E.3). To do this, we fit the empirical estimation of the potential to a quartic polynomial $U(x, t) = a_0(t) + a_1(t)x + a_2(t)x^2 + a_3(t)x^3 + a_4(t)x^4$, where $a_i(t)$ are time dependent parameters. At every time t from 0 to τ , with time step $\Delta t = 1$ ms, the empirical values of $U(x, t)$ are fitted to a quartic potential using a nonlinear least squares weighed fit. The data is weighed as $w(x, t) = e^{-U(x,t)/kT}$, i.e., the data that are in the bottom of the wells are favored by the fit. All the data points that exceed in more than $V_{\text{top}} = 10$ kT the global minimum of the potential are not considered in the fit. In Fig. 6.7 we show snapshots of the empirical values and fitted curves of the energy landscape generated by the two optical traps close to the SB transition.

The entropy production associated to a single stochastic trajectory in the SB has two contributions [cf. Eq. (6.2)]

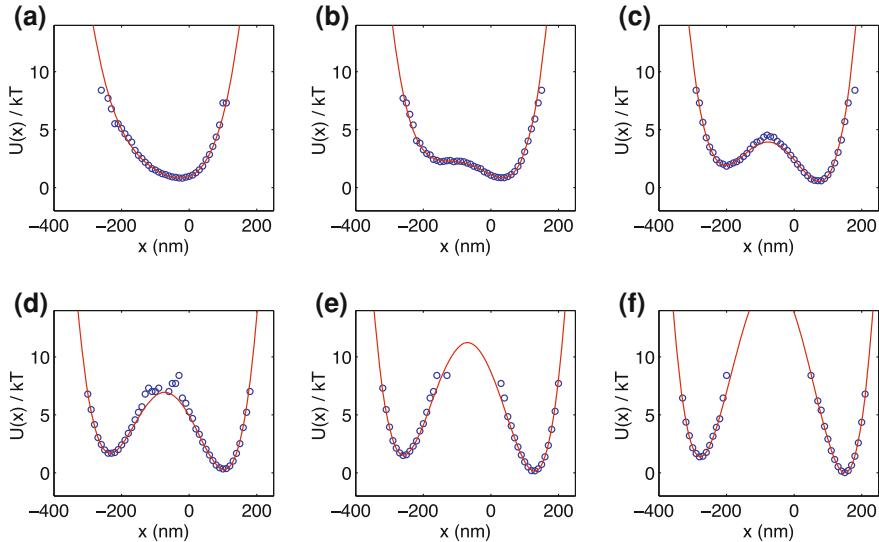


Fig. 6.7 Empirical potential energy (blue circles) and fit to a quartic potential (red curve) at different times t during the symmetry breaking experiment. At different stages of the protocol shown in Fig. 6.5: **a** $t = 1.7$ s. **b** $t = 1.8$ s. **c** $t = 1.9$ s. **d** $t = 2$ s. **e** $t = 2.1$ s. **f** $t = 2.2$ s. The data is centered in such a way that the moving trap moves from left to right. The global minimum of the empirical potential is set to zero

$$S_{\text{prod}}^{(\text{SB})} = \Delta S - \frac{Q^{(\text{SB})}}{T}. \quad (6.10)$$

First, the system entropy change can be measured as $\Delta S = S_{\text{fin}} - S_{\text{ini}}$, since it is a state function. The initial system entropy is

$$S_{\text{ini}} = -k \int_{\Gamma} dx \rho_{\text{ini}}(x, t) \ln \rho_{\text{ini}}(x, t), \quad (6.11)$$

where the integration is done over the whole phase space. The initial phase space density $\rho_{\text{ini}}(x, t)$ is obtained using the data from the whole interval from 0 to $\tau_1 = 0.5$ s. The system entropy at the end of the process depends on which trap the particle ends,

$$S_{\text{fin}, i} = -k \int_{\Gamma_i} dx \rho_{\text{fin}, i}(x, t) \ln \rho_{\text{fin}, i}(x, t), \quad (6.12)$$

where Γ_i is the phase space accessible to the particle when it ends in the trap i after the SB and $\rho_{\text{fin}, i}(x, t)$ is the final phase space density measured only with the statistics of trajectories that end in the i trap in the SB. The final density is also obtained using the data points of the last 0.5 s of the trajectories, from $\tau - \tau_1$ to τ . We calculate the

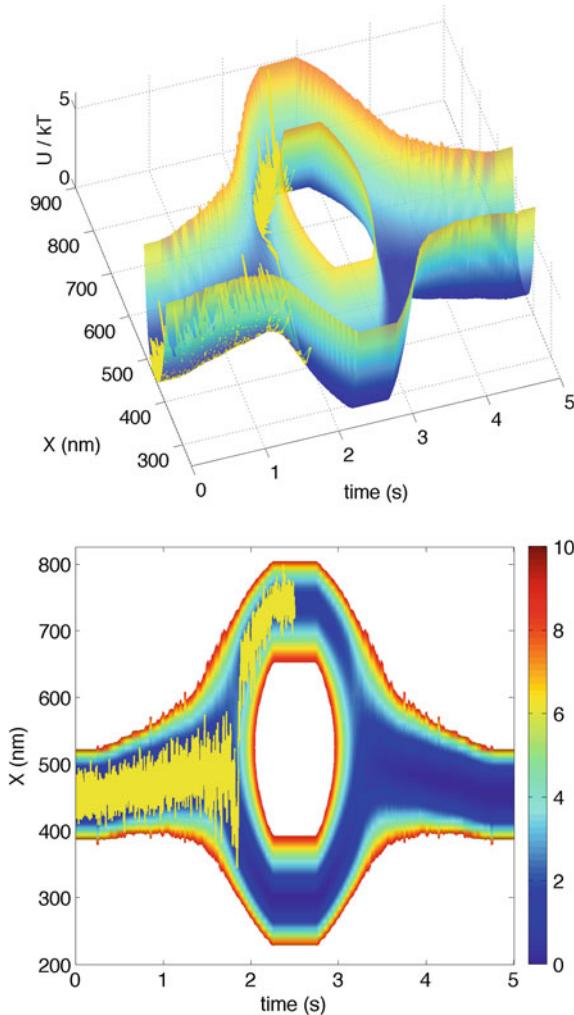
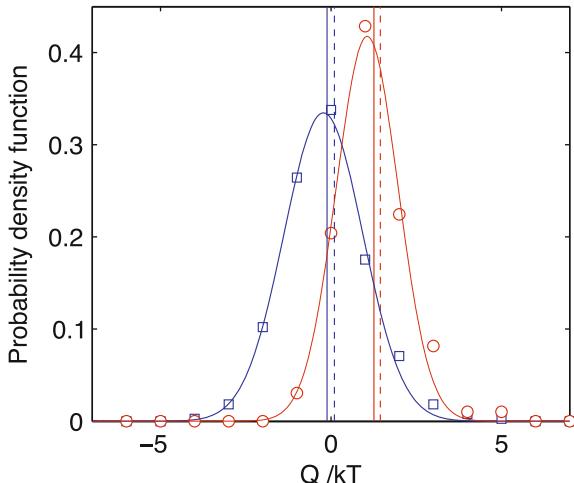


Fig. 6.8 Fitted potential $U(x, t)$ as a function of time (x -axis, in seconds) and position (y -axis in nanometers) and a sample trajectory in the SB in which the particle ends in the M trap. *Top* Surface 3d plot with the trajectory in the SB shown in yellow and the value of the potential $U(x, t)$ plotted in the z -axis. *Bottom* 2d plot with the value of the potential indicated in the color bar. White areas correspond to values of the potential above $V_{\text{top}} = 10 kT$

heat transferred to the Brownian particle along an individual trajectory in the SB, $x_{\tau/2} = \{x(t)\}_{t=0}^{\tau/2}$, using the formalism described in Sect E.3 equivalent to Sekimoto's stochastic thermodynamics described in Sect. 1.4.2,

$$Q[x_{\tau/2}]^{(\text{SB})} = \sum_{i=0}^{(\tau/2)/\Delta t} \delta Q(i \Delta t), \quad (6.13)$$

Fig. 6.9 Empirical probability density function of the heat (in units of kT) for trajectories that end in the F (blue open squares) and M (red open circles) trap. Solid curves are Gaussian fits, with $R^2 = 0.995$ for the F case and $R^2 = 0.990$ for the M case. Vertical lines are the empirical averages of the two ensembles and dashed vertical lines correspond to $\Delta S_i/k - \ln p_i$



where the heat exchanged in $[t, t + \Delta t]$, $\delta Q(i\Delta t)$, is measured using (E.23). In Fig. 6.8 we show an example of a stochastic trajectory of the particle moving in the trajectory potential. The averages of heat, system entropy and entropy production to F and M ensembles are obtained as the mean value of these quantities restricted to the case where the particle chooses the F or M trap. In Fig. 6.9 we show the probability distribution of the microscopic heat for both F and M trajectories, which are Gaussian in good approximation. The mean of the heat satisfies

$$\langle S_{\text{prod}} \rangle_i^{(\text{SB})} = \Delta S_i - \frac{\langle Q \rangle_i^{(\text{SB})}}{T} \geq k \ln p_i, \quad (6.14)$$

and therefore

$$\beta \langle Q \rangle_i^{(\text{SB})} \leq \frac{\Delta S_i}{k} - \ln p_i. \quad (6.15)$$

Measuring the system entropy change and the heat transfer, we can obtain the average entropy production up to time $t \leq \tau$ for individual trajectories. We first measure the entropy production along the SB process, i.e., from $t = 0$ to $t = \tau/2$. In Fig. 6.10 we show the average entropy production in the SB process as a function of $\ln p_i$. We include experimental data from both F and M averages in both fast ($\tau_2 = 2$ s) and slow ($\tau_2 = 5.5$ s) protocols. The predicted dependence of the entropy production with respect to the probability of choice is verified in the experiment and the negative entropy production associated to SB transitions is observed as well. Our experimental series include cases where p_i ranges from 5 to 95 %. The average entropy production is closer to $\ln p_i$ when the probability is higher and there is more statistics available to measure entropy production accurately.

We also study the SR by using the data from trajectories starting in $t = \tau/2$ and ending in $t = \tau$ in our experimental protocol. We measure the entropy production

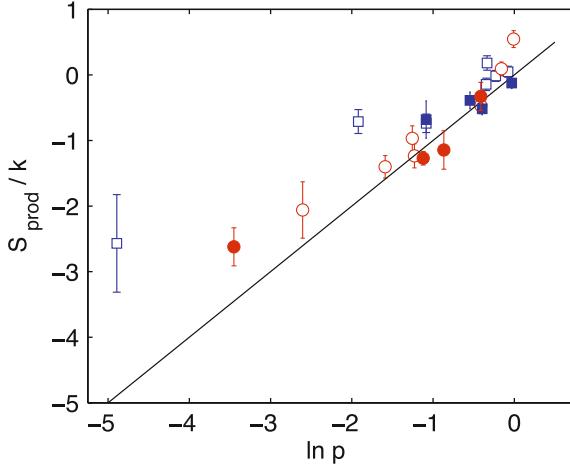


Fig. 6.10 Entropy production in the symmetry breaking as a function of the probability p_i of adopting instance i . Open symbols are obtained using the fast protocol ($\tau_2 = 2$ s) and filled symbols using the slow protocol ($\tau_2 = 5.5$ s). Blue squares represent the ensemble averages of S_{prod} for trajectories in which the bead ends in the fixed trap, whereas red circles are the corresponding values for realizations in which the bead ends in the moving trap. The black line corresponds to the theoretical prediction, $\ln p$. Error bars are only statistical and obtained using a statistical significance of 90 %

for the SR process implemented in the same time series as those studied in the SB in Fig. 6.10. In Fig. 6.11, we show the value of the entropy production as a function of $\ln \tilde{p}_i$ ($= \ln p_i$ in this case) for $i = F, M$ in both slow and fast experimental protocols. Our experimental results agree within experimental errors with the theoretical prediction given by Eq. (6.8).

In addition to these measurements, we also compute the value of the entropy production from $t = 0$ to $t \leq \tau/2$ (i.e., in the SB process) as a function of time. We define the accumulated averages of entropy production $\langle S_{\text{prod}} \rangle_F^{(\text{SB})}(t)$ and $\langle S_{\text{prod}} \rangle_M^{(\text{SB})}$ as

$$\langle S_{\text{prod}}(t) \rangle_i^{(\text{SB})} = \int_0^t dt' \langle dS_{\text{prod}}(t') \rangle_i, \quad (6.16)$$

where

$$\langle dS_{\text{prod}}(t) \rangle_i^{(\text{SB})} = S_i(t + \Delta t) - S_i(t) - \frac{\langle \delta Q(t) \rangle_i^{(\text{SB})}}{T}, \quad (6.17)$$

is the average entropy production in the interval $[t, t + \Delta t]$ when the system chooses trap i . We measure system entropy along the process, $S_i(t)$, using Seifert's definition of nonequilibrium system entropy [18], that is, replacing the empirical distribution $\rho(x, t)$ at any time t during the process in Eq. (6.11) or (6.12). In Fig. 6.12 we plot the

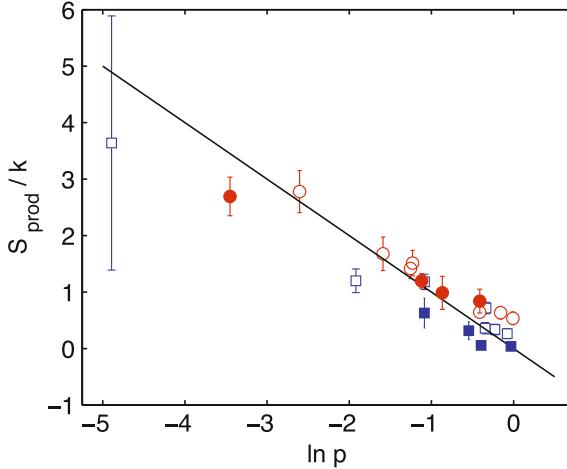


Fig. 6.11 Entropy production in the symmetry restoration as a function of the probability \tilde{p}_i of adopting instance i in the symmetry breaking experiment. Open symbols are obtained using the fast protocol ($\tau_2 = 2$ s) and filled symbols using the slow protocol ($\tau_2 = 5.5$ s). *Blue squares* represent the ensemble averages of S_{prod} for trajectories in which the bead ends in the fixed trap, whereas *red circles* are the corresponding values for realizations in which the bead ends in the moving trap. The *black line* corresponds to the theoretical prediction, $-\ln p$. Error bars are only statistical and obtained using a statistical significance of 90 %

difference $\langle S_{\text{prod}} \rangle_F^{(\text{SB})}(t) - \langle S_{\text{prod}} \rangle_M^{(\text{SB})}(t)$ as a function of time t along the SB process. The majority of the difference between the ensemble averages of entropy production appears in the region where the energy barrier appears in the symmetry breaking (cf. Fig. 6.7), which is the effective region where the symmetry is broken. The value of the difference at the end of the SB increases with p_F which is expected since the difference between ensemble averages is of the order of $\ln p_F/p_M = \ln p_F/(1-p_F)$. As a future application, the curves in Fig. 6.12 could be used to test the recently introduced *Differential Landauer principle* in [6], where the entropy production rate in a physical memory recording and erasure is studied and analytical expressions of $\dot{S}_{\text{prod}}(t)$ in terms of mutual information creation or destruction are obtained.

We notice that we have focused on testing experimentally Eqs. (6.2) and (6.8) rather than their corresponding equations expressed in terms of the dissipative work, Eqs. (6.1) and (6.7). This is due to the fact that our experimental setup lacks of a direct force measurement device. Although the force map of a single trap is known (see Fig. 6.4b), experimental observation showed that the potential felt by the bead when the two traps are separated does not coincide with the sum of two equal potentials obtained from the force map. This issue might be solved by using a different experimental setup equipped with a device allowing a direct measurement of the force, and therefore of the work.

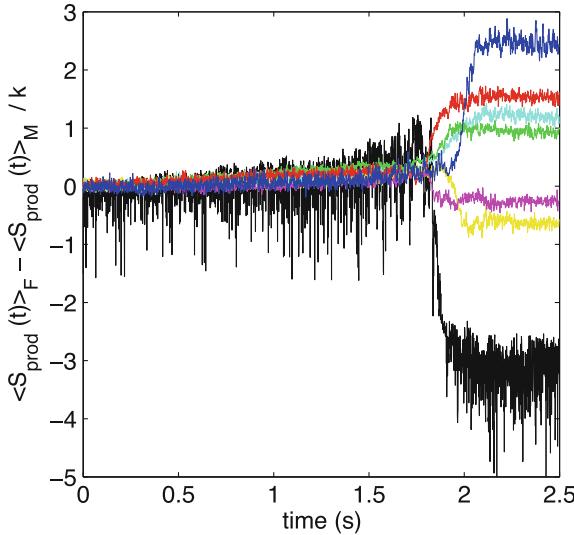


Fig. 6.12 Cumulated entropy production $\langle S_{\text{prod}}(t) \rangle_F - \langle S_{\text{prod}}(t) \rangle_M$ (in units of k) as a function of time t in the symmetry breaking process. The curves correspond to the experimental data of the fast protocol shown in Fig. 6.10. The probability of the bead to choose the F trap is $p_F = 0.01$ (black), $p_F = 0.15$ (yellow), $p_F = 0.34$ (magenta), $p_F = 0.71$ (cyan), $p_F = 0.71$ (green), $p_F = 0.80$ (red) and $p_F = 0.93$ (blue)

6.4 Building a Szilard Engine

As an illustration of the implications of these results, we construct a Szilard engine that extracts energy from the thermal bath, combining the SB and SR protocols. The engine can be implemented with an adequate combination of SB and symmetry restoration processes where the lower bound for the work done in the cycle (6.9), $kT \sum_i p_i \ln(p_i / \tilde{p}_i^i)$, is negative. The maximum work that can be extracted, or the minimum (most negative) value of the work is attained for $p_i = 1/2$ and $\tilde{p}_i = 1$, for $i = F, M$ as in the original Szilard cycle. The best we can do in our experiment is $p_F = 0.35$, $p_M = 0.65$, probabilities obtained for an external voltage $V_{\text{ext}} = 2$ V, and $\tilde{p}_M^M = 0.99$ for $V_{\text{ext}} = 4$ V and $\tilde{p}_F^F = 0.93$ for $V_{\text{ext}} = 0$ V. Our Szilard engine consists then on the following feedback protocol:

1. Symmetry breaking: The SB is induced with an external voltage external voltage $V_{\text{ext}} = 2$ V.
2. Symmetry restoration with feedback: If the particle end in the fixed trap (blue curves) we restore the symmetry with no external field, $V_{\text{ext}} = 2$ V. If the particle end in the moving trap (red curves) we restore the symmetry with $V_{\text{ext}} = 4$ V.
3. Reset of the potential: The cycle should be completed tuning quasistatically the external voltage back to its initial value $V_{\text{ext}} = 2$ V.

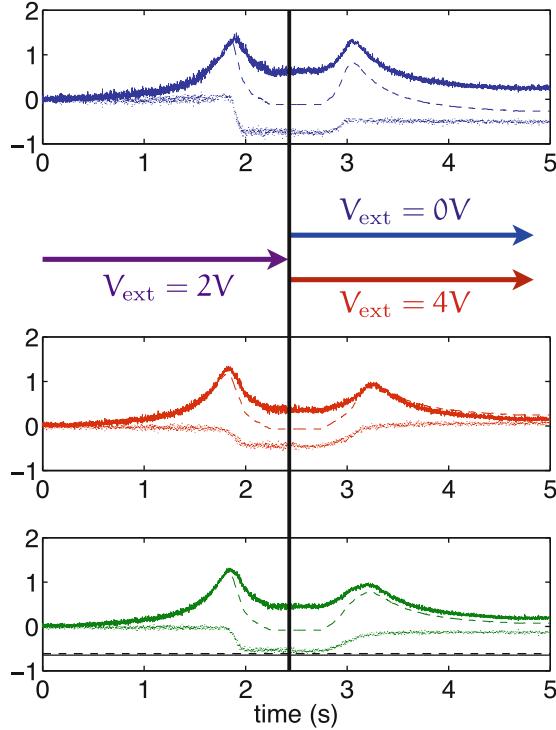


Fig. 6.13 Experimental realization of the Szilard engine. We show the average heat (solid lines) the Shannon entropy of the trajectory distribution (dashed lines), and the average entropy production (dotted lines). Heat is plotted in units of kT and entropies in units of k . The upper plot (blue curves) corresponds to averages over trajectories that end in the fixed trap, the middle plot (red curves) to averages over trajectories that end in the moving trap, and the lower plot shows averages over all trajectories. The feedback protocol is indicated by the arrows. The symmetry breaking is created with an external voltage $V_{\text{ext}} = 2 \text{ V}$ which induces probabilities $p_F = 0.35$, $p_M = 0.65$. When the particle chooses the fixed trap (blue) the symmetry is restored decreasing the voltage to $V_{\text{ext}} = 0 \text{ V}$, and so biasing the potential towards the fixed trap ($\tilde{p}_F = 0.93$). When the particle ends in the moving trap, the symmetry is restored increasing the voltage to $V_{\text{ext}} = 4 \text{ V}$, and biasing the potential towards the moving trap ($\tilde{p}_M = 0.99$). We show the bounds from the KPB theorem in the bottom panel: $-kT H(p_i)$ (black solid line) and $-\sum_i p_i \ln \frac{p_i}{\tilde{p}_i}$ (black dashed line)

This last step has not been implemented in the experiment, but it is assumed that it can be realized with arbitrary small entropy production. Figure 6.13 shows the average heat (solid curves), the change of Shannon entropy of the probability distribution of the particle (dashed curves) and the average entropy production (dotted curves). The averages are taken over trajectories that end in the fixed trap (upper plot, blue curves), the moving trap (middle plot, red curves), and over all trajectories (lower plot, green curves). We observe that using a feedback protocol such that $\tilde{p}_i^j > p_i$ we obtain a negative value for the average $\langle S_{\text{prod}} \rangle_F$ and a positive but small value for $\langle S_{\text{prod}} \rangle_M$

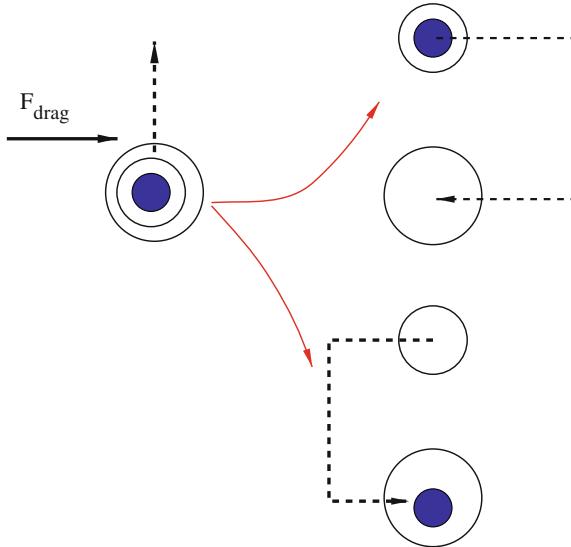


Fig. 6.14 Optimal Szilard engine design proposal. A Brownian particle (*blue filled circle*) is trapped with two optical traps (*black circles*) within an external uniform force F_{drag} in the x direction. The F trap is represented by the *bigger circle* whereas the M trap by the *smaller circle*. The M trap can be moved in both x and y axes. The M trap is separated from the F vertically, which implies that $p_F = p_M = 1/2$. If the bead chooses the M trap, we the M trap re-enters to its original position from the *right*, yielding $\tilde{p}_M^M \simeq 1$ when the external force F_{drag} is large. If the bead chooses the F trap, the re-entering of the M trap at the end of the process is done from the *left*, yielding $\tilde{p}_F^F = 1$ for strong external forcing

when measuring the entropy production in the whole cycle. In addition, the average over all the trajectories is also negative $\langle S_{\text{prod}} \rangle = \langle S_{\text{prod}} \rangle^{(\text{SB})} + \langle S_{\text{prod}} \rangle^{(\text{SR})} < 0$. Notice however that the negative average entropy production is compensated by the information entropy since $\langle S_{\text{prod}} \rangle + kT H(p_i) > 0$ [17]. More precisely, the total entropy production in the cycle averaged over all trajectories is $\langle S_{\text{prod}} \rangle/k = -0.12$ which is in accordance with the bound provided by the KPB (6.9), since $\langle S_{\text{prod}} \rangle/k > -\sum_i p_i \ln \frac{p_i}{\tilde{p}_i} = -0.62$ (see bottom panel in Fig. 6.13). The maximum work that could be extracted for the SB probabilities p_i is also indicated in Fig. 6.13, $H(p_i) = -0.65$, which also bounds from below the average entropy production.

As a further application, we introduce a new experimental design to construct the original Szilard engine using optical tweezers. Such engine could be implemented with a Brownian particle trapped with two optical traps that generate exactly the same optical potential. If the two traps are separated by moving one of the traps in a direction that is perpendicular to the external field, the particle chooses the $i = F, M$ trap with probability $p_i = 1/2$ in both cases, as illustrated in Fig. 6.14. After the choice, the following feedback process is done: If the particle chooses the M trap, the moving trap is moved to its original position opposing the flow. If the particle chooses the F trap, the moving trap is moved to its original position in the direction

of the flow. With this feedback process, the probabilities to choose the i trap in the backward process when the particle chooses the i trap in the forward process is close to 1 when the flow is large.

6.5 Numerical Simulations

As a complement to our experiment, we study the energetics of the symmetry breaking in the case where it is possible to measure directly the force exerted by the traps on the Brownian particle. We simulate a Brownian particle trapped with a dual optical trap and study the relation between the work dissipation and the probability of choice in a symmetry breaking process [Eq. (6.1)] analogous to our experimental protocol shown in Fig. 6.5.

We simulate the motion of a Brownian sphere of diameter $d = 1.2 \mu\text{m}$ immersed in water at room temperature, where $kT = 4 \text{ pN nm}$. The particle can move in 2 dimensions, x and y , and it is trapped with two optical tweezers of stiffness κ whose potential is

$$V(x, y) = -\kappa r^2 e^{-(x^2+y^2)/2r^2}, \quad (6.18)$$

where $r = d/2$ is the radius of the particle. This potential produces a force in the x direction

$$F_x(x, y) = -\frac{\partial V(x, y)}{\partial x} = -\kappa x e^{-(x^2+y^2)/2r}, \quad (6.19)$$

and in y direction, $F_y = -\kappa y e^{-(x^2+y^2)/2r}$. This force map is linear close to the center of the particle, $F_x \simeq -\kappa x$ for $x \ll r$ and also in the y direction. The nonlinear factor in the potential, $e^{-(x^2+y^2)/2r}$, is necessary to produce a bistable potential when separating two traps.

To study the energetics of the SB, we simulate a physical process that is analogous to the experimental protocol described in Fig. 6.5. The particle is trapped with two optical traps whose force map was described above. One of the two traps, F , is held fixed in $(0, 0)$ along the process and the other trap M can be moved in x direction. Initially, the center of the two traps (F and M) are in $(0, 0)$ for $\tau_1 = 2 \text{ s}$, so the bead relaxes to equilibrium. Then, the center of the M trap is moved along x -axis from $(0, 0)$ to $(L, 0)$ at constant velocity v_{trap} and let again relax to equilibrium during 2 s. A constant force which resembles the electric force in our experiment, F_e , is applied in the x direction in our model to tune the probability of the particle to end in the M or F trap after the process. We model the dynamics of the Brownian particle with the overdamped Langevin equation. When the M trap is separating at constant velocity v_{trap} from the origin, the equations of motion of the center of the particle are

$$\begin{aligned} \gamma \dot{x}_t &= F_e + F_{F,x}(x_t, y_t) + F_{M,x}(x_t - v_{\text{trap}}t, y_t) + \xi_{x,t}, \\ \gamma \dot{y}_t &= F_{F,y}(x_t, y_t) + F_{M,y}(x_t - v_{\text{trap}}t, y_t) + \xi_{y,t}, \end{aligned} \quad (6.20)$$

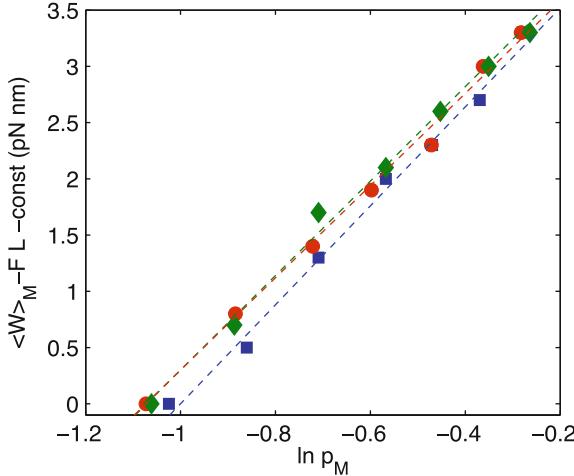


Fig. 6.15 Work dissipation in simulations of a symmetry breaking for different number of realizations. $\langle W \rangle_M - F_e L$ (in pN nm) as a function of $\ln p_M$ for values of F_e ranging from -0.02 pN to 0.04 pN. Each point in the figure is obtained measuring the average work and the probability p_M in a simulation for a fixed value of F_e . In every simulation, the M trap is moved at $v_{\text{trap}} = 300$ nm/s in time steps of $\Delta t = 10^{-3}$ ms. The average of the work is done over 1,000 realizations (blue squares), 2,000 realizations (red circles) and 3,000 realizations (green diamonds). Dashed lines are linear fits of slope (in pN nm) 4.42 (1,000 realizations, blue), 4.13 (2,000 realizations, red) and 4.17 (4,000 realizations, green)

where γ is the friction coefficient of the particle in water, which can be approximated following Stokes' law by $\gamma = 6\pi\eta d$, η being the viscosity of water at room temperature, $\eta = 0.89$ mPa s. The terms $\xi_{x,t}$ and $\xi_{y,t}$ are Gaussian white noise terms of zero mean and correlation $\langle \xi_t \xi'_t \rangle = 2kT\gamma\delta(t - t')$ that account for the thermal fluctuations.

We simulate our model using Heun's numerical integration method described in Sect A.2. For different values of the external force from -0.02 to 0.04 pN we measure the average work done when the particle ends in the M trap, $\langle W \rangle_M$, where the average is done over more than 1,000 trajectories. The work done in a single stochastic trajectory is given by Sekimoto's formula (1.59). In this case, the control parameter is the center of the M trap $\lambda(t) = x_{\text{trap},t} = v_{\text{trap}}t$, and Eq. (1.59) yields

$$\begin{aligned}
 W[x_\tau, y_\tau] &= \int_0^\tau \frac{\partial V_M(x_t - x_{\text{trap},t}, y_t)}{\partial x_{\text{trap},t}} \circ v_{\text{trap}} dt, \\
 &= \int_0^\tau F_M(x_t - x_{\text{trap},t}, y_t) \circ v_{\text{trap}} dt,
 \end{aligned} \tag{6.21}$$

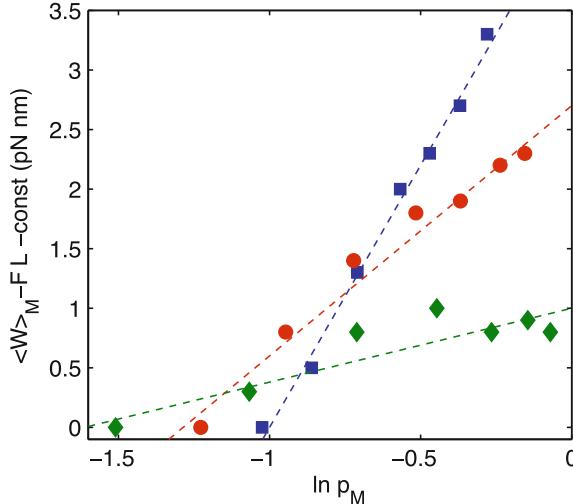


Fig. 6.16 Work dissipation in simulations of a symmetry breaking when the temperature of the environment. $\langle W \rangle_M - F_e L$ (in pN nm) as a function of $\ln p_M$ for values of F_e ranging from -0.02 pN to 0.04 pN. Each point in the figure is obtained measuring the average work and the probability p_M in a simulation for a fixed value of F_e . In every simulation, the M trap is moved at $v_{\text{trap}} = 300$ nm/s in time steps of $\Delta t = 10^{-3}$ ms. The temperature of the thermal bath is $kT = 4$ pN nm (blue squares), $kT = 2$ pN nm (red circles) and $kT = 1$ pN nm (green diamonds). Dashed lines are linear fits whose color corresponds to the data that is fitted, the slopes being (in pN nm) equal to 4.42 (blue, $kT = 4$ pN nm), 2.09 (red, $kT = 2$ pN nm) and 0.69 (green, $kT = 1$ pN nm)

where \circ denotes Stratonovich product and $\{x_\tau, y_\tau\} = \{x(t), y(t)\}_{t=0}^\tau$ is a stochastic trajectory of the center of the particle starting in $t = 0$ (after the initial $\tau_1 = 2$ s in equilibrium) and ending in $\tau = L/v_{\text{trap}}$. The average of the work over M trajectories minus the free energy difference is related to the probability to choose the M trap according to Eq. (6.1). If the trapping potential of the two traps is quadratic close to the center, we can calculate analytically the free energy difference between the final (F or M) and the initial state of the particle. In Sect E.2 we show that the average work done over the M trajectories for weak external forces ($F_e \ll \kappa L$) satisfies

$$\langle W \rangle_M - F_e L \geq kT \ln p_M + C, \quad (6.22)$$

where C stands for a constant term that does not depend on the value of the external force (see Sect E.2). Equation (6.22) results from applying (6.1), $\langle W \rangle_F - \Delta F_M \geq kT \ln p_M$, to our model by assuming that the trapping potential is quadratic close to the center of the traps.

We test the validity of (6.22) using traps of the same stiffness $\kappa = 280$ pN/ μm and moving apart the traps a distance $L = 3\mu\text{m}$, which ensures that F_e is one order of magnitude below κL . The velocity of the M trap is set to $v_{\text{trap}} = 300$ nm/s. In Fig. 6.15 we show the dependence of $\langle W \rangle_M - F_e L$ with $\ln p_M$ when the average of

the work is done over $N_r = 2,000, 4,000$ and $5,000$ realizations. The slope of the curve differs from kT in less than a 10 % in all the cases, which proves the validity of Eq. (6.22) for this system. The equation is also tested for different temperatures of the water ranging from $kT = 1$ to 4 pN nm as shown in Fig. 6.16.

Equation (6.1) relating average dissipation and probability of choice in a symmetry breaking could be tested experimentally using optical tweezers and using a direct measurement of the force [1, 9] of the two traps on the Brownian particle, or at least of the force exerted by the M trap. Using (6.21), one can measure the work done on the particle using the value of the force exerted by the M trap and the velocity of the center of the M trap in the experimental protocol. This result could be applied to infer the probability of a protein or a DNA molecule to adopt a given configuration from nonequilibrium measurements of the average work dissipated when the molecule changes its conformation. One candidate for such experiment is the protein *ribonuclease H* found in *E. coli* which unfolds from a native state to two different conformations (unfolded and intermediate) which can be distinguished from inspection of the force-extension curves [4].

6.6 Conclusions

The energetics of a generic symmetry breaking (SB) or a symmetry restoration (SR) can be reproduced by the quantitative relationship between irreversibility and dissipation for nonequilibrium processes that occur in the microscopic scale, as we have shown in Sect. 6.1. Our equation is universal since it does not depend on the nature of the (microscopic) physical system or of the mechanism that favors any option i and it is valid for any probability p_i . Our formula correctly predicts that a shrinkage in the accessible phase space leads to a negative entropy production. Moreover, a similar expression is found for a SR, which, together with the equation for the SB, form the core of the thermodynamics of choice. Our two formulas reproduce Landauer principle, the energetics of the Szilard engine and recent expressions obtained in the framework of the generalization of the second law to feedback-controlled processes.

An experimental test of our formulas is done in Sect. 6.3. We have designed an experiment where a Brownian microscopic dielectric particle is trapped with a dual optical tweezer where one of the traps is separated from the other (details of the experimental setup are described in Sect. 6.2). Using an external DC electric field, we were able to tune the probability of the particle to end in any of the traps, and therefore to implement a SB for different values of p_i . A SR is implemented by retracing the path of the moving trap to its original position. By measuring the average entropy production and the probability to end in the two traps, we reproduce our main formulas derived in Sect. 6.1. The experimental test covers a wide range of probabilities to choose an option, from 5 to 95 %.

Our formulas have relevant implications in the thermodynamics of choice of small systems. For example, they can be applied to estimate the probability of a protein to choose between different conformations using the data of the work dissipated

along its conformational changes. Moreover, our results suggest that one can design a Szilard engine with the concatenation of an SB with an SR using our setup. In Sect. 6.4, we have explored this possibility by designing such an engine, where the average entropy production after a cycle is negative (but still is positive when taking into account the information entropy).

As a further application, we have studied in Sect. 6.5 the relationship between average work dissipation and the choice probability in an SB using data from simulations of a Brownian particle trapped with two optical tweezers that are described by non-harmonic potentials. We have tested our main equation relating dissipated work with the probability of choice and discussed a further application of our formula to infer the probability of a protein to choose a specific conformation from only work measurements.

References

1. A. Alemany, A. Mossa, I. Junier, F. Ritort, Nat. Phys. **8**, 688–694 (2012)
2. A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, E. Lutz, Nature **483**, 187–189 (2012)
3. F. Beunis, F. Strubbe, K. Neyts, D. Petrov, Phy. Rev. Lett. **108**, 016101 (2012)
4. C. Cecconi, E.A. Shank, C. Bustamante, S. Marqusee, Science **309**, 2057–2060 (2005)
5. H. Felgner, O. Müller, M. Schliwa, Appl. Opt. **34**, 977–982 (1995)
6. L. Granger, H. Kantz, Europhys. Lett. (EPL) **101**, 50004 (2013)
7. J.M. Horowitz, J.M.R. Parrondo, New J. Phys. **13**, 123019 (2011)
8. M. Jahnel, M. Behrndt, A. Jannasch, E. Schffler, S.W. Grill, Opt. Lett. **36**, 1260 (2011)
9. I. Junier, A. Mossa, M. Manosas, F. Ritort, Phys. Rev. Lett. **102**, 070602 (2009)
10. R. Kawai, J.M.R. Parrondo, C.V. den Broeck, Phys. Rev. Lett. **98**, 80602 (2007)
11. H.S. Leff, A.F. Rex, D.L. Hogenboom, Am. J. Phys. **60**, 282 (1992)
12. P. Maragakis, M. Spichty, M. Karplus, J. Phys. Chem. B **112**, 6168–6174 (2008)
13. I.A. Martínez, D. Petrov, Appl. Opt. **51**, 5973–5977 (2012)
14. I.A. Martínez, D. Petrov, Appl. Opt. **51**, 5522–5526 (2012)
15. P. Mestres, D. Petrov, Eur. Biophys. J. **40**, 1081–1085 (2011)
16. J.M.R. Parrondo, Chaos: an interdisciplinary. J. Nonlinear Sci. **11**, 725–733 (2001)
17. T. Sagawa, M. Ueda, Phys. Rev. Lett. **102**, 250602 (2009)
18. U. Seifert, Phys. Rev. Lett. **95**, 40602 (2005)
19. S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, M. Sano, Nat. Phys. **6**, 988–992 (2010)
20. K. Visscher, S.P. Gross, S.M. Block, IEEE J. Sel. Top. Quantum Electron. **2**, 1066–1076 (1996)

Chapter 7

Effective Heating with Random Forces

The use of optical tweezers to study the behavior of microscopic systems driven out of equilibrium has been extended in the last decades, as discussed in the previous chapters. Apart from driving Brownian particles under different nonequilibrium scenarios using forces of the amplitude of the fluctuating forces exerted by the environment, it is also possible to increase the amplitude of the fluctuations by exerting random forces on the particle, as shown experimentally in [12]. In the opposite direction, cooling techniques have also been developed and sub-milikelvin kinetic temperatures have been experimentally achieved using Brownian particles recently [16]. In this chapter, we show how it is possible to tune experimentally the effective kinetic temperature of a polystyrene microscopic sphere immersed in water by using optical tweezers and random electric fields. Our experimental technique allows one to control the effective temperature of a Brownian particle from room temperature to 3000 K (i.e. ten times the vaporization temperature of water) in timescales of the order of \sim ms. Our setup can be applied to sample energy landscapes where Brownian particles cannot overcome high energy barriers at room temperature.

When random forces are applied to microscopic systems, the fluctuations of the system, which are affected by both the thermal force and the external random force, are of greater amplitude than if the random external force were not applied. In this situation, the system is said to be in a thermal bath of an *effective* temperature that is higher than the temperature of the surrounding thermal bath. This effective temperature can be increased, for example, using a laser whose wavelength is close to the absorption peak of the environment [22]. Using this technique, one can only increase the effective temperature of a particle up to the vaporization temperature of the material that forms the environment. We adopt a different approach, consisting on exerting random forces on a Brownian particle that increase the effective (kinetic) temperature of the particle without increasing the temperature of the environment.

Our experiment consists of a Brownian polystyrene spheric bead immersed in water that is trapped with an optical tweezer. The bead is charged, and external

To the memory of D. Petrov, who passed away on February 3rd, 2014.

random electric fields are applied with two electrodes, which increase the amplitude of the fluctuations of the position of the bead. With our setup, Brownian particles respond to a temperature change in $\sim \text{ms}$ and can achieve a maximum temperature of 3000 K without heating the surrounding water. We study both the fluctuations and the response of the bead to nonequilibrium perturbations.

This chapter is organized as follows. Section 7.1 is an introduction to the rest of the chapter where we discuss the state of the art of the cooling and heating experimental techniques applied to microscopic systems. In Sect. 7.2 we show the experimental setup that we use to tune the temperature of Brownian particles. Sections 7.3 and 7.4 show our main results concerning the variation of the effective temperature of the particles in an equilibrium and nonequilibrium situations, respectively. In Sect. 7.5 we show how our experiment can be used to accelerate the sampling of the potential created by a dual trap as well as to build microscopic heat engines, and in Sect. 7.6 we discuss all the results of this chapter.

7.1 Controlling the Temperature of Small Systems

Controlling the temperature of small systems is an emerging topic in physics. During the last years, there has been an increasing interest in cooling small systems in order to observe quantum properties that only arise at low temperatures. For example, a cooled optically-trapped system in vacuum can be used to search for non-Newtonian gravity forces at small scales or to produce quantum superpositions of living organisms [16]. Since first Ashkin's experiments [1], where the feedback stabilization of optically trapped particle was suggested, cooling of the center-of-mass motion of an optically trapped microspheres has advanced rapidly, and now the cooling below to several millikelvin of microscopic spheres has become possible using active optical feedback [16]. This technique has also been applied to nano particles [10]. Such a controllable attenuation of the motion is an important step towards new experiments in different areas of physics. However, we now suggest to go in the opposite direction by controlling the motion of optically trapped microscopic spheres, namely to increase in a controllable way the Brownian fluctuations. We are therefore interested in *heating* instead of cooling the center-of-mass motion of a Brownian particle.

The Brownian motion of microspheres in liquid has offered a model system for experimental studies of processes where thermal fluctuations are important. Out of equilibrium processes in biophysics and colloidal systems [2, 6, 8], Kramers transitions [19], and a micrometer-sized stochastic engine [4] are just several examples of such processes. We present an experimental technique that allows one to increase the amplitude of the Brownian fluctuations in a controllable way with millisecond response time, extending significantly the range of parameters used in modeling of equilibrium and out-of-equilibrium processes. Our experiment does not require any additional source of heat like optical beams or micro ovens.

Both the temperature of the thermal bath T and its viscous friction coefficient γ affect the amplitude of Brownian fluctuations. We are not aware of methods allowing

to change dynamically the viscosity, however, the temperature can be controlled by heating the Brownian sphere or its surroundings. Spherical colloids [22] or nanorods [23] were selectively heated by external sources of a light which is not absorbed by the solvent. In [4, 18], the temperature in the surroundings of a dielectric sphere was varied using an optical beam whose wavelength matches an absorption peak of the solvent. The maximum temperature that can be reached using these methods in aqueous experiments is restricted by the vaporization temperature of the surrounding fluid. In all the mentioned experiments, the temperature increase did not exceed several tenths of kelvin and response times were of the order of several ten of milliseconds.

In equilibrium, fluctuations of the position and of the velocity of a Brownian particle depend on the temperature via the equipartition theorem. Therefore designing experiments to enhance or reduce the fluctuations would result on an increase or decrease of the effective kinetic temperature of a particle. Active optical feedback permitted to cool down the kinetic temperature of optically trapped microspheres [10, 16], and the same technique may increase the kinetic temperature if the active control is tuned to increase the sphere displacements. Here we suggest an alternative and simpler method, based on combining an optical trap with a source of additional external forces acting on the trapped sphere. With our technique, the Brownian sphere could fluctuate in water in one of the directions with the center-of-mass motion corresponding to the temperature of several thousand kelvin, however, the viscous coefficient of the liquid and the temperature, describing the center-of-mass motion in two other directions, remain the same as at the temperature of thermal bath.

Consider an overdamped Brownian particle trapped with an optical tweezer of stiffness κ in a fluid at temperature T . For simplicity, we only consider the dynamics in one dimension, x , but the particle can in principle move in three dimensions. The optical trapping potential in x -axis, $U(x)$, is assumed to be quadratic and centered at x_0 :

$$U(x) = \frac{1}{2}\kappa(x - x_0)^2. \quad (7.1)$$

If an external force $\zeta(t)$ acts on the particle and it is applied in x direction, then the position along the direction of the external force, $x(t)$, obeys the overdamped Langevin equation [cf. Eq. (1.49)]

$$\gamma \dot{x}(t) = -\kappa[x(t) - x_0] + \xi(t) + \zeta(t), \quad (7.2)$$

where γ is the friction coefficient of the Brownian particle in the fluid and $\xi(t)$ is a Gaussian white noise with zero mean and correlation $\langle \xi(t)\xi(t') \rangle = 2kT\gamma\delta(t - t')$, which models thermal fluctuations. We assume that the external force $\zeta(t)$ is also described by a stochastic process independent of the thermal noise, with zero mean and correlation $\langle \zeta(t)\zeta(t') \rangle = \sigma^2\Gamma(t - t')$, where $\Gamma(\tau)$ is a normalized function, $\int_{\mathbb{R}} \Gamma(\tau) d\tau = 1$, peaked at $\tau = 0$, and σ is the amplitude of the external noise. If the external noise is white, i.e., $\Gamma(\tau) = \delta(\tau)$, then the particle is subject to an effective noise $\xi_{\text{eff}}(t) = \xi(t) + \zeta(t)$, which is a random Gaussian process with zero mean and

correlation $\langle \xi_{\text{eff}}(t)\xi_{\text{eff}}(t') \rangle = 2kT_{\text{kin}}\gamma\delta(t-t')$, where T_{kin} is defined as

$$T_{\text{kin}} = T + \frac{\sigma^2}{2k\gamma}. \quad (7.3)$$

Hence the position of the particle in x -axis fluctuates corresponding to a kinetic temperature T_{kin} which is always higher than the temperature T of the thermal bath. Therefore, an external random force exerted on the sphere increases the kinetic temperature of the sphere without real heating of the surrounding fluid, which remains at temperature T . The motion of the particle in x -axis is in every respect identical to the motion of a Brownian particle immersed in a thermal bath at temperature T_{kin} (this holds for underdamped particles as well). However, this is valid only if the external noise has white power spectral density (PSD).

7.2 Experimental Setup

We now discuss the experimental setup that we designed to tune the effective temperature of Brownian microscopic spheres. The experiment was done by Ignacio A. Martínez and Dmitry Petrov in ICFO, Institute of Photonic Sciences.¹ The experimental setup is analog to the one used to test our analytical results in the energetics of symmetry breaking, described in Sect. 6.2. The setup includes the following principal parts shown and explained in Fig. 7.1.

Polystyrene micro-spheres (G. Kisker-Products for Biotechnology, <http://www.kisker-biotech.com/>), polystyrene microparticles PPs-1.0 with a diameter of $1.00 \pm 0.05 \mu\text{m}$) were diluted in distilled de-ionized water to a final concentration of a few spheres per ml.

The spheres were inserted into a custom-made electrophoretic fluid chamber with two electrodes connected to a computer-controlled electric generator and an amplifier. The chamber was previously described in [3, 20]. The spheres were trapped above the bottom surface of the chamber at a distance $\sim 20 \mu\text{m}$ controlled by the piezoelectric stage. The value of the friction coefficient γ was corrected due to the proximity of the chamber surface using the Faxen law as described in [25].

Electrical signals applied with the electrodes were produced by the generator fed with a noisy sequence close to Gaussian white noise. The sequence was generated using independent distributed random variables $\zeta_{n\delta t}$, with $\langle \zeta_{n\delta t} \rangle = 0$, $\langle \zeta_{n\delta t}^2 \rangle = A$, and $\langle \zeta_{n\delta t} \zeta_{m\delta t} \rangle = 0$ for $n \neq m$, A being the variance of the noise [28]. The time interval δt is the clock time of the generator, which is $\delta t = 0.2 \text{ ms}$ in the experiments. Notice that δt is a lower bound to the actual correlation time of the external force, which also depends on the response characteristics of the amplifier.

¹ Full address: ICFO—Institut de Ciències Fotòniques, Optical Tweezers Group, Mediterranean Technology Park, Av. Carl Friedrich Gauss, num. 3, 08860 Castelldefels (Barcelona), Spain.

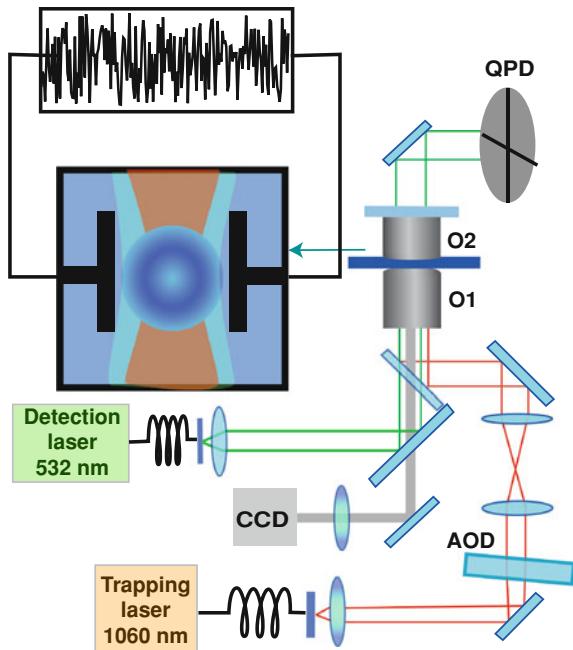


Fig. 7.1 Experimental setup An acousto-optical deflector (AOD) (ISOMET LS55 NIR) steers a 1060 nm optical beam from a laser coupled into a single-mode fiber (ManLight, ML10-CW-P-OEM/TKS-OTS, maximal power 3 W). The AOD modulation voltage is obtained from an arbitrary waveform generator (TaborElectronics WW1071) controlled by a LabView program. The beam deflected by the AOD, is expanded and inserted through a oil-immersed objective O1 (Nikon, CFI PL FL 100X NA 1.30) into a custom-made fluid chamber. An additional 532 nm optical beam from a laser coupled to a single-mode fiber (OZOptics) is collimated by a ($\times 10$, $NA = 0.10$) microscope objective and passes through the trapping objective. The forward scattered detection beam is collected by a ($\times 10$, $NA = 0.10$) microscope objective O2, and its back focal-plane field distribution is analyzed by a quadrant position detector (QPD) (New Focus 2911) at an acquisition rate of 20 kHz. A 532 nm band pass filter in front of the QPD blocks beams with wavelengths different from the detection beam wavelength. The AOD permits the control of the position of the beam focus. A fluid chamber with microspheres was placed on a piezoelectric-controlled calibrated stage (PiezosystemJena, Tritor 102) allowing the 3D translation. Two electrodes are connected to the two ends of the fluid chamber to exert random forces on the microspheres. The electrodes are connected to an electric generator that can be fed by noisy signals with controllable PSD. The inset of the figure sketches an optically trapped sphere between the electrodes

The spheres are trapped using an infrared ($\lambda_{\text{trap}} = 1060 \text{ nm}$) laser trap described in Fig. 7.1. The orientation of the AOD can be controlled externally, which allows one to modify the position of the trap center x_0 . A second green detection laser ($\lambda = 532 \text{ nm}$) is used to monitor the position of the particles. A position detection system equipped with a 532 nm band pass filter is placed in the back-focal plane of the objective. The position detection is done by a quadrant position detector (QPD), which measures the intensity of the detection laser, and has a 1-nm sensitivity, which

is above the resolution of CCD cameras ($\sim 5\text{--}10\text{ nm}$). At last, the kinetic temperature of the optically trapped sphere is obtained from the analysis of the output signals of the QPD.

The system needs two calibration procedures previously described in Sect. 6.2: calibration of the position detection system and calibration of the force exerted by the trap on the sphere. The displacements of the sphere produces a unique and linear QPD response only within a restricted region of the sphere displacement, usually several hundred nanometers. Since an increase of the amplitude of the displacements by application of the additional noise is expected, the linear range of the detection system and the force map of the optical trap need to be characterized.

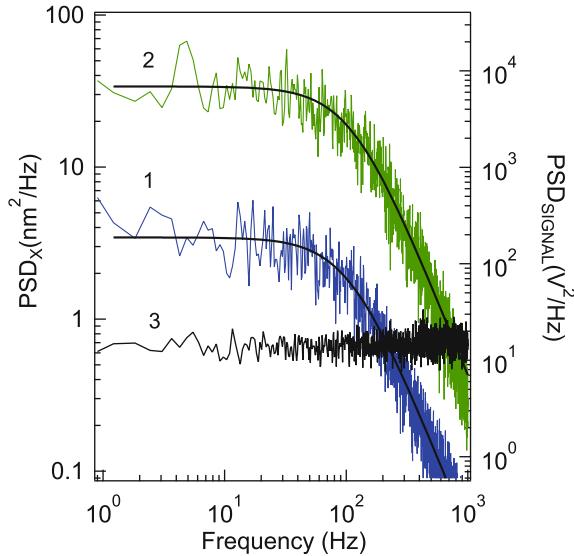
First, we calibrated the QPD measuring the ratio between QPD output signals (in Volts) and real displacements of the probe (in nm) by using natural thermal fluctuations of the probe as described in [27]. The trap was calibrated in the absence of the additional force. Using a relatively strong trap where the Brownian fluctuations did not overcome several tens of nanometers we first obtained the *Power Spectrum Density* (PSD) of the position of the particle. The PSD is defined as the Fourier transform of the autocorrelation of the position $\text{PSD}(f) = \mathcal{F}[\langle x(0)x(t) \rangle]$, f being the frequency. Using the PSD one can not only characterize parameters of the trap such as the stiffness but also detect any misalignment in the trapping setup or the detection system [27]. For a Brownian particle in a harmonic trap and described by an overdamped Langevin equation, the PSD of the position can is a Lorentzian function (see Sect. F.1). In our experiment, we fit the experimental PSD to a Lorentzian function

$$\text{PSD}(f) = \frac{1}{2\pi^2\gamma} \frac{kT}{f^2 + f_c^2}, \quad (7.4)$$

where $f_c = \kappa/(2\pi\gamma)$ is the corner frequency. From the fit to (7.4) we obtain the stiffness of the trap κ and the calibration coefficient $S_{\text{QPD}}(\text{nm/V})$ relating the QPD output signal with the absolute displacement of the sphere as described in [27]. In our experiment, $\kappa = 6\text{ pN}/\mu\text{m}$ and $S_{\text{QPD}} = 1280\text{ nm/V}$. We remark that the displacements of the sphere from the trap center did not exceed the linear range of the optical trap in all our experiments, and therefore the harmonic approximation, given by (7.1), is valid to describe the optical potential felt by the Brownian particle.

We then calibrated the AOD by measuring the ratio between the amplitude of the AOD input modulation signal V_{AOD} (in Volts) and the corresponding sphere displacement (in nm) using the previously measured QPD calibration factor S_{QPD} . We moved the trapped sphere in the range of several ten nanometers changing the AOD modulation signal and measured the output signal of the QPD. This provided us the calibration factor $S_{\text{AOD}}(\text{nm}/V_{\text{AOD}})$ for the small range of the sphere displacement. We then confirmed using image analysis that this factor holds for the AOD modulation amplitudes corresponding to the sphere displacements within the range $\pm 3\text{ }\mu\text{m}$. This calibration procedure allows us to track the position of the trap center in time, $x_0(t)$, with sub-nanometer precision.

Fig. 7.2 Effect of the addition of a random force on the PSD
 Power spectrum density of the position of the sphere obtained from a 10 s time series without (blue curve, 1) and with (green curve, 2) the additional stochastic force. Solid black lines correspond to the Lorentzian fits. The power spectral density of the input noisy signal measured at the electrodes of the fluid cell is also shown (black line, right y-axis, 3)



The calibration procedures described above allow to express the experimental data in the physical units for both force exerted by the trap on the particle (pN), the sphere displacement (nm) and the position of the trap center (nm).

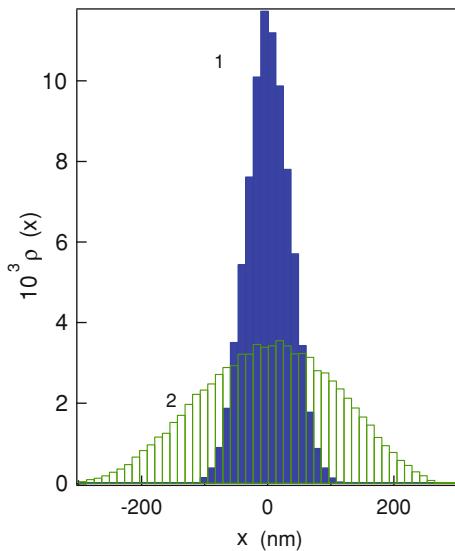
The stability of the electric charge of the bead was previously tested in Sect. 6.2, where we applied a sinusoidal voltage at 1 Hz to the electrodes of the electrophoretic chamber and observed that the amplitude of the position of the particle did not change more than 6 % during at least 3000 s (see Fig. 5.1c).

7.3 Brownian Sphere in a Stationary Trap with Additional Noise

We first investigate how the addition of an external random force alters the dynamics of a Brownian sphere trapped with an optical tweezer whose center is held fixed. We focus on the analysis on the fluctuations of the position of the particle, whose amplitude increases when adding an external random force. We discuss how one can define an effective kinetic temperature in this situation and how this temperature can be tuned using our setup.

We analyze the time traces of the position of the trapped sphere when the trap center is held fixed, for two cases: (i) in the absence of any external force $\zeta(t) = 0$ and (ii) at the maximum value of the amplitude of the external noise, where the position of the particle does not exceed the linear range of the QPD. PSDs of the position of the sphere are shown in Figs. 7.2 and 7.3 illustrates the probability density of the sphere position using data recordings of ~ 10 s. Notice that the data acquisition

Fig. 7.3 Effect of the addition of a random force on the position histograms Probability density of the position of the sphere without (1) and with the additional stochastic force (2) corresponding to the PSD curves (1) and (2) shown in Fig. 7.2



time exceeds in several orders of magnitude the relaxation time of the particle in the trap $\tau_r \sim \text{ms}$, which is measured from the PSD as we show below. Therefore using a 10 s experiment we are sampling the position of the particle after it has reached an stationary state. As shown in Fig. 7.3, the maximal sphere Brownian excursions without the additional noise do not exceed $\pm 50 \text{ nm}$, however with the additional stochastic force the maximal amplitude can reach $\pm 250 \text{ nm}$. Even in this case the sphere displacements do not exceed the linear ranges of both the position detection system and the harmonic approximation of the optical trapping potential holds valid.

Our generator allowed us to obtain an electrical signal with flat spectrum for frequencies up to 1 kHz. As it can be seen in Fig. 7.2, the PSD of the position of the particle in the presence of the additional force is also Lorentzian and the corner frequency f_c does not change, as expected, since it is a function only of κ and γ . We notice that $f_c \sim 100 \text{ Hz}$ which implies that the relaxation time of the bead is $\tau_r \sim 1/f_c \sim 10 \text{ ms}$. On the other hand, the height of the PSD increases due to the external force. Therefore, the PSD of the position of the bead when the field is applied is in every respect identical to the PSD of a Brownian sphere trapped by the same tweezer but immersed in a thermal bath at a temperature that is larger than the temperature of the water. This temperature increase can also be inferred from the modification of stationary probability density of the position of the sphere, $\rho_{ss}(x)$, depicted in Fig. 7.3, where it is shown that the presence of an external random force broadens the distribution of the position.

We now discuss the results shown in Figs. 7.2 and 7.3, which concern the modification of the PSD and the position histograms of the Brownian sphere in a fixed trap when applying an additional random force. As we have seen, the addition of a random external force affects both the autocorrelation and the histogram of the

position of the bead. Consequently, we can define *two* kinetic temperatures, either using the data of the PSD or $\rho_{ss}(x)$. As we show below, these two kinetic temperatures can differ if the spectrum of the additional noise is not white.

From the Langevin equation (7.2), we can calculate analytically both the PSD and the stationary probability distribution $\rho_{ss}(x)$ for an external noise $\zeta(t)$ with arbitrary correlation $\langle \zeta(t)\zeta(t') \rangle = \sigma^2 \Gamma(t-t')$, with $\int_{-\infty}^{\infty} \Gamma(t)dt = 1$ (see details in Sect. F.2). The PSD for such a system is given by:

$$\text{PSD}(f) = \frac{1}{4\pi^2\gamma^2} \frac{2\gamma kT + \sigma^2 \tilde{\Gamma}(f)}{f^2 + f_c^2}, \quad (7.5)$$

where $\sigma^2 \tilde{\Gamma}(f)$ is the Fourier transform of the correlation function of the external noise. If $\tilde{\Gamma}(f)$ is constant for frequencies much higher than the corner frequency of the trap f_c , then the PSD (7.5) is approximately a Lorentzian function with the same corner frequency. We can define the following effective kinetic temperature,

$$T_{\text{PSD}} \equiv \frac{2\pi^2\gamma f_c^2 \text{PSD}(0)}{k} = T + \frac{\sigma^2}{2k\gamma}. \quad (7.6)$$

Notice that, with this definition, T_{PSD} does not depend on the shape of the correlation function of the noise, $\Gamma(t)$.

On the other hand, we can define the following kinetic temperature derived from width of the probability density of the position,

$$T_{\text{hist}} \equiv \frac{\kappa \langle x^2 \rangle_{\text{ss}}}{k}, \quad (7.7)$$

that is, the temperature that enters into equipartition theorem, $\langle U \rangle = \frac{1}{2}\kappa \langle x^2 \rangle_{\text{ss}} = \frac{1}{2}kT_{\text{hist}}$. The stationary probability density $\rho_{ss}(x)$ can be calculated analytically for a Brownian particle described by the overdamped Langevin equation (7.2) (see Sect. F.2). The probability density in the stationary state is found to be Gaussian with zero mean and dispersion

$$\langle x^2 \rangle_{\text{ss}} = \frac{kT}{\kappa} + \frac{\sigma^2}{\gamma\kappa} \int_0^{\infty} dt e^{-\kappa t/\gamma} \Gamma(t), \quad (7.8)$$

yielding

$$T_{\text{hist}} = T + \frac{\sigma^2}{\gamma k} \int_0^{\infty} dt' \Gamma(t') e^{-t'/\tau_r}, \quad (7.9)$$

$\tau_r = \gamma/\kappa$ being the relaxation time of the particle in the trap. If $\Gamma(t)$ is peaked around $t = 0$ and the correlation time of the noise, given by $\tau_c \equiv \int_0^{\infty} dt t \Gamma(t)$, is

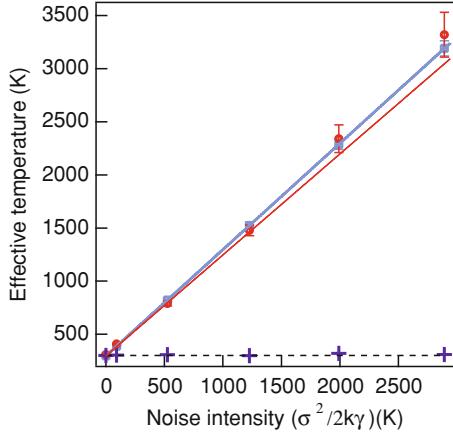


Fig. 7.4 Equilibrium effective kinetic temperatures as functions of $T_e = \sigma^2/(2k\gamma)$. Filled symbols correspond to experimental values of the effective temperature measured from the fluctuations and the response of the bead along the x-axis (direction of the additional noise): T_{PSD} (blue filled squares) and T_{hist} (red filled circles). Statistical errors are plotted in the error bars. Lines represent the analytical values obtained for non-white external noise $\zeta(t)$ with correlation function given by Eq. (7.11) for T_{PSD} (7.6) (blue solid line) and T_{hist} (7.9) (red dashed line). We also show the values of the effective temperature measured from the fluctuations along the y-axis, $T_{hist,y}$ (purple “+”). Statistical errors are smaller than the size of the symbols. Black dashed line is set to room temperature $T = 300\text{K}$

small compared to τ_r , T_{hist} is approximatively (see Sect. F.2)

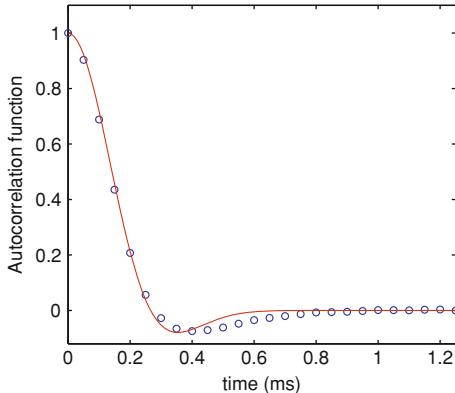
$$T_{hist} \simeq T + \frac{\sigma^2}{2k\gamma} \left(1 - \frac{2\tau_c}{\tau_r} \right). \quad (7.10)$$

Notice that T_{hist} , unlike T_{PSD} defined as (7.6), does depend on the correlation time of the noise. The difference between T_{PSD} and T_{hist} is proportional to τ_c/τ_r , therefore it vanishes if the external noise is a Gaussian white noise. In our experiments, $\tau_r = 1.4$ ms and $\tau_c = 0.26$ ms, being τ_c obtained experimentally from the spectral analysis of the voltage at the electrodes (see below).

We analyze quantitatively how the kinetic temperature of the microscopic sphere changes when we tune the amplitude of the external random force. For this purpose, we measure T_{kin} and T_{PSD} experimentally for different values of the noise intensity. Both effective temperatures increase with the intensity of the external force as shown in Fig. 7.4. Notice however that the temperature along *y-axis* remains unaltered when changing the noise intensity.

We now test if Eqs. (7.6) and (7.9) relating T_{PSD} and T_{hist} with the correlation function of the external noise, respectively, reproduce the experimental results. First, we obtain T_e for the different experimental data from T_{PSD} assuming that (7.6) is valid. The analytical value of T_{hist} depends not only on the amplitude but also on the correlation function of the external random force, whose shape is now discussed.

Fig. 7.5 Autocorrelation function of the random force directly measured in the electrodes as a function of time (in milliseconds): experimental data (blue squares) and fit to $\tilde{\Gamma}(t) = e^{-(t/\tau_c)^2} \cos\left(\frac{t}{\tau_o}\right)$ (red curve)



In practical situations, Gaussian white noise cannot be implemented and only noisy signals with flat spectrum up to a finite frequency can be designed experimentally. In our experiment, the spectrum of the external random force is flat only up to 1 kHz, as shown in Fig. 7.2. We also measured the autocorrelation function of the signal on the electrodes $V(t)$, defined in [5], $\tilde{\Gamma}(t) = \langle V(t)V(t+t) \rangle / \langle V(t)V(t') \rangle$, which can be fitted to the function $\tilde{\Gamma}(t) = e^{-(t/\tau_c)^2} \cos(t/\tau_o)$, being $\tau_c = 0.26\text{ ms}$ and $\tau_o = 0.17\text{ ms}$, as shown in Fig. 7.5. The correlation of the noise is related to $\tilde{\Gamma}(t)$ by $\Gamma(t) = N\tilde{\Gamma}(t)$, where N is a normalization constant such that $\int_{-\infty}^{\infty} \Gamma(t)dt = N \int_{-\infty}^{\infty} \tilde{\Gamma}(t)dt = 1$. The full expression of the correlation function of the additional noise in our experiment is

$$\Gamma(t) = \frac{e^{tc^2/4t_o^2}}{\sqrt{\pi}} \frac{e^{-(t/\tau_c)^2}}{\tau_c} \cos(t/\tau_o). \quad (7.11)$$

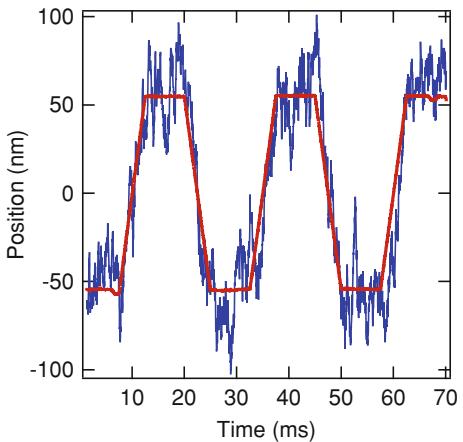
Using (7.11) in (7.9) we calculate analytically the value of T_{hist} as a function of the noise intensity which we show in Fig. 7.4.

In the figure, we plot the effective temperatures T_{PSD} and T_{hist} as functions of $T_e = \sigma^2/2k\gamma$, and we observe that T_{hist} and T_{PSD} coincide within statistical errors. For $\Gamma(t)$ defined in Eq. (7.11), we obtain using (7.9) $T_{\text{hist}} = T + 0.95 T_e$ which is slightly smaller than $T_{\text{PSD}} = T + T_e$ (7.6) and fits the experimental data.

7.4 A Nonequilibrium Process

It would be interesting to apply our experimental technique to implement nonequilibrium processes at the microscopic scale. To do so, a complete characterization of the effective thermal bath might be done, by including measurements from the response of the Brownian particle to a nonequilibrium driving.

Fig. 7.6 Position of the trap center (black line) and the position of the *sphere* as a function of time without external electric field (blue curve)



We now study how the presence of the additional random force alters the dynamics of the microscopic sphere in a nonequilibrium experiment. Fluctuation theorems can be used to study the response of the system to any nonequilibrium driving. Crooks fluctuation theorem (CFT) is an excellent candidate to validate the applicability of the technique to nonequilibrium processes, since it establishes a temperature-dependent relationship between the probability density function (PDF) of the work W along a process arbitrarily far from equilibrium, $\rho_F(W)$, and the PDF of the work in the time-reversed process, $\rho_B(W)$ [see Eq. (1.70)]. For our technique to be applicable to nonequilibrium situations, the temperature that enters in CFT should be equal to that obtained in equilibrium conditions.

We study the dynamics of the system under a nonequilibrium driving in which the trap center is displaced at a constant velocity. We have implemented the non-equilibrium protocol depicted in Fig. 7.6. In this experiment, the stiffness was set to $\kappa = 26 \text{ pN}/\mu\text{m}$, and the acquisition frequency to $f_{\text{acq}} = 3 \text{ kHz}$. We now describe the experimental protocol: In the first step (forward process), the trap center is moved from $x_0 = -L$ to $x_0 = +L$ at constant velocity $v = 2L/\tau$, where $L = 55 \text{ nm}$ and $\tau = 6.25 \text{ ms}$. We then let the sphere relax to the new equilibrium position, keeping the trap center at $x_0 = +L$ for the same time $\tau = 6.25 \text{ ms}$, which is larger than the relaxation time of the trap, $\tau_r = \gamma/\kappa \sim 0.5 \text{ ms}$. Then the trap is moved back from $x_0 = +L$ to $x_0 = -L$ with the same velocity $-v$ (backward process) and it is held fixed at $x_0 = -L$ for the same time τ . This protocol was repeated about 10^4 times. The position of the Brownian particle in the absence of noise follows the trap center along this protocol as shown in Fig. 7.6. When increasing the noise intensity, the fluctuations of the position increase, as in the equilibrium experiments (data not shown).

In each repetition of the nonequilibrium process, we calculate the work in the forward and time-reversed processes using the stochastic trajectories of the position of the bead using Sekimoto's expression for the work associated to a trajectory of an

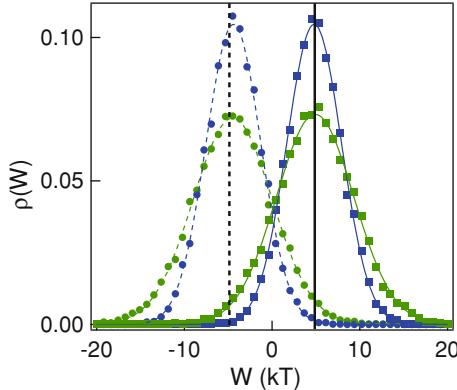


Fig. 7.7 Probability density function of the work obtained in the forward process, $\rho_F(W)$ and probability density function of minus the work in the backward process, $\rho_B(-W)$. Squares correspond to the forward process, $\rho_F(W)$, and the circles to the backward process, $\rho_B(-W)$, where blue and green symbols are obtained without and with additional external force (using the same noise amplitude as in the case of the curve (2) in Fig. 7.2), respectively. Solid blue and green curves are Gaussian fits. Vertical black lines represent the analytical value of the average work in the forward process calculated using Eq. (7.12) (solid line) and the average of $-W$ in the backward process (dashed line) which is calculated in an analogous way to Eq. (7.12)

overdamped Brownian particle (1.59). In this particular case, the control parameter is the position of the trap center, $\lambda(t) = x_0(t)$. The work distributions in forward and backward processes $\rho_F(W)$ and $\rho_B(-W)$ are Gaussian as shown in Fig. 7.7. Moreover, $\rho_F(W)$ and $\rho_B(-W)$ intersect at $W^* \simeq 0$, in accordance with the prediction of CFT [$W^* = \Delta F = 0$, cf. Eq. (1.73)]. When adding the external noise, both forward and backward work distributions broaden despite their mean value is the same as in the absence of the additional force. The intersection between the distributions $\rho_F(W)$ and $\rho_B(-W)$ is $W^* \simeq 0 = \Delta F$ in this case.

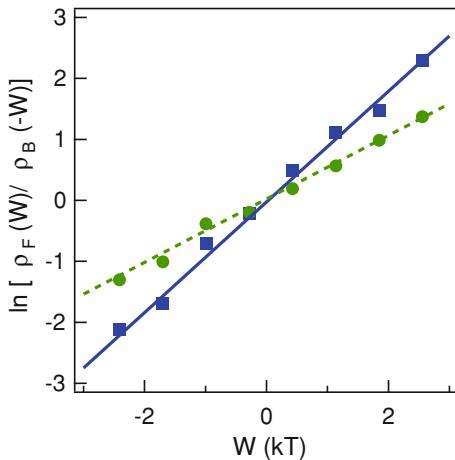
The value of the average work coincides with the analytical prediction, derived in Sect. F.3,

$$\langle W \rangle = \gamma v^2 \tau \left[1 + \frac{\tau_r}{\tau} (e^{-\tau/\tau_r} - 1) \right]. \quad (7.12)$$

Notice that there is no contribution of the external noise to the average work as long as the external noise enters into Langevin equation as an additive noise. Therefore, the addition of the external random field does not increase the average dissipation of the system, and the energy flow associated to this field can be considered as heat.

We now investigate the validity of Crooks Fluctuation Theorem (CFT) for different values of the amplitude of the additional noise. Figure 7.8 shows that $\ln [\rho_F(W)/\rho_B(-W)]$ depends linearly on W with and without additional noise, as expected by the CFT (1.70). In the absence of external noise, CFT is satisfied in our experiment since the slope of $\ln [\rho_F(W)/\rho_B(-W)]$ as a function of W/kT is 1 within experimental errors. With external noise, the slope of $\ln [\rho_F(W)/\rho_B(-W)]$

Fig. 7.8 Experimental test of Crooks fluctuation theorem $\ln [\rho_F(W)/\rho_B(-W)]$ as a function of W/kT without (blue squares) and with additional noise (green circles) with amplitude corresponding respectively to the PSD curves (1) and (2) shown in Fig. 7.2. The solid lines are linear fits of the experimental data



as a function of W/kT is always smaller than 1 and it decreases when increasing the amplitude of the external noise.

We introduce a new effective temperature, *Crooks Temperature* T_C , to check the consistency of our effective thermal bath for nonequilibrium processes. We define T_C as the *nonequilibrium* temperature that enters into CFT,

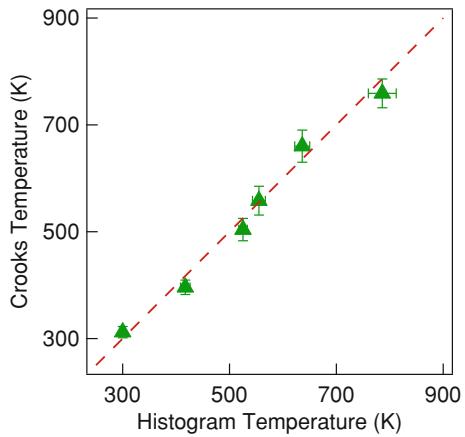
$$\ln \frac{\rho_F(W)}{\rho_B(-W)} = \frac{1}{kT_C}(W - \Delta F). \quad (7.13)$$

Clearly, in the absence of the external force, $T_C = T$, whereas a decrease of the slope in Fig. 7.8 when the external noise is added can be interpreted as an *increase* of the effective Crooks temperature. Therefore, nonequilibrium measurements also agree qualitatively with the behavior observed in equilibrium, where the effective kinetic temperature was shown to increase with noise intensity. More precisely, in the absence of external noise, we obtain $T_{\text{hist}} = (310 \pm 3)K$, $T_{\text{PSD}} = (297 \pm 7)K$ and $T_C = (283 \pm 1)K$, where the errors are purely statistical and therefore a lower bound to the real error.

When the external force is added, T_C increases with the noise intensity as shown in Fig. 7.9 and coincides with T_{hist} within experimental errors for different values of the noise intensity.

The result shown in Fig. 7.9 implies that our technique can be extended to the realization of nonequilibrium non-isothermal processes in the microscale.

Fig. 7.9 Effective Crooks temperature as a function of temperature obtained from histograms of the position



7.5 Applications of Our Setup

We now focus on the applications that could be implemented using our setup. Increasing the amplitude of the fluctuations of a Brownian particle might serve to sample energy landscapes where some regions are difficult to access at room temperature because of the presence of high energy barriers. On the other hand, an accurate control of the effective temperature could be used to design microscopic heat engines. We now discuss in detail these two applications.

7.5.1 Sampling Energy Landscapes and Kramers Transitions

Sampling energy landscapes where high energy barriers are present and ergodicity is broken has attracted the interest of many disciplines in physics. In the recent years, single molecule experiments have allowed to explore the energy landscape of macromolecules, showing that the energetic difference between different conformations of RNA [17] or proteins [7, 21] can be of the order of $\sim 10 \text{ kT}$. For example, in Dietz et al. [9] measured the energy landscape of a GFP (Green Fluorescence Protein) using single molecule techniques, showing that barriers of $\simeq 20 \text{ kT}$ are present in the landscape. In the theoretical framework, different algorithms to sample such kind of energy landscapes have been introduced such as Montecarlo methods [13] or umbrella sampling [26].

Let us consider an overdamped Brownian particle in a double-well potential where the two wells are separated by a barrier of height $\Delta E \gg kT$. In equilibrium, the probability of the particle to jump from one well to the other is of the order of $e^{-\Delta E/kT}$ and therefore very small. If an external random force with associated effective temperature $T_{\text{kin}} > T$ is applied, the jumping probability increases to

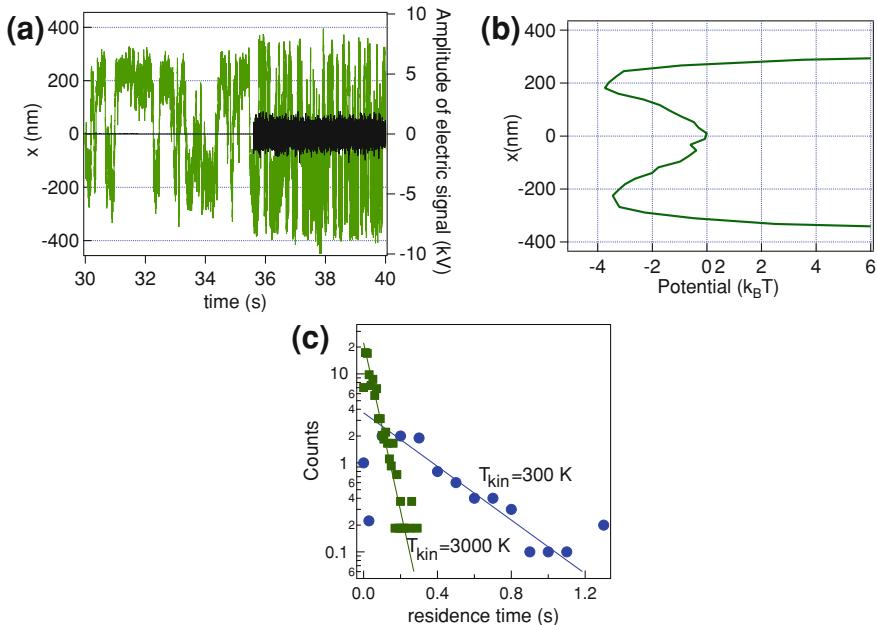


Fig. 7.10 Modification of Kramers transitions with our experimental setup **a** 10-s position traces (green, left axis) of the sphere in the double-well potential as a function of time and voltage on the electrodes (black, right axis) as a function of time. The additional noisy signal was switched on at $t_{\text{switch}} = 35.5$ s. **b** The trapping potential obtained at room temperature from the stationary distribution of the position before the noise was added. **c** Probability of the residence time of the Kramers transitions at room temperature (blue circles) and at 3000 K (green squares). The values of the amplitude of the additional force coincides with those used in (1) and (2) cases in Fig. 7.2

$e^{-\Delta E/kT_{\text{kin}}} > e^{-\Delta E/kT}$ and the Brownian particle can sample the energy landscape faster than in the absence of the external random force. This simple technique to sample high energy barriers can be implemented using our setup. This situation can be experimentally reproduced by trapping a Brownian particle with a dual optical tweezer. To observe thermally activated escape over a potential barrier, i.e. Kramers transitions [15], the optical traps of the dual-well trapping potential have to be close each other, and thermal energy has to be large enough to drive the sphere over the potential barrier between the two neighboring optical traps. The probability distribution of residence times, i.e. of the intervals of time between escape events from trap to trap depends on the height of the potential barrier and on the temperature of the system [15]. Experimental results at room temperature are well described by the Kramers theory [24].

We studied how the addition of the external random force affects to Kramers transitions of a microsphere trapped with a dual optical tweezer. We created a double-well potential with controllable distance between two traps and depths of the traps using a time-sharing protocol of the AOD. In Fig. 7.10a we show a 10 s frame of the

time traces of the $1\mu\text{m}$ diameter sphere moving in the double-well potential before and after the additional noise signal is switched on. When adding the external force, the transitions between the two traps are observed more frequently. In this situation, the potential created by the dual trap can be sampled faster. In Fig. 7.10b we show the effective potential generated by the dual trap, obtained as $U(x)/kT = -\ln \rho_{ss}(x)$ using the stationary probability distribution of the position before adding the external force, $\rho_{ss}(x)$. Figure 7.10c shows the residence time probability distribution without and with additional noise signals. As predicted by Kramers [15], the residence time probability decays exponentially with the residence time. Notice that in the presence of the additional noise, the probability distribution decays faster, which is consistent with an increase of the effective temperature.

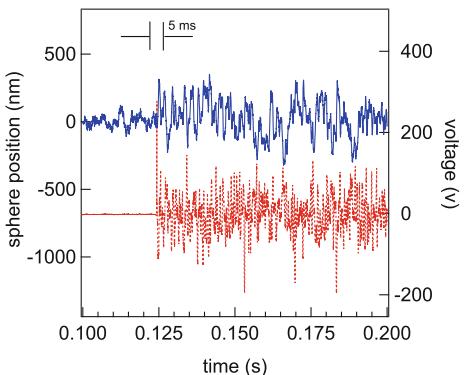
7.5.2 Building Microscopic Heat Engines

Since the origin of thermodynamics, heat engines, i.e. devices where a system operates between two thermal baths at different temperature, have been widely studied. The engine of a car, where fuel is in contact with a hot reservoir (the motor) and a cold reservoir (the surrounding air) is a real-life example of a heat engine that performs work. The construction of heat engines in the microscopic scale has been an open problem until recently. Blickle and Bechinger [4] were able to design the experimental version of Stirling engine, where a cyclic process is implemented with two isothermal processes at different temperatures and two isochoric processes. In their experiment, a microscopic sphere is trapped with an optical tweezer. The temperature is controlled using laser heating techniques and the stiffness of the trap plays the role of the inverse of the volume in original Stirling engine.

Such microscopic engines require an accurate control of the temperature to work in optimal conditions. Our experimental setup is a good candidate to be applied as a temperature control mechanism in a microscopic engine. For this purpose, our setup might be able to control the temperature of the system at fast timescales. We designed an experimental protocol to measure how fast one can switch the effective kinetic temperature of a Brownian particle. We studied the temporal response of the amplitude of the fluctuations of the position of the particle when we change the amplitude of the additional noise abruptly. Figure 7.11 shows that the standard deviation of the position of the sphere (and hence the kinetic temperature) changes with the same rate as the electric field amplitude, so that only the relaxation time of the trap ($\tau_r = \gamma/k = 1.4\text{ ms}$ in our experimental conditions) is the limiting factor.

Moreover, the broadening and compressing of the histograms of the position are reversible when we switch on and off the noisy electric field. This fact opens the possibility to implement thermodynamic cycles where a Brownian particle is immersed in a thermal bath of controllable temperature, and therefore to build microscopic heat engines in the spirit of the first microscopic heat engine designed by Blickle and Bechinger [4]. One could for example implement adiabatic processes by designing a protocol where $\rho(x)$, and therefore entropy of the system along a stochastic

Fig. 7.11 Standard deviation of the position of the *sphere* (left axis) and amplitude of electric signal (right axis) as a function of time at an abrupt change of the electric field



trajectory (1.63), $S(t) = -k \int dx \rho(x; t) \ln \rho(x; t)$, does not change in time. The density $\rho(x)$ can be kept constant in time with our setup by changing the stiffness of the trap and the intensity of the external random force using active feedback.

7.6 Conclusions

In this chapter, we have studied how one can tune the amplitude of the Brownian fluctuations of a microscopic sphere using an experimental setup that is similar to the one used in Chap. 6 to study the energetics of a symmetry breaking. We have studied the dynamics of a Brownian sphere in equilibrium and nonequilibrium situations and discussed the possible applications of our experimental setup.

Our experimental setup was described in Sect. 7.2. We trapped a Brownian microscopic charged sphere in water with an optical tweezer. An external random electric field whose spectrum is flat up to high frequencies is applied to the charged particle and plays the role of an external random force exerted to the microscopic particle. When the optical trap is held fixed and the external field is switched on, both position histograms and PSD suggest that the particle behaves like if it were immersed in a thermal bath at an effective temperature that is higher than room temperature (see Sect. 7.3).

We also studied in Sect. 7.4 the response of the particle to a nonequilibrium process in the presence of the external random field. When the field is switched on, Crooks fluctuation theorem is verified by replacing the room temperature by an effective temperature that is above room temperature. This temperature equals to the effective temperatures obtained from the PSD of the position histograms with the trap held fixed.

Our experiment allows one to control the effective temperature of a Brownian sphere within the range 300–3000 K, that is, we can reach effective temperatures that are 10 times greater than the room temperature. The temperature can be changed in a millisecond response time within the range 300–3000 K, which opens the possibility

to exploit our technique to design novel artificial micro and nanoengines, as discussed in Sect. 7.5. In addition, the setup can be used as a novel experimental technique to sample energy landscapes where ergodicity is broken because of the presence of high-energy barriers.

References

1. A. Ashkin, J. Dziedzic, *Appl. Phys. Lett.* **30**, 202–204 (1977)
2. A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, E. Lutz, *Nature* **483**, 187–189 (2012)
3. F. Beunis, F. Strubbe, K. Neyts, D. Petrov, *Phys. Rev. Lett.* **108**, 016101 (2012)
4. V. Blickle, C. Bechinger, *Nat. Phys.* **8**, 143–146 (2011)
5. G.E. Box, G.M. Jenkins, G.C. Reinsel, *Time Series Analysis: Forecasting and Control* (Wiley, Hoboken, 2011)
6. C. Bustamante, J. Liphardt, F. Ritort Farran, *Phys. Today* **58**(7), 43–48 (2005)
7. C. Cecconi, E.A. Shank, C. Bustamante, S. Marqusee, *Science* **309**, 2057–2060 (2005)
8. S. Ciliberto, S. Joubaud, A. Petrosyan, *J. Stat. Mech. Theory Exp.* **2010**, P12003 (2010)
9. H. Dietz, M. Rief, *Proc. Nat. Acad. Sci. USA* **101**, 16192–16197 (2004)
10. J. Gieseler, B. Deutscher, R. Quidant, L. Novotny, *Phys. Rev. Lett.* **109**, 103603 (2012)
11. C. Gosse, V. Croquette, *Biophys. J.* **82**, 3314–3329 (2002)
12. J.R. Gomez-Solano, L. Bellon, A. Petrosyan, S. Ciliberto, *EPL (Europhys. Lett.)* **89**, 60003 (2010)
13. W.K. Hastings, *Biometrika* **57**, 97–109 (1970)
14. C. Jarzynski, O. Mazonka, *Phys. Rev. E* **59**, 6448 (1999)
15. H.A. Kramers, *Physica* **7**, 284–304 (1940)
16. T. Li, S. Kheifets, M.G. Raizen, *Nat. Phys.* **7**, 527–530 (2011)
17. H. Ma, D.J. Proctor, E. Kierzek, R. Kierzek, P.C. Bevilacqua, M. Gruebele, *J. Am. Chem. Soc.* **128**, 1523–1530 (2006)
18. H. Mao, J. Ricardo, S.B. Arias-Gonzalez, I. Smith, J.R. Tinoco, C. Bustamante, *Biophys. J.* **89**, 1308–1316 (2005)
19. L.I. McCann, M. Dykman, B. Golding, *Nature* **402**, 785–787 (1999)
20. P. Mestres, D. Petrov, *Eur. Biophys. J.* **40**, 1081–1085 (2011)
21. J.N. Onuchic, Z. Luthey-Schulten, P.G. Wolynes, *Ann. Rev. Phys. Chem.* **48**, 545–600 (1997)
22. R. Radünz, D. Rings, K. Kroy, F. Cichos, *J. Phys. Chem. A* **113**, 1674–1677 (2009)
23. P. Ruijgrok, N. Verhart, P. Zijlstra, A. Tchebotareva, M. Orrit, *Phys. Rev. Lett.* **107**, 037401 (2011)
24. A. Simon, A. Libchaber, *Phys. Rev. Lett.* **68**, 3375–3378 (1992)
25. K. Svoboda, S.M. Block, *Ann. Rev. Biophys. Biomol. Struct.* **23**, 247–285 (1994)
26. G.M. Torrie, J.P. Valleau, *J. Comput. Phys.* **23**, 187–199 (1977)
27. K. Visscher, S.P. Gross, S.M. Block, *IEEE J. Sel. Top. Quantum Electron.* **2**, 1066–1076 (1996)
28. G. Volpe, S. Perrone, J.M. Rubi, D. Petrov, *Phys. Rev. E* **77**, 051107 (2008)

Part IV

Conclusions

Chapter 8

Conclusions and Outlook

In this thesis, we have studied the relationship between irreversibility and dissipation for microscopic systems that reach a nonequilibrium steady state (NESS), in the framework of stochastic thermodynamics. We have found a tool to estimate the dissipation of a microscopic system in the NESS from the statistics of a single stationary trajectory produced by the system. Our technique measures the Kullback-Leibler divergence (KLD) between the probability to observe a (*forward*) trajectory and the probability to observe the time reversed (or *backward*) trajectory, and it does not require to know any physical detail of the system to estimate the entropy produced by the system during the process. We have derived analytical expressions for the KLD for some specific stochastic processes and applied our technique to simulations as well as to experimental data acquired from a biological system. A specific nonequilibrium process, namely, a symmetry breaking, is also studied as a relevant example where the quantitative relationship between irreversibility and dissipation holds. We have found universal results concerning the energetics of a symmetry breaking at the microscopic level, and checked these results experimentally using optical tweezers. Our experimental setup has also allowed us to design a protocol to tune the kinetic temperature of Brownian particles. We now outline the main results obtained in this work:

1. **New estimators of the KLD rate.** We have introduced new estimators of the KLD rate between the probability to observe a trajectory and the probability to observe the corresponding time-reversal trajectory. For discrete trajectories, using statistics of blocks or sequences of m -data is the most feasible technique to estimate the KLD rate as long as there is sufficient statistics, as described in Sect. 3.1.1. When there is not sufficient statistics, i.e. when a finite-data sequence is not sampled in both forward and backward trajectories, the best strategy is to use our estimator based on compression algorithms defined in Sect. 3.1.2, which outperforms previous estimators of the KLD based on compression [6]. On the other hand, we have also introduced two different techniques to measure the KLD rate for real-valued or continuous time series. First, we have designed a novel estimator that transforms the time series into a directed graph whose

in-going and out-going degree distributions capture the irreversibility of a continuous time series (see Sect. 3.3). This method is applied to several stochastic processes that serve as benchmark cases, showing that it is possible to distinguish between reversible and irreversible stochastic processes as well as between conservative and dissipative chaotic processes. We introduced in Sect. 3.2.2 a second estimator of the KLD rate for continuous correlated time series which uses an asymmetric functional to produce a new uncorrelated series whose KLD is easier to calculate. Such functional is the residual of an auto regressive model. Unlike when using the visibility technique, the KLD estimator using the residual function can be closely related to the average dissipation of a microscopic system in the NESS, which is analyzed using biological data.

2. **Estimating dissipation from single stationary trajectories of discrete systems.** We have first applied our technique to a discrete flashing ratchet model (see Chap. 4). We have compared the average dissipation of the model with the KLD rate between trajectories of the ratchet that contain partial or full information of the system. We have shown that, even when having partial information of the system, our estimators are able to distinguish between equilibrium and NESS. When an external force stalls the ratchet, no net current of particles is observed and linear irreversible thermodynamics fail to predict that the system is dissipating energy. However, using our technique, we correctly predict that the system is producing entropy despite no net current is observed. In this simple example, the KLD, and not the current, is revealed as the actual fingerprint of irreversibility in microscopic systems.
3. **Detecting active processes in biological systems.** In Chap. 5, we have applied our technique to biological (continuous) data. With our KLD rate estimator based on the use of residuals we have been able to distinguish between active and passive spontaneous oscillations of ear hair bundles from bullfrog's sacculus. Our technique only requires the measurement of spontaneous oscillations of the position of the top of the hair bundles during ~ 100 s, outperforming significantly the requirements of previous works [3]. In addition, our KLD rate estimator indicates the minimum irreversibility to sustain spontaneous oscillations in bullfrog's ear hair bundles, which is of the order of the energy consumption rate from ATP hydrolyzation by a single adaptation motor. Therefore our method serves not only to detect the presence of active processes but also to bound the entropy produced by such processes when they are present.
4. **Universal features of the energetics of the symmetry breaking.** The energetics of a symmetry breaking and a symmetry restoration can be linked to the relationship between dissipation and irreversibility in the microscopic scale. When a microscopic system breaks a symmetry, the accessible phase space of the system is reduced to a subset of the full phase space. Even when the symmetry is broken quasi statically, the symmetry breaking is accompanied by a (negative) entropy production. We have derived a formula that relates the average entropy production when the system chooses an instance with the probability of the system to choose that option. Our formula is valid in the quasi static limit and it is universal: it does not depend on the mechanism that biases the probability to choose any option.

In addition, we have obtained an equivalent formula for the symmetry restoration process, which together with the symmetry breaking result, can be used to derive Landauer's principle and the energetics of the Szilard engine as discussed in Chap. 6. For a complete validation of our formula, we have checked our results experimentally, by trapping a Brownian particle in a dual optical tweezer and using external electric fields to tune the probability of the bead to choose between the two traps.

5. **Building experimental microscopic thermodynamic engines.** The development of experimental techniques in this thesis has provided us a tool to design a variety of thermodynamic processes in the microscopic scale. Using the dual optical trap, we have been able to reconstruct a Szilard engine as a concatenation of a symmetry breaking and a symmetry restoration, as discussed in Chap. 6. On the other hand, we have also designed an experiment to tune the amplitude of the Brownian fluctuations of a microscopic charged sphere. The setup is discussed in Chap. 7 and consists of a Brownian particle trapped by an optical tweezer and an external random electric field of tunable amplitude. With our setup, we are able to control the kinetic temperature of a microscopic particle from room temperature to 3000 K in a fast timescale (\sim ms). This technique opens the possibility of implementing heat engines in the microscopic scale such as a non-quasistatic Carnot engine.

In summary, we have performed a complete study on the relationship between dissipation and irreversibility for microscopic system in the NESS. Our study includes analytical results, simulations and experimental verifications. Our results provide new insights in stochastic thermodynamics and nonequilibrium statistical mechanics but also in time series analysis, information theory or biophysics. Apart from the aforementioned applications, our work opens the possibility to analyze new problems in stochastic thermodynamics. Here we indicate some of the possible future research lines related to our work.

1. **Application of our technique to different biological data.** Once our technique to estimate the dissipation in microscopic systems has been validated using data from bullfrog's ear hair cells, the method could be applied to other biological data. For example, one could analyze if the time asymmetry observed in the heartbeat of an aging heart is related to the dissipation rate of the heart [1]. Also, EEG recordings could be used to estimate the entropy produced in neural processes, or use the KLD to distinguish between different (active or passive) oscillations in red blood cells [2].
2. **Extension of our method to quantum systems.** All the results of this thesis are in the realm of classical physics and therefore their validity extends to small systems of size where quantum effects are negligible. The quantification of the relationship between irreversibility and dissipation for quantum system has not been considered yet. However, the extension of the KLD to quantum systems has already been discussed in [5].
3. **New experimental insights in stochastic thermodynamics.** Our experimental techniques can be applied to design new microscopic engines such as a

version of a Szilard engine simpler than the experiment done in Ref. [4]. The implementation of adiabatic processes in the microscopic scale—and therefore the construction of a Carnot engine—is also possible by using active feedback when tuning the kinetic temperature of a Brownian particle with our experimental setup described in Chap. 7. In addition, the dynamics of trapped Brownian particles under the action of random forces with arbitrary spectrum opens new questions that could be addressed with new experimental observations.

References

1. M. Costa, A.L. Goldberger, C.-K. Peng, Phys. Rev. Lett. **95**, 198102 (2005)
2. D.J. López, M. Egido-Gabas, I. López-Montero, J.V. Bustó, J. Casas, M. Garnier, F. Monroy, B. Larijani, F.M. Goñi, A. Alonso, Biophys. J. **102**, 2077 (2012)
3. P. Martin, A. Hudspeth, F. Jülicher, Proc. Natl. Acad. Sci. **98**, 14380–14385 (2001)
4. S. Toyabe, T. Sagawa, M. Ueda, E. Muneyuki, M. Sano, Nat. Phys. **6**, 988–992 (2010)
5. V. Vedral, Rev. Mod. Phys. **74**, 197 (2002)
6. J. Ziv, N. Merhav, IEEE Trans. Inf. Theory **39**, 1270–1279 (1993)

Appendix A

Appendix to Chapter 1

A.1 Gaussian White Noise

Let us consider the Brownian motion of a particle in one dimension x . We consider the *Wiener process*, $W(t)$, that is the continuous limit of the following discrete Brownian motion in one dimension

$$x(t + \Delta t) = x(t) + \eta_t \Delta x, \quad (\text{A.1})$$

where η_t is a random number that can take two values, $\eta_t = \pm 1$ each of them with probability $p = 1/2$, and Δx is the amplitude of the jumps of the particle. The initial condition is assumed to be $x(0) = 0$. Equation (A.1) models the motion of a Brownian particle in one dimension that starts in $x(0) = 0$ and jumps randomly to the left and to the right with equal probability. The Wiener process $W(t)$ is the limit of this Brownian motion when $\Delta x \rightarrow 0$, $\Delta t \rightarrow 0$ with $\sigma^2 = \Delta x^2 / \Delta t$ finite. The position of the Wiener process satisfies the following properties:

1. $W(t)$ is Gaussian distributed with zero mean and standard deviation $\sigma \sqrt{t}$, $W(t) \sim \mathcal{N}(0, \sigma \sqrt{t})$.
2. $W(t)$ is Markovian, i.e. the value of $W(t)$ is independent on the value of $W(t')$ for any $t' < t$.
3. The increment of $W(t)$ is independent on its value, i.e. $W(t + \Delta t) - W(t)$ is independent of $W(t)$.

We can investigate which is the value of the correlation of the Wiener process $\langle W(t)W(t') \rangle$ by taking into account the above properties. Because of the 3rd property,

$$\langle [W(t + \Delta t) - W(t)]W(t) \rangle = 0, \quad (\text{A.2})$$

or equivalently

$$\langle W(t + \Delta t)W(t) \rangle = \langle W(t)W(t) \rangle = \sigma^2 t, \quad (\text{A.3})$$

where we have used the 1st property in the second equality. For any t and t' ,

$$\langle W(t)W(t') \rangle = \sigma^2 \min(t, t'). \quad (\text{A.4})$$

Gaussian white noise is formally defined as the derivative of a Wiener process

$$\xi(t) = \frac{dW(t)}{dt}. \quad (\text{A.5})$$

However, since $W(t)$ is stochastic, its time derivative cannot be computed, and the above definition is not useful for generating a stochastic series of a Gaussian white noise process. However, since $\xi(t)$ is obtained from a Gaussian process, $\xi(t)$ is also a Gaussian random variable. Its average and correlation can be computed from the properties of the Wiener process. Since $\langle W(t) \rangle = 0$,

$$\langle \xi(t) \rangle = \frac{d}{dt} \langle W(t) \rangle = 0. \quad (\text{A.6})$$

The correlation of the white noise can be obtained taking into account (A.4)

$$\langle \xi(t)\xi(t') \rangle = \frac{d}{dt} \frac{d}{dt'} \langle W(t)W(t') \rangle = \frac{d}{dt} \sigma^2 \theta(t' - t) = \sigma^2 \delta(t - t'). \quad (\text{A.7})$$

Therefore, a Gaussian white noise of amplitude σ is a stochastic process defined by the following three properties

1. It is Gaussian distributed.
2. It has zero mean $\langle \xi(t) \rangle = 0$.
3. Its correlation is such that $\langle \xi(t)\xi(t') \rangle = \sigma^2 \delta(t - t')$.

A.2 Stochastic Differential Equations

Let us consider stochastic differential equations of the form

$$\dot{x}(t) = f(x(t)) + g(x(t))\xi(t), \quad (\text{A.8})$$

for any generic functions f and g , and $\xi(t)$ being a Gaussian white noise (see Appendix A.1). The solution of these equations have the form

$$x(t) = x(0) + \int_0^t f(x(t'))dt' + \int_0^t g(x(t'))dW(t'). \quad (\text{A.9})$$

Because of the stochastic nature of $W(t)$, the calculation of the last term of (A.9) is cumbersome. We have to give a prescription to the integral

$$\int_0^t g(x(t')) dW(t') = \lim_{\delta t \rightarrow 0} \sum_i g(t_i^*) [W(t_i + \Delta t) - W(t_i)]. \quad (\text{A.10})$$

There are two prescriptions commonly used to calculate this integral introduced by Ito and Stratonovich:

- **Ito:** $t_i^* = t_i$.
- **Stratonovich:** $t_i^* = t_i + \frac{\Delta t}{2}$.

Therefore, the integrals of the type $\int g(x(t')) dW(t')$ can be evaluated according to these two different conventions. In general, the following notation [1] is used: For Ito calculus, (A.10) is evaluated using the following discretization scheme,

$$g(x(t')) \cdot dW(t') \rightarrow g(x(t')) [W(t' + \Delta t) - W(t')], \quad (\text{A.11})$$

whereas for Stratonovich calculus

$$g(x(t')) \circ dW(t') \rightarrow \frac{g(x(t' + \Delta t)) + g(x(t'))}{2} [W(t' + \Delta t) - W(t')]. \quad (\text{A.12})$$

Depending on the prescription used, the values of integrals that involve stochastic terms may differ. In the case in which the noise is *additive*, that is, when the term that multiplies $\xi(t)$ does not depend on $x(t)$ ($g(x(t)) = \text{const}$), Ito and Stratonovich calculus yield the same result. In any other case, the two prescriptions may lead to different values. As an example, let us consider the following differential equations

$$\dot{x}(t) = \xi(t), \quad (\text{A.13})$$

$$\dot{y}(t) = x(t)\xi(t). \quad (\text{A.14})$$

The solution to this equation is (see Appendix A.1)

$$y(t) = \int_0^t W(t') dW(t'). \quad (\text{A.15})$$

We now analyze which is the value of the ensemble average of $y(t)$ to different realizations. In Ito formalism,

$$\langle y^{\text{Ito}}(t) \rangle = \lim_{\Delta t \rightarrow 0} \sum_i \langle W(t_i) [W(t_i + \Delta t) - W(t_i)] \rangle = 0, \quad (\text{A.16})$$

where we have used (A.2). On the other hand, in the Stratonovich prescription,

$$\begin{aligned}\langle y^{\text{Strat}}(t) \rangle &= \lim_{\Delta t \rightarrow 0} \sum_i \langle W(t_i + \Delta t/2)[W(t_i + \Delta t) - W(t_i)] \rangle \\ &= \lim_{\Delta t \rightarrow 0} \sum_i \sigma^2[t_i + \Delta t/2 - t_i] = \frac{\sigma^2 t}{2}.\end{aligned}\quad (\text{A.17})$$

Notice that in Stratonovich calculus usual rules of calculus do apply. In this example,

$$y^{\text{Strat}}(t) = \int_0^t W(t') dW(t') = \frac{W(t')^2}{2} \Big|_0^t = \frac{W(t)^2}{2}. \quad (\text{A.18})$$

From the above (exact) solution we can calculate the average $\langle y^{\text{Strat}}(t) \rangle = \langle W(t)^2/2 \rangle = \sigma^2 t/2$, recovering (A.17). In Ito calculus, usual rules of calculus do not apply. This can be seen intuitively in this example. The term,

$$\begin{aligned}\langle W(t_i + \Delta t)[W(t_i + \Delta t) - W(t_i)] \rangle &= \frac{[W(t_i + \Delta t)^2 - W(t_i)^2]}{2} \\ &\quad - \frac{[W(t_i + \Delta t) - W(t_i)]^2}{2}.\end{aligned}$$

The first term leads to the result from standard calculus $W(t)^2/2$ and the second vanishes if $W(t)$ is differentiable since $[W(t_i + \Delta t) - W(t_i)]^2 \sim \Delta t^2$. However, for the Wiener process, $[W(t_i + \Delta t) - W(t_i)]^2 = \sigma^2 \Delta t$, and the contribution of the second term cannot be neglected. In Ito calculus, the terms that go like ΔW^2 are replaced by $\sigma^2 \Delta t$. As a consequence, the following property, known as *Ito's lemma*, holds in Ito's convention: If $x(t)$ follows (A.8), and $A = A(x)$ is any function of $x = x(t)$, then

$$\dot{A}(t) = \frac{\partial A}{\partial x} [f(x(t)) + g(x(t))\xi(t)] + \frac{\sigma^2}{2} \frac{\partial^2 A}{\partial x^2} g(x(t))^2. \quad (\text{A.19})$$

Another paradigmatic case study in which Ito and Stratonovich calculus do not coincide is $\langle x(t)\xi(t) \rangle$. In Ito's formalism

$$\langle x(t)\xi(t) \rangle^{\text{Ito}} = 0, \quad (\text{A.20})$$

whereas in the Stratonovich prescription, $\langle x(t)\xi(t) \rangle$ does not vanish in general and it is calculated using Novikov's theorem [2],

$$\langle x(t)\xi(t) \rangle^{\text{Strat}} = \int_0^t ds \langle \xi(t)\xi(s) \rangle \left\langle \frac{\delta x(t)}{\delta \xi(s)} \right\rangle, \quad (\text{A.21})$$

where $\frac{\delta x(t)}{\delta \xi(s)}$ stands for functional the functional derivative of $x(t)$ with respect to $\xi(s)$.

A.3 Fokker-Planck Equation

Let us consider the following generalized overdamped Langevin equation

$$\dot{x} = f(x) + g(x) \cdot \xi, \quad (\text{A.22})$$

where the stochastic product is taken in Ito's sense. By Ito's lemma, any function $A = A(x)$ follows

$$\dot{A} = \frac{\partial A}{\partial x} [f(x) + g(x)\xi(t)] + \frac{\sigma^2}{2} \frac{\partial^2 A}{\partial x^2} g(x)^2. \quad (\text{A.23})$$

We now focus on calculating the average of $\langle \dot{A} \rangle$. We notice that in Ito's convention $\langle g(x)\xi \rangle = 0$ for any function g (A.20). Therefore,

$$\langle \dot{A}(t) \rangle = \left\langle A'(x)f(x) + \frac{\sigma^2}{2}g(x)^2A''(x) \right\rangle, \quad (\text{A.24})$$

where we have simplified the notation by using $x = x(t)$ and $A'(x) = \frac{\partial A}{\partial x}$. Any average can be expressed in terms of the probability density function $\rho(x, t)$, $\langle A \rangle = \int dx \rho(x, t)A$, where the integration is done over all the space. Taking into account this, and using the notation $\partial_\alpha = \frac{\partial}{\partial \alpha}$ and $\partial_\alpha^2 = \frac{\partial^2}{\partial \alpha^2}$ for any variable α , we can rewrite (A.24)

$$\int A(x) \frac{\partial}{\partial t} \rho(x, t) = \int dx \rho(x, t) \left[A'(x)f(x) + \frac{\sigma^2}{2}g(x)A''(x) \right]. \quad (\text{A.25})$$

Now, integrating by parts and assuming that $\rho(x, t)$ vanishes in $x = \pm\infty$ at any time t ,

$$\int A(x) \frac{\partial}{\partial t} \rho(x, t) = \int dx A(x) \left[-\partial_x(f(x)\rho(x, t)) + \frac{\sigma^2}{2} \partial_x^2(g(x)\rho(x, t)) \right]. \quad (\text{A.26})$$

Since Ito's lemma holds for any function $A(x)$, the above equality implies

$$\partial_t \rho(x, t) = -\partial_x \left[f(x)\rho(x, t) \right] + \frac{\sigma^2}{2} \partial_x^2 \left[g(x)^2 \rho(x, t) \right]. \quad (\text{A.27})$$

which is known as the Fokker-Planck equation for the probability density $\rho(x, t)$. Fokker-Planck equation can be seen as a continuity equation for the probability current $j(x, t)$

$$\partial_t \rho(x, t) = -\partial_x j(x, t), \quad (\text{A.28})$$

with

$$j^{\text{Ito}}(x, t) = \left[f(x) - \frac{\sigma^2}{2} \partial_x g(x)^2 \right] \rho(x, t). \quad (\text{A.29})$$

In the case of Stratonovich calculus, the definition of the probability current changes slightly,

$$j^{\text{Strat}}(x, t) = \left[f(x) - \sigma^2 g(x) \partial_x g(x) \right] \rho(x, t). \quad (\text{A.30})$$

For additive noise, both conventions lead to the same result. If $g(x) = C$ with $C = \text{const}$,

$$j^{\text{Ito}}(x, t) = j^{\text{Strat}}(x, t) = \left[f(x) - \sigma^2 C^2 \right] \rho(x, t). \quad (\text{A.31})$$

References

1. K. Sekimoto, *Stochastic Energetics*, vol. 799 (Springer, Berlin, 2010)
2. E.A. Novikov, Sov. Phys. J. Exp. Theor. Phys. **20**, 1290–1294 (1965)

Appendix B

Appendix to Chapter 2

B.1 Shannon Entropy is Time Invariant Under Liouville Evolution

We now prove that Shannon entropy of an isolated system, in units of k ,

$$S(t) = - \int dqdp \rho(q, p; t) \ln \rho(q, p; t), \quad (\text{B.1})$$

does not change in time, i.e. $\partial_t S(t) = 0$, along an isolated process where the phase space density evolves according to Liouville's equation (1.28)

$$\frac{\partial \rho(z, t)}{\partial t} = \mathcal{L}\rho(z, t), \quad (\text{B.2})$$

where $\mathcal{L}\rho = \{H, \rho\} = \partial_q H \partial_p \rho - \partial_p H \partial_q \rho$, H being the Hamiltonian of the system.

The explicit time derivative of the Shannon entropy yields

$$\begin{aligned} \partial_t S(t) &= - \int dqdp \partial_t \rho(q, p; t) \ln \rho(q, p; t) - \int dqdp \partial_t \rho(q, p; t) \\ &= - \int dqdp \partial_t \rho(q, p; t) \ln \rho(q, p; t), \end{aligned} \quad (\text{B.3})$$

since $\int dqdp \partial_t \rho(q, p; t) = \partial_t \int dqdp \rho(q, p; t) = \partial_t 1 = 0$. We now use Liouville's theorem (B.2) in (B.3),

$$\begin{aligned} \partial_t S(t) &= - \int dqdp \mathcal{L}\rho(q, p; t) \ln \rho(q, p; t) \\ &= - \underbrace{\int dqdp \partial_q H \partial_p \rho \ln \rho}_{\dot{S}_1} + \underbrace{\int dqdp \partial_p H \partial_q \rho \ln \rho}_{\dot{S}_2}. \end{aligned} \quad (\text{B.4})$$

Integrating by parts the first term in (B.5),

$$\begin{aligned}\dot{S}_1 &= - \int dq \left[\partial_q H \rho \ln \rho \Big|_{-\infty}^{\infty} - \int dp \partial_p \partial_q H \rho \ln \rho - \int dp \partial_q H \partial_p \rho \right] \\ &= \int dq dp [\partial_p \partial_q H \rho \ln \rho + \partial_q H \partial_p \rho],\end{aligned}\quad (\text{B.5})$$

where we have assumed that ρ has finite support and it vanishes in $p = \pm\infty$ as does $\rho \ln \rho$ using the convention $0 \ln 0 = 0$. The second term in (B.5) can be obtained by adding a minus sign and changing $q \rightarrow p$, and vice versa, in (B.5),

$$\dot{S}_2 = - \int dq dp [\partial_q \partial_p H \rho \ln \rho + \partial_p H \partial_q \rho]. \quad (\text{B.6})$$

If we assume that $\partial_q \partial_p H = \partial_p \partial_q H$ and we replace (B.5) and (B.6) in (B.5), we obtain

$$\partial_t S(t) = \int dq dp [\partial_q H \partial_p \rho - \partial_p H \partial_q \rho] \quad (\text{B.7})$$

$$= \int dq dp \mathcal{L} \rho = \int dq dp \partial_t \rho = 0, \quad (\text{B.8})$$

which completes the proof.

B.2 Properties of the Kullback–Leibler Divergence

The Kullback–Leibler divergence or relative entropy between two probability distributions of a random continuous variable X is defined by

$$D[p(x)||q(x)] = \int_{\mathcal{X}} dx p(x) \ln \frac{p(x)}{q(x)}, \quad (\text{B.9})$$

where the integral runs over all the values that the value X can take, \mathcal{X} . We use the conventions $0 \ln 0 = 0$, $0 \ln \frac{0}{q} = 0$ and $p \ln \frac{p}{0} = \infty$. The latter case implies that if there is a value x of the variable X such that $p(x) \neq 0$ and $q(x) = 0$, the KLD diverges. The relative entropy $D[p(x)||q(x)]$ is only finite if the support set of p is contained in the support set of q .

Notice that the definition of KLD can be extended to distribution of any number n of random variables. We define by X_1, X_2, \dots, X_n a set of n different random variables, which can be simplified using the notation $X_1^n \equiv X_1, \dots, X_n$. In this case, we denote by $p(x_1^n) dx_1^n = p(x_1, \dots, x_n) dx_1 \cdots dx_n$ the probability to observe the n random variable to take the values $[x_1, x_1 + dx_1], \dots, [x_n, x_n + dx_n]$. The KLD between two probability distributions of the set of random variables X_1^n is defined as

$$D[p(x_1^n)||q(x_1^n)] = \int_{\mathcal{X}_1^n} dx_1^n p(x_1^n) \ln \frac{p(x_1^n)}{q(x_1^n)}. \quad (\text{B.10})$$

In case of a single random variable of many dimensions such as $\mathbf{x} = (x, y, z)$, the KLD between two distributions of such a variable can be expressed using (B.10) as the KLD between the joint probability distributions of the components, $p(x, y, z)$ and $q(x, y, z)$ in this example. We now introduce some of the most relevant properties of the KLD that are valid for both single and multivariate probability density functions of discrete and continuous random variables.

We now consider the discrete version of (B.9) to prove the main properties of the KLD [1]. For a discrete random variable X which can take values in a set $X = \{x \in \mathcal{X}\}$, the KLD between two distributions of the variable X , p and q is

$$D[p(x)||q(x)] = \sum_{x \in \mathcal{X}} p(x) \ln \frac{p(x)}{q(x)}. \quad (\text{B.11})$$

We now show and prove some properties of the KLD between two probability distributions of discrete variables that are extensive to the KLD between distributions of continuous random variables as well as to the KLD between probability distributions of more than one variable.

Positivity: For any p and q ,

$$D[p(x)||q(x)] \geq 0, \quad (\text{B.12})$$

with equality if and only if $p = q$ for all x .

Proof Let us consider $-D[p(x)||q(x)]$ where we denote by \mathcal{X} the support of X and by A the support set of p , which is contained in \mathcal{X} , $A = \{x|p(x) > 0\}$. The following inequalities hold

$$-D[p(x)||q(x)] = \sum_{x \in A} p(x) \ln \frac{q(x)}{p(x)} \quad (\text{B.13})$$

$$\leq \ln \sum_{x \in A} p(x) \frac{q(x)}{p(x)} = \ln \sum_{x \in A} q(x) \quad (\text{B.14})$$

$$\leq \ln \sum_{x \in \mathcal{X}} q(x) = \ln 1 = 0. \quad (\text{B.15})$$

Since $-D[p(x)||q(x)] \leq 0 \Rightarrow D[p(x)||q(x)] \geq 0$. Notice that the inequality in (B.14) is consequence of Jensen's inequality for the (concave) function $\ln(x)$, which satisfies $\langle \ln x \rangle \leq \ln \langle x \rangle$ for any value of the argument x .

Chain rule: Kullback–Leibler divergence decreases upon coarse graining,

$$D[p(x, y) \parallel q(x, y)] \geq D[p(x) \parallel q(x)]. \quad (\text{B.16})$$

Proof

$$\begin{aligned} D[p(x, y) \parallel q(x, y)] &= \sum_{x,y} p(x, y) \ln \frac{p(x, y)}{q(x, y)}, \\ &= \sum_{x,y} p(x, y) \ln \frac{p(x)p(y|x)}{q(x)q(y|x)}, \\ &= \sum_{x,y} p(x, y) \ln \frac{p(x)}{q(x)} + \sum_{x,y} p(x, y) \ln \frac{p(y|x)}{q(y|x)}, \\ &= D[p(x) \parallel q(x)] + D[p(y|x) \parallel q(y|x)], \\ &\geq D[p(x) \parallel q(x)]. \end{aligned} \quad (\text{B.17})$$

The inequality in (B.17) is because the KLD $D[p(y|x) \parallel q(y|x)]$ is positive.

The properties that are mentioned above apply not only to distributions of independent random variables but also to *dependent* random variables. A paradigmatic case of dependent random variables is to consider the outcomes of a stochastic process with memory. The KLD can be extended to distributions of stochastic trajectories of a random variable X . The extension is done in a similar way as in the definition of the KLD between probability distributions of n different random variables. Let us consider a finite sequence of n symbols x_1, x_2, \dots, x_n which we denote by x_1^n . The extension is done. Let $p(x_1^n) = p(x_1, x_2, \dots, x_n)$ be a joint probability density of observing in the stochastic process n of the continuous random variable X to take the values x_1, x_2, \dots, x_n respectively. Let us consider another stochastic process of the same continuous random variable X drawn according to a different probability distribution q , and let $q(x_1^n) = q(x_1, x_2, \dots, x_n)$ be the joint probability distribution to observe the sequence x_1, x_2, \dots, x_n in the stochastic process drawn according to q . The KLD between the two probability distributions is expressed in the same way as the KLD between multivariate random variables (B.10),

$$D[p(x_1^n) \parallel q(x_1^n)] = \int dx_1^n p(x_1^n) \ln \frac{p(x_1^n)}{q(x_1^n)}, \quad (\text{B.18})$$

In [2] we introduced the following compact notation

$$D_n^X[p \parallel q] \equiv D[p(x_1^n) \parallel q(x_1^n)], \quad (\text{B.19})$$

where the subscript n stands for the number of data sampled to distinguish between p and q and the superscript X denotes the variable that is distributed according to p and q . D_n^X is known as the n -time relative entropy between the distributions p and q .

of the random variable X . In case of stochastic process it is also interesting to define the KLD *rate* per data,

$$d^X[p||q] = \lim_{n \rightarrow \infty} \frac{D_n^X[p||q]}{n}, \quad (\text{B.20})$$

which saturates when n is of order of the memory of the stochastic process.

B.3 Kullback–Leibler Divergence Between Hidden Markov Chains Using Replica Trick

B.3.1 Replica Trick

Replica trick is a mathematical technique that was introduced to calculate free energies in spin glasses [3]. Consider a Hamiltonian that describes the interaction of N subsystems which are coupled by random strengths $\sigma_{i,j}$ that are distributed according to a specific probability distribution. One example is a spin glass with random coupling strengths between every pair of spins, such as the Sherrington–Kirkpatrick model [4] where coupling strengths are random and Gaussian distributed, $p(\sigma_{i,j}) \propto \exp(-\sigma_{i,j}^2/2)$. For such a system, it is interesting to calculate the free energy per subsystem. Let Z be the partition function of the N -spin system. The average free energy per site over all the possible values of the coupling, f , is given by

$$f = -k_B T \lim_{N \rightarrow \infty} \frac{1}{N} \langle \ln Z \rangle. \quad (\text{B.21})$$

The average $\langle \ln Z \rangle$ over all possible configurations for these kind of models cannot be calculated analytically in general. However the following trick

$$\langle \ln Z \rangle = \lim_{n \rightarrow 0} \frac{d}{dn} \ln \langle Z^n \rangle, \quad (\text{B.22})$$

simplifies the calculation of (B.21) since $\langle Z^n \rangle$ can be calculated explicitly in general. By replacing (B.22) into (B.21),

$$f = -k_B T \lim_{N \rightarrow \infty} \left(\lim_{n \rightarrow 0} \frac{d}{dn} \phi_N(n) \right), \quad (\text{B.23})$$

where

$$\phi_N(n) = \frac{1}{N} \ln \langle Z^n \rangle, \quad (\text{B.24})$$

can be considered as the free energy per site of a system formed by n replica of the original system. Because of this, the technique to replace the average $\langle \ln Z \rangle$ by the right hand side of Eq. (B.22) is often called *replica trick*.

B.3.2 KLD of Hidden Markov Chains Using Replica Trick

The semi-analytical calculation of the KLD rate for a specific case of hidden Markov chains was discussed in Sect. 2.5.2. We now introduce a new technique to calculate the KLD d^X in Eq. (2.46) using the replica trick [2, 5]. To this end, we first consider the expression of d^X in terms of Shannon and cross entropy rates, $d^X = h_r^X - h^X$. We define the matrix resulting from the multiplication of m transition matrices [defined in Eq. (2.50)] chosen according to x_1^m as

$$\mathbf{T}(x_1^m) = \prod_{i=1}^{m-1} \mathbf{T}(x_i, x_{i+1}), \quad (\text{B.25})$$

Shannon entropy rate h^X can be rewritten by substituting (B.25) into (2.51),

$$h^X = - \lim_{m \rightarrow \infty} \frac{1}{m} \langle \ln \text{Tr} \mathbf{T}(x_1^m) \rangle. \quad (\text{B.26})$$

The analytical calculation of the average $\langle \ln \text{Tr} \mathbf{T}(x_1^m) \rangle$ is cumbersome and it can only be done semi-analytically, as we explained in Sect. 2.5.2. However, we can express this average in terms of $\langle \text{Tr} \mathbf{T}(x_1^m) \rangle$, which can be calculated analytically by using the replica trick [cf. Eq. (B.22)]

$$\langle \ln \text{Tr} \mathbf{T}(x_1^m) \rangle = \lim_{\alpha \rightarrow 0} \frac{d}{d\alpha} \ln \langle [\text{Tr} \mathbf{T}(x_1^m)]^\alpha \rangle. \quad (\text{B.27})$$

Reference [6] shows how to apply this technique when $\mathbf{T}(x_1^m)$ is equal to a product of random matrices which are chosen following a Markovian process. In our case, an underlying Markovian process defined by the two random variables X and Y , defines the order of the matrices that are multiplied in $\mathbf{T}(x_1^m)$. We now apply the technique described in [6] to calculate h^X . If we define the *generalized Lyapunov exponent* of degree α [7] as

$$L_\alpha^X = \lim_{m \rightarrow \infty} \frac{1}{m} \ln \langle [\text{Tr} \mathbf{T}(x_1^m)]^\alpha \rangle, \quad (\text{B.28})$$

which is analogous to the free energy per site described in (B.24). With this definition and taking into account the replica trick for this problem (B.27), Shannon entropy rate (B.26) is given by

$$h^X = - \lim_{\alpha \rightarrow 0} \frac{d}{d\alpha} L_\alpha^X. \quad (\text{B.29})$$

Now we consider the following property: Given a matrix \mathbf{A} and a positive integer α , $(\text{Tr}\mathbf{A})^\alpha = \text{Tr}(\mathbf{A}^{\otimes\alpha})$, where $\mathbf{A}^{\otimes\alpha} = \underbrace{\mathbf{A} \otimes \mathbf{A} \otimes \cdots \otimes \mathbf{A}}_{\alpha \text{ times}}$. Using this property, the average in (B.28) reads

$$\langle [\text{Tr}\mathbf{T}(x_1^m)]^\alpha \rangle = \langle \text{Tr}[\mathbf{T}(x_1^m)^{\otimes\alpha}] \rangle = \text{Tr} \langle \mathbf{T}(x_1^m)^{\otimes\alpha} \rangle. \quad (\text{B.30})$$

Since the tensor power of a product of matrices factorizes, $(\mathbf{ABC})^{\otimes\alpha} = \mathbf{A}^{\otimes\alpha} \mathbf{B}^{\otimes\alpha} \mathbf{C}^{\otimes\alpha}$, Eq. (B.30) can be rewritten,

$$\langle [\text{Tr}\mathbf{T}(x_1^m)]^\alpha \rangle = \text{Tr} \sum_{x_1^m, y_1^m} \prod_{i=1}^{m-1} \mathbf{T}(x_i, x_{i+1})_{y_i, y_{i+1}} \mathbf{T}(x_i, x_{i+1})^{\otimes\alpha}. \quad (\text{B.31})$$

We now define a block matrix $\mathcal{T}(\alpha)$, where each block is a transition matrix $\mathbf{T}(x_1, x_2)^{\otimes\alpha+1}$. The matrix elements of $\mathcal{T}(\alpha)$ are therefore:

$$\mathcal{T}(\alpha)_{x_1, y_1, x_2, y_2} = [\mathbf{T}(x_1, x_2)^{\otimes\alpha+1}]_{y_1, y_2}. \quad (\text{B.32})$$

Using (B.31) and (B.32) in (B.28), we see that L_α^X is dominated by the largest eigenvalue of $\mathcal{T}(\alpha)$ which we call $\tau(\alpha)$,

$$L_\alpha^X = \lim_{m \rightarrow \infty} \frac{1}{m} \ln \text{Tr}[\mathcal{T}(\alpha)^{m-1}] = \ln \tau(\alpha), \quad (\text{B.33})$$

yielding,

$$h^X = - \lim_{\alpha \rightarrow 0} \frac{d}{d\alpha} \ln \tau(\alpha). \quad (\text{B.34})$$

The above limit cannot be calculated analytically because the tensor powers in $\mathcal{T}(\alpha)$ are only defined for integer values of α . Therefore we approximate the limit $\alpha \rightarrow 0$ by an estimation of the slope of L_α^X as a function of α close to $\alpha = 0$ [7],

$$\hat{h}^X = 2L_1^X - \frac{L_2^X}{2} = 2 \ln \tau(1) - \frac{\ln \tau(2)}{2}. \quad (\text{B.35})$$

We obtain an equivalent result for h_r^X by replacing $\mathbf{T}(x_1^m)$ in (B.26) by the product of transition matrices but ordered according to the time-reversed series $x_m^1, \mathbf{T}(x_m^1)$. Defining the following matrix

$$\mathcal{T}_r(\alpha)_{x_1, y_1, x_2, y_2} = [\mathbf{T}(x_2, x_1)^T \otimes \mathbf{T}(x_1, x_2)^{\otimes\alpha}]_{y_1, y_2}, \quad (\text{B.36})$$

and $\tau_r(\alpha)$ being the largest eigenvalue of $\mathcal{T}_r(\alpha)$, we get

$$h_r^X = - \lim_{\alpha \rightarrow 0} \frac{d}{d\alpha} \ln \tau_r(\alpha). \quad (\text{B.37})$$

In practice, we also need to approximate the limit $\alpha \rightarrow 0$ in the above expression using Eq. (B.35) but replacing τ by τ_r ,

$$\hat{h}_r^X = 2 \ln \tau_r(1) - \frac{\ln \tau_r(2)}{2}. \quad (\text{B.38})$$

Finally, the estimation of d^X for this kind of series using replica trick, is obtained with the difference between Eqs. (B.38) and (B.35),

$$\hat{d}^X = \hat{h}_r^X - \hat{h}^X = 2 \ln \frac{\tau_r(1)}{\tau(1)} + \frac{1}{2} \ln \frac{\tau(2)}{\tau_r(2)}. \quad (\text{B.39})$$

References

1. T.M. Cover, J.A. Thomas, *Elements of Information Theory* (Wiley, New York, 2006)
2. É. Roldán, J.M.R. Parrondo, Phys. Rev. E **85**, 031129 (2012)
3. J. Van Hemmen, R. Palmer, J. Phys. A Math. Gen. **12**, 563 (1979)
4. D. Sherrington, S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975)
5. É. Roldán, J.M.R. Parrondo, Phys. Rev. Lett. **105**, 150607 (2010)
6. M. De Oliveira, A. Petri, Phys. Rev. E **53**, 2960 (1996)
7. A. Crisanti, G. Paladin, A. Vulpiani, *Products of Random Matrices in Statistical Physics (Springer Series in Solid State Sciences)* (Springer, Berlin, 1993)

Appendix C

Appendix to Chapter 3

C.1 Outgoing Degree Distribution in the Visibility Graph of Uncorrelated Random Series

We now proof Eq. (3.34) for real valued random uncorrelated series. We study the value of the *out* degree distribution, since the analysis for the *in* distribution is completely equivalent. Let x be an arbitrary data of the aforementioned series. The probability that the horizontal visibility of x is interrupted by a data x_r on its right is independent of the distribution that generates the series, $\rho(x)$,

$$\Phi_1 = \int_{-\infty}^{\infty} \int_x^{\infty} \rho(x)\rho(x_r)dx_r dx = \int_{-\infty}^{\infty} \rho(x)[1 - F(x)]dx = \frac{1}{2}, \quad (\text{C.1})$$

$$\text{where } F(x) = \int_{-\infty}^x \rho(x')dx'.$$

The probability $p_{\text{out}}(k)$ of the data x being capable of exactly seeing k data can be expressed as

$$p_{\text{out}}(k) = q_{\text{out}}(k)\Phi_1 = \frac{1}{2}q_{\text{out}}(k), \quad (\text{C.2})$$

where $q_{\text{out}}(k)$ is the probability of x seeing at least k data. $q_{\text{out}}(k)$ can be recurrently calculated via

$$q_{\text{out}}(k) = q_{\text{out}}(k-1)(1 - \Phi_1) = \frac{1}{2}q_{\text{out}}(k-1). \quad (\text{C.3})$$

Since the first data at least sees the second data, $q_{\text{out}}(1) = 1$, the following expression is obtained

$$q_{\text{out}}(k) = \left(\frac{1}{2}\right)^{k-1}, \quad (\text{C.4})$$

which together with Eq. (C.2) concludes the proof of Eq. (3.34),

$$P_{\text{out}}(k) = \left(\frac{1}{2}\right)^k. \quad (\text{C.5})$$

We notice that an analogous derivation holds for the *in* case.

C.2 KLD Visibility Estimator for a Logistic Map

In this section, we prove analytically that the time series generated by the logistic map (3.35) with $\mu = 4$ are irreversible at the level of the *in* and *out* degree distributions of their associated DHVg, that is, $d_{\text{vis}} > 0$. We notice that the KLD between two distributions is zero if and only if the distributions are the same in the entire support. Therefore, if we want to prove that d_{vis} is strictly positive, it is sufficient to find that $P_{\text{in}}(k) \neq P_{\text{out}}(k)$ for some value of the degree k . Here we take advantage of this fact to provide a rather general recipe to prove that a chaotic system is irreversible.

Consider a time series x_1^n with a joint probability distribution $\rho(x_1, x_2, \dots, x_n)$ and support (a, b) , and denote x_{i-1}, x_i, x_{i+1} three (ordered) generic data of the series. By construction,

$$\begin{aligned} P_{\text{out}}(k=1) &= P(x_i \leq x_{i+1}) = \int_a^b dx_i \int_{x_i}^b dx_{i+1} \rho(x_i, x_{i+1}), \\ P_{\text{in}}(k=1) &= P(x_{i-1} > x_i) = \int_{x_i}^b dx_{i-1} \int_a^b dx_i \rho(x_{i-1}, x_i). \end{aligned} \quad (\text{C.6})$$

The probability that $k_{\text{out}} = 1$ ($k_{\text{in}} = 1$) is actually the probability that the series increases (decreases) in one step. This probability is independent of time, because we consider stationary series. If the chaotic map is of the form $x_{i+1} = F(x_i)$, it is Markovian, and the preceding equations simplify due to Markov property (2.36):

$$\begin{aligned} P_{\text{out}}(k=1) &= \int_a^b dx_i \int_{x_i}^b dx_{i+1} \rho(x_i) \rho(x_{i+1}|x_i), \\ P_{\text{in}}(k=1) &= \int_a^b dx_i \int_{x_i}^b dx_{i-1} \rho(x_{i-1}) \rho(x_i|x_{i-1}). \end{aligned} \quad (\text{C.7})$$

For chaotic dynamical systems whose trajectories are in the attractor, there exists an invariant probability measure that characterizes the long-term fraction of time spent by the system in the various regions of the attractor. In the case of the Logistic map,

$$F(x_i) = \mu x_i(1 - x_i), \quad (\text{C.8})$$

with parameter $\mu = 4$, the attractor is the whole interval $[0, 1]$ and the probability measure $\rho(x)$ corresponds to [1]

$$\rho(x) = \frac{1}{\pi \sqrt{x(1-x)}}. \quad (\text{C.9})$$

Now, for a deterministic system, the transition probability is simply

$$\rho(x_{i+1}|x_i) = \delta(x_{i+1} - F(x_i)), \quad (\text{C.10})$$

where $\delta(x)$ is the Dirac delta distribution. Equation (C.8) for the Logistic map with $\mu = 4$ and $x \in [0, 1]$ read

$$\begin{aligned} P_{\text{out}}(k=1) &= \int_0^1 dx_i \int_{x_i}^1 dx_{i+1} \rho(x_i) \delta(x_{i+1} - F(x_i)), \\ P_{\text{in}}(k=1) &= \int_0^1 dx_i \int_{x_i}^1 dx_{i-1} \rho(x_{i-1}) \delta(x_i - F(x_{i-1})). \end{aligned} \quad (\text{C.11})$$

Notice that, using the properties of the Dirac delta distribution, the integral $\int_{x_i}^1 \delta(x_{i+1} - F(x_i)) dx_{i+1}$ is equal to one if and only if $F(x_i) \in [x_i, 1]$, what happens iff $0 < x_i < 3/4$, and it is zero otherwise. Therefore the only effect of this integral is to restrict the integration range of x_i to be $[0, 3/4]$. The first equation in (C.12) reduces to

$$P_{\text{out}}(k=1) = \int_0^{3/4} dx_i \rho(x_i) = \frac{2}{3}. \quad (\text{C.12})$$

On the other hand,

$$\int_{x_i}^1 dx_{i-1} \rho(x_{i-1}) \delta(x_i - F(x_{i-1})) = \sum_{\substack{x_k^* | F(x_k^*) = x_i}} \rho(x_k^*) / |F'(x_k^*)|, \quad (\text{C.13})$$

that is, the sum over the roots x^* of the equation $F(x) = x$, iff $F(x_{i-1}) > x_0$. Since $x_{i-1} \in [x_i, 1]$ in the latter integral, the condition $F(x_{i-1}) > x_0$ is verified if $0 < x_i < 3/4$. In fact, if $0 < x_i < 3/4$ there is always a *single* value of $x_{i-1} \in [x_i, 1]$ such that $F(x_{i-1}) = x_i$, so the sum restricts to the adequate root. The particular value is $x_i^* = (1 + \sqrt{1 - x_i})/2$. Making use of these piecewise solutions and Eq. C.10, we finally have

$$P_{\text{in}}(k=1) = \int_0^{3/4} dx_i \frac{\rho(x_i^*)}{4\sqrt{1-x_i}} = \frac{1}{2} \int_0^{3/4} dx_i \rho(x_i) = \frac{1}{3}. \quad (\text{C.14})$$

We conclude that $P_{\text{out}}(k) \neq P_{\text{in}}(k)$ for the Logistic map and hence the KLD measure based on degree distributions is positive.

Recall that $P_{\text{out}}(k=1) = 2/3$ is the probability that the series exhibits a positive jump ($x_i > x_{i-1}$) once in the attractor. These positive jumps must be smaller in size than the negative jumps because, once in the attractor, $\langle x_i \rangle$ is constant. The irreversibility captured by the difference between $P_{\text{out}}(k=1)$ and $P_{\text{in}}(k=1)$ is then the asymmetry of the probability distribution of the slope $x_i - x_{i-1}$ of the original time series. The KLD of the degree distributions given by (3.31) clearly goes beyond this simple signature of irreversibility and can capture more complex and long-range traits.

Reference

1. J. Sprott, G. Rowlands, Int. J. Bifurcat. Chaos **11**, 1865–1880 (2001)

Appendix D

Appendix to Chapter 5

D.1 Numerical Integration of Stochastic Differential Equations

We are interested in integrating numerically equations of the type

$$\dot{x}(t) = f(x(t); t) + g(x(t); t)\xi(t), \quad (\text{D.1})$$

for any generic functions f and g that depend on x and could also depend explicitly on t . We focus on the one-dimensional case for simplicity. In the above equation, $\xi(t)$ is a Gaussian white noise of intensity σ (see Sect. A.1). We want to obtain a numerical scheme that approximates the solution of this kind of equations has the form

$$x(t) = x(0) + \int_0^t f(x(t'); t')dt' + \int_0^t g(x(t'); t')dW(t'). \quad (\text{D.2})$$

We now show two simple numerical schemes that perform this task when (D.1) is taken in Ito or Stratonovich sense [1]. Interestingly, the two methods apply also to both Ito and Stratonovich case in case of additive noise.

D.1.1 Euler's Method

Euler's method uses the simplest integration scheme, where the value of x at time $t + \Delta t$ —with Δt is smaller than any characteristic time of the system—is given by

$$x_{t+\Delta t} = x_t + f(x_t; t)\Delta t + g(x_t; t)\Delta W. \quad (\text{D.3})$$

In the above equation, ΔW is the increment of the Wiener process in Δt , which is given by $\Delta W = \sigma\sqrt{\Delta t}\mathcal{N}(0, 1)$, where $\mathcal{N}(0, 1)$ is a random number generated

by a Gaussian distribution of zero mean and unit variance. This numerical scheme converges to the Ito solution of (D.1) [1].

D.1.2 Heun's Method

Heun's scheme is a refinement of Euler's method

$$\begin{aligned} x_{t+\Delta t} = & x_t + \frac{1}{2} [f(x_t; t) + f(\tilde{x}_{t+\Delta t}; t + \Delta t)] \Delta t \\ & + \frac{1}{2} [g(x_t; t) + g(\tilde{x}_{t+\Delta t}; t + \Delta t)] \Delta W, \end{aligned} \quad (\text{D.4})$$

where $\tilde{x}_{t+\Delta t}$ is the predicted value of x at time $t + \Delta t$ using Euler's scheme,

$$\tilde{x}_{t+\Delta t} = x_t + f(x_t; t) \Delta t + g(x_t; t) \Delta W. \quad (\text{D.5})$$

The numerical integration method given by Eqs. (D.4) and (D.5) converges to the Stratonovich solution of (D.1) as indicated in [1].

Reference

1. W. Rümelin, SIAM J. Numer. Anal. **19**, 604–613 (1982)

Appendix E

Appendix to Chapter 6

E.1 Energetics of Symmetry Breaking and Extended Fluctuation Theorems

The extended fluctuation relation [1, 2] expresses a symmetry of the work distribution when ergodicity is broken. Let F be a nonequilibrium forward process where a system initially in equilibrium, is driven out of equilibrium during a time τ by switching a control parameter λ from λ_A to λ_B . The system is described by a Hamiltonian that depends on the value of the control parameter, H_λ . In the backward process B the system is initially in equilibrium with Hamiltonian H_B and the control parameter evolves in the reverse way, from λ_B to λ_A . The system is assumed to be in contact with a thermal bath at temperature T throughout the process. The microstate of the system at any time t is denoted by $z(t)$. We define by $a(z)$ any function of the microstate of the system in the beginning of the forward process and by $b(z)$ of the microstate in the beginning on the reverse process. Let $p_F(W, a \rightarrow b)$ be the joint probability distribution to measure a work W in the forward process provided that $a(z) = a$ in the beginning of the forward process and $b(z) = b$ in the end of the forward process. The joint probability distribution $p_F(-W, b^* \rightarrow a^*)$ is the probability to perform a work $-W$ in the backward process given that, in the beginning of the backward process, $b(z) = b^*$ and in the end of the backward process $a(z) = a^*$, where $*$ stands for a change of sign in all momenta. The following relation holds [1],

$$p_F(W, a \rightarrow b)e^{-\beta W} = p_B(-W, b^* \rightarrow a^*)e^{-\beta \Delta F}, \quad (\text{E.1})$$

where $\Delta F = F_b - F_a$ is the free energy difference between the states described by a and b , and therefore $e^{-\beta \Delta F} = Z_b/Z_a$. The above equation can be factorized,

$$p_F(W|a \rightarrow b)p_F(a \rightarrow b) = p_B(-W|b^* \rightarrow a^*)p_B(b^* \rightarrow a^*)e^{-\beta(W-\Delta F)}, \quad (\text{E.2})$$

where $p_F(a \rightarrow b)$ and $p_B(b^* \rightarrow a^*)$ are the probabilities to observe the transitions $a \rightarrow b$ ($b^* \rightarrow a^*$) in the forward (backward) process. The distributions $p_F(W|a \rightarrow b)$

and $p_B(-W|b^* \rightarrow a^*)$ measure the work distribution conditioned to the path $a \rightarrow b$ and $b^* \rightarrow a^*$ respectively.

In the symmetry breaking, the system is initially in equilibrium in a reference state 0 and ends in one of the two traps F or M . The probability to choose the trap $\alpha = F, M$ in the forward process is denoted by $p_F(0 \rightarrow \alpha) = p_\alpha$. The restoring-symmetry probability is $p_B(\alpha \rightarrow 0) = 1$, since the system always ends in the reference state in the reverse process. Therefore, for the symmetry breaking, the Maragakis's fluctuation theorem yields,

$$p_F(W|\alpha)p_\alpha = p_B(-W|\alpha)e^{-\beta(W-\Delta F_\alpha)}, \quad (\text{E.3})$$

where $p_F(W|\alpha)$ is the probability to observe a work W in the forward process conditioned to choose the trap α , and $\Delta F_\alpha = F_{\text{fin},\alpha} - F_{\text{ini}}$. Taking logarithms in the above equation,

$$W - \Delta F_\alpha = kT \ln p_\alpha + kT \ln \frac{p_F(W|\alpha)}{p_B(-W|\alpha)}, \quad (\text{E.4})$$

which shows that $p_F(W|\alpha)$ and $p_B(-W|\alpha)$ satisfy a fluctuation theorem similar to Crooks fluctuation theorem. These probability distributions should cross in $W_\alpha^* = \Delta F + \ln p_\alpha$. Averaging over all the realizations that end in the α trap,

$$\langle W \rangle_\alpha - \Delta F_\alpha = kT \ln p_\alpha + kT D[p_F(W|\alpha)||p_B(-W|\alpha)] \geq kT \ln p_\alpha. \quad (\text{E.5})$$

Equation (E.5) implies the inequality (6.1) relating the average dissipation and the choice probability in a symmetry breaking. The KLD $D[p_F(W|\alpha)||p_B(-W|\alpha)]$ is the correction term that appears when the process is not done quasistatically. When the process is done reversibly, this KLD vanishes, and

$$\langle W \rangle_\alpha - \Delta F_\alpha = kT \ln p_\alpha, \quad (\text{E.6})$$

which coincides with the equality in (6.1).

E.2 Symmetry Breaking with Harmonic Traps

We now study the energetics of the symmetry breaking of a process where a Brownian particle is immersed in a thermal bath at temperature T in a bistable potential that is modulated following a protocol analogous to the one described in Sect. 6.3. A movable optical trap M , is initially overlapped with a fixed trap F with their centers located in $x = 0$ for a longer time than the relaxation time of the traps so the particle relaxes to equilibrium. The M trap is moved in one dimension from $x = 0$ to $x = L$ at constant velocity v and let again to relax in $x = L$. The separation distance L is assumed to be large enough to avoid that the particle feels the force extorted by the

F trap if it ends in the M trap, and vice versa. During the process an electrostatic fields induces a force $\mathbf{F}_{\text{elec}} = -F_{\text{elec}}\mathbf{e}_x$, F_{elec} being the modulus of the force.

In the quasistatic limit, the average work done along the trajectories in which the particle ends in the $\alpha = F, M$ trap is equal to [3]

$$\langle W \rangle_\alpha - \Delta F_\alpha = kT \ln \frac{p_\alpha}{\tilde{p}_\alpha}, \quad (\text{E.7})$$

where ΔF_α is the free energy difference between the final and the initial equilibrium states of the particle and p_α (\tilde{p}_α) the probability of the particle to choose the trap α in the forward (backward) process.

To obtain the relation between dissipation and irreversibility in such symmetry breaking process, we derive analytic expressions for the initial and free energy differences as well as for \tilde{p}_α . The initial free energy of this process is

$$F_{\text{init}} = -kT \ln Z_{F+M}, \quad (\text{E.8})$$

where Z_{F+M} is the partition function of the particle in the combined trap. The final free energy is

$$F_{\text{fin}} = -kT \ln(Z_F + Z_M), \quad (\text{E.9})$$

where Z_α is the partition function of the particle in a single trap $\alpha = F, M$ in its final location. Notice that they can depend on the electrostatic force F_{elec} and the distance L .

Let p_α be the probability that the particle ends in trap $\alpha = F, M$ after the separation. In the backward process the probability to find the particle in the α trap is just the probability to find it there in the initial condition (recall that in the backward process starts with the two traps separated). Since the initial condition is the canonical state, this probability is

$$\tilde{p}_\alpha = \frac{Z_\alpha}{Z_F + Z_M}, \quad (\text{E.10})$$

Inserting $\Delta F = F_{\text{fin}} - F_{\text{init}}$ using (E.8) and (E.9), and replacing \tilde{p}_α by (E.10) in (E.7), one obtains

$$\langle W \rangle_\alpha + kT \ln(Z_F + Z_M) - kT \ln Z_{F+M} = kT \ln \frac{p_\alpha (Z_F + Z_M)}{Z_\alpha}, \quad (\text{E.11})$$

yielding

$$\langle W \rangle_\alpha + kT \ln Z_\alpha - kT \ln Z_{F+M} = kT \ln p_\alpha. \quad (\text{E.12})$$

To calculate the partition functions we will assume that the traps generate harmonic potentials of stiffness κ centered in x_α , $U_\alpha(x) = \frac{\kappa}{2}(x - x_\alpha)^2$, where x_α is the position of the center of the trap, $x_F = 0$ and $x_M(t) = vt$. In the presence of the electrostatic force, the potential changes by a factor $-F_{\text{elec}}x$,

$$U_\alpha(x) = \frac{\kappa}{2}(x - x_\alpha)^2 - F_{\text{elec}}x. \quad (\text{E.13})$$

Neglecting inertia, the partition functions read

$$-kT \ln Z_F = -kT \ln \sqrt{\frac{2\pi}{\beta\kappa}} - \frac{F_{\text{elec}}^2}{2\kappa}, \quad (\text{E.14})$$

$$-kT \ln Z_M = -kT \ln \sqrt{\frac{2\pi}{\beta\kappa}} - \frac{F_{\text{elec}}^2}{2\kappa} + F_{\text{elec}}L, \quad (\text{E.15})$$

$$-kT \ln Z_{F+M} = -kT \ln \sqrt{\frac{\pi}{\beta\kappa}} - \frac{F_{\text{elec}}^2}{4\kappa}. \quad (\text{E.16})$$

If we replace Eqs. (E.14), (E.15) and (E.16) in (E.12) we obtain

$$\langle W \rangle_M - F_{\text{elec}}L + \frac{F_{\text{elec}}^2}{4\kappa} + kT \ln \sqrt{2} = k_B T \ln p_M, \quad (\text{E.17})$$

$$\langle W \rangle_F + \frac{F_{\text{elec}}^2}{4\kappa} + kT \ln \sqrt{2} = k_B T \ln p_F. \quad (\text{E.18})$$

For small forces, i.e., $F_{\text{elec}}^2/4\kappa \ll F_{\text{elec}}L$, or equivalently $F_{\text{elec}} \ll \kappa L$,

$$\langle W \rangle_M - F_{\text{elec}}L = kT \ln p_M + \text{const.} \quad (\text{E.19})$$

E.3 Measuring Entropy Production in the Symmetry Breaking Experiment

We now show how to measure the stochastic work, the heat and the entropy production from time series obtained experimentally in the symmetry breaking experiment described in Chap. 6, where the sampling time is finite and equal to Δt .

Consider a Brownian particle in a potential $U(x, t)$, changing in time, and temperature T . The position of the particle at time t is denoted by x_t . The change of potential energy in an interval of time $[t, t + \Delta t]$ is

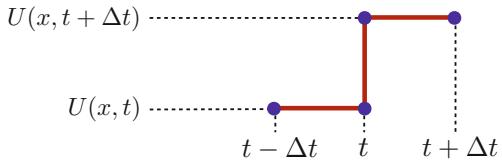
$$\Delta U(t) = U(x_{t+\Delta t}, t + \Delta t) - U(x_t, t), \quad (\text{E.20})$$

which can be written as

$$\Delta U(t) = \delta W(t) + \delta Q(t), \quad (\text{E.21})$$

where the work and the heat transferred in Δt are given by

Fig. E.1 Change in the potential energy in discrete steps



$$\delta W(t) = U(x_t, t + \Delta t) - U(x_t, t), \quad (\text{E.22})$$

$$\delta Q(t) = U(x_{t+\Delta t}, t + \Delta t) - U(x_t, t + \Delta t). \quad (\text{E.23})$$

These expressions are exact if the potential changes in discrete steps. Suppose that the potential is constant and equal to $U(x, t)$ between $t - \Delta t$ and t (see Fig. E.1). We assume that the potential changes abruptly to $U(x, t + \Delta t)$ at time t and (6.2) is constant up to $t + \Delta t$ and so on. The change introduces an energy $\delta W(t) = U(x_t, t + \Delta t) - U(x_t, t)$. Then the particle moves from x_t to $x_{t+\Delta t}$ in the constant potential $U(x, t + \Delta t)$. The change in energy $\delta Q(t) = U(x_{t+\Delta t}, t + \Delta t) - U(x_t, t + \Delta t)$ is therefore due to the thermal bath.

If the potential changes smoothly, the above expressions are still a good approximation for the work and heat. In fact, they reduce to Sekimoto's expressions for $\Delta t \rightarrow 0$,

$$\delta W(t) = \frac{\partial U(x_t, t)}{\partial t} \Delta t, \quad (\text{E.24})$$

$$\delta Q(t) = \left. \frac{\partial U(x, t)}{\partial x} \right|_{x=x_t} \circ \Delta x_t. \quad (\text{E.25})$$

References

1. P. Maragakis, M. Spichty, M. Karplus, J. Phys. Chem. B **112**, 6168–6174 (2008)
2. I. Junier, A. Mossa, M. Manosas, F. Ritort, Phys. Rev. Lett **102**, 070602 (2009)
3. R. Kawai, J.M.R. Parrondo, C.V. den Broeck, Phys. Rev. Lett. **98**, 80602 (2007)

Appendix F

Appendix to Chapter 7

F.1 Power Spectrum Density of a Brownian Particle Trapped with a Harmonic Trap

Let us consider an overdamped Brownian particle that moves in one dimension and that is trapped by a harmonic trap defined by the optical potential $V(x) = \frac{1}{2}\kappa x^2$. The position of the Brownian particle obeys the overdamped Langevin equation

$$\gamma \dot{x}(t) = -\kappa x(t) + \xi(t), \quad (\text{F.1})$$

where $\xi(t)$ models the thermal noise, i.e. it is a Gaussian white noise with zero mean $\langle \xi(t) \rangle = 0$ and correlation

$$\langle \xi(t) \xi(t') \rangle = 2\gamma k_B T \delta(t - t'). \quad (\text{F.2})$$

The Fourier transform of (F.1) is

$$i\omega \tilde{x}(\omega) = -\kappa \tilde{x}(\omega) + \tilde{\xi}(\omega), \quad (\text{F.3})$$

where $\tilde{x}(\omega) = \int_{-\infty}^{\infty} dt e^{-i\omega t} x(t)$, ω being the angular frequency and $i = \sqrt{-1}$ is the imaginary unit. The PSD of the position of the Brownian particle is defined as

$$\text{PSD}_x(\omega) = |\tilde{x}(\omega)|^2, \quad (\text{F.4})$$

From (F.3) and (F.2) we obtain

$$\text{PSD}_x(\omega) = \frac{2k_B T \gamma}{\omega^2 \gamma^2 + \kappa^2}. \quad (\text{F.5})$$

If we define the *corner frequency* as $\omega_c = \kappa/\gamma$ one immediately gets

$$\text{PSD}_x(\omega) = \frac{2k_B T}{\gamma} \frac{1}{\omega^2 + \omega_c^2}. \quad (\text{F.6})$$

Equivalently using $f = 2\pi\omega$ and $f_c = 2\pi\omega_c$

$$\text{PSD}_x(f) = \frac{k_B T}{2\pi^2 \gamma} \frac{1}{f^2 + f_c^2}, \quad (\text{F.7})$$

which proves Eq. (7.4).

F.2 Effective Kinetic Temperature from the Position Histograms

We now prove the analytical expression (7.9) that relates T_{hist} with the properties of the additional random force applied in the experimental setup used in Chap. 7.

We consider a Brownian particle whose motion is described by the Langevin equation (7.2),

$$\gamma \dot{x}(t) = -\kappa x(t) + \xi(t) + \zeta(t), \quad (\text{F.8})$$

where $\xi(t)$ and $\zeta(t)$ are Gaussian noises with zero average $\langle \xi(t) \rangle = \langle \zeta(t) \rangle = 0$ and their correlation functions are

$$\langle \xi(t) \xi(t') \rangle = 2k_B T \gamma \delta(t - t'), \quad (\text{F.9})$$

$$\langle \zeta(t) \zeta(t') \rangle = \sigma^2 \Gamma(t - t'). \quad (\text{F.10})$$

The solution of Eq. (F.8) is:

$$x(t) = x(0)e^{-t/\tau_r} + \frac{e^{-t/\tau_r}}{\gamma} \int_0^t ds e^{s/\tau_r} [\xi(s) + \zeta(s)], \quad (\text{F.11})$$

where $\tau_r = \gamma/\kappa$ is the relaxation time in the trap. If we multiply $x(t)$ by Eq. (F.8), and we do the average over different trajectories, we get

$$\frac{\gamma}{2} \frac{d\langle x^2 \rangle}{dt} = -\kappa \langle x^2 \rangle + \langle \xi x \rangle + \langle \zeta x \rangle, \quad (\text{F.12})$$

where $\langle \xi x \rangle$ and $\langle \zeta x \rangle$ can be calculated using Novikov's theorem:

$$\langle \zeta(t) x(t) \rangle = \int_0^t ds \langle \zeta(t) \zeta(s) \rangle \left\langle \frac{\delta x(t)}{\delta \zeta(s)} \right\rangle, \quad (\text{F.13})$$

where $\frac{\delta x(t)}{\delta \zeta(s)}$ stands for functional the functional derivative of $x(t)$ with respect to $\zeta(s)$. Using Eq. (F.11), $\left\langle \frac{\delta x(t)}{\delta \zeta(s)} \right\rangle = \frac{1}{\gamma} e^{-(t-s)/\tau_r}$, therefore

$$\langle \zeta(t)x(t) \rangle = \frac{\sigma^2}{\gamma} \int_0^t ds \Gamma(t-s) e^{-(t-s)/\tau_r}, \quad (\text{F.14})$$

whereas for the thermal noise

$$\langle \xi(t)x(t) \rangle = \int_0^t ds \langle \xi(t)\xi(s) \rangle \left\langle \frac{\delta x}{\delta \xi(s)} \right\rangle = k_B T. \quad (\text{F.15})$$

Using Eqs. (F.14) and (F.15), Equation (F.12) can be rewritten as follows,

$$\frac{\gamma}{2} \frac{d\langle x^2 \rangle}{dt} = -\kappa \langle x^2 \rangle + k_B T + \frac{\sigma^2}{\gamma} \int_0^t dt' \Gamma(t') e^{-t'/\tau_r}. \quad (\text{F.16})$$

In the steady state $t \rightarrow \infty$ and $d\langle x^2 \rangle/dt = 0$. In this limit, the above equation yields (7.8)

$$\kappa \langle x^2 \rangle_{ss} = k_B T + \frac{\sigma^2}{\gamma} \int_0^\infty dt' \Gamma(t') e^{-t'/\tau_r}. \quad (\text{F.17})$$

By using equipartition theorem, we get T_{hist} as a function of the mean square displacement in the steady state,

$$T_{\text{hist}} = \frac{\kappa \langle x^2 \rangle_{ss}}{k_B} = T + \frac{\sigma^2}{\gamma k_B} \int_0^\infty dt' \Gamma(t') e^{-t'/\tau_r}, \quad (\text{F.18})$$

which proves Eq. (7.9).

We notice that $\Gamma(t)$ has a characteristic timescale given by τ_c , so it can be expressed as a function of t/τ_c , say $\Gamma(t) = \Gamma(t/\tau_c)$. The correlation of the noise decays in this timescale, which makes that the integral in the above equation can be expressed, by using the change of variable $s = t/\tau_r$,

$$\int_0^\infty dt' \Gamma(t'/\tau_c) e^{-t'/\tau_r} = \tau_r \int_0^\infty ds \Gamma\left(\frac{\tau_r}{\tau_c}s\right) e^{-s}. \quad (\text{F.19})$$

If $\tau_c \ll \tau_r$, the exponential decays much slower with s than the correlation of the noise in the units above, therefore it can be approximated by $e^{-s} \simeq 1 - s$,

$$\int_0^\infty dt' \Gamma(t'/\tau_c) e^{-t'/\tau_r} \simeq \tau_r \int_0^\infty ds \Gamma\left(\frac{\tau_r}{\tau_c}s\right) (1-s) = \frac{1}{2} - \frac{\tau_c}{\tau_r}. \quad (\text{F.20})$$

For the last equality, we have used that $\Gamma(t)$ is normalized, $\int_{-\infty}^\infty dt \Gamma(t/\tau_c) = 1$, that it is symmetric around $t = 0$ and the definition of the correlation time of the noise $\tau_c = \int_0^\infty dt t \Gamma(t/\tau_c)$. By using Eqs. (F.20) into (F.18), we prove Eq. (7.10)

$$T_{\text{hist}} \simeq \frac{\sigma^2}{2\gamma k_B} \left(1 - \frac{2\tau_c}{\tau_r}\right). \quad (\text{F.21})$$

F.3 Average Work Exerted to a Dragged Brownian Particle

We now calculate the average of the work done on an overdamped Brownian particle trapped with a harmonic potential when the trap centre is moved at a constant velocity v for a period of time t . In this situation, the work done to move a trap that creates a moving quadratic potential $V(x(t), x_0(t)) = \frac{\kappa}{2}(x - x_0(t))^2$ is equal to

$$W = \int_0^t -\kappa(x - vt') v dt' = -\kappa v \int_0^t y(t') dt', \quad (\text{F.22})$$

where we have used the definition of the work in Eq. (1.59) and $x_0(t) = v t$ as control parameter, and we have introduced the variable $y(t) = x(t) - vt$. $y(t)$ satisfies the following Langevin equation

$$\gamma \dot{y}(t) = -\kappa y(t) - \gamma v + \xi(t) + \zeta(t), \quad (\text{F.23})$$

whose solution is

$$y(t) = y(0)e^{-t/\tau_r} + \frac{e^{-t/\tau_r}}{\gamma} \int_0^t ds e^{s/\tau_r} [-\gamma v + \xi(s) + \zeta(s)]. \quad (\text{F.24})$$

The average over trajectories is given by

$$\langle y(t) \rangle = \langle y(0) \rangle e^{-t/\tau_r} - v \tau_r [1 - e^{-t/\tau_r}] = -v \tau_r [1 - e^{-t/\tau_r}], \quad (\text{F.25})$$

where $\langle y(0) \rangle = \langle x(0) \rangle = 0$, since the bead is initially in equilibrium oscillating around the trap center located in $x = 0$. The average work reads

$$\langle W \rangle = -\kappa v \int_0^t \langle y(t') \rangle dt' = \gamma v^2 t + \frac{\gamma^2 v^2}{\kappa} [e^{-t/\tau_r} - 1], \quad (\text{F.26})$$

which is the expression that we introduced in Eq. (7.12).

Publications

Some ideas and figures have appeared or will appear in the following publications of the author.

1. *Estimating dissipation with single stationary trajectories*. É. Roldán and J.M.R. Parrondo, *Phys. Rev. Lett.* **105**, 150607 (2010).
2. *Entropy production and Kullback–Leibler divergence between stationary trajectories of discrete systems*. É. Roldán and J.M.R. Parrondo, *Phys. Rev. E* **85**, 031129 (2012).
3. *Time series irreversibility: a visibility graph approach*. L. Lacasa, Á. Núñez, É. Roldán, J.M.R. Parrondo and B. Luque, *Eur. Phys. Journal B* **85**, 6, 1–11 (2012).
4. *Effective heating to several thousand kelvin of an optically trapped sphere in a liquid*. I.A. Martínez, É. Roldán, J.M.R. Parrondo and Dmitri Petrov, *Phys. Rev. E* **87**, 032159 (2013)
5. *Universal features in the energetics of symmetry breaking*. É. Roldán, I.A. Martínez, J.M.R. Parrondo and Dmitri Petrov, *Nature Phys.* **10**, 457–461 (2014).
6. *Measuring kinetic energy changes in the mesoscale with low acquisition rates*. É Roldán, I.A. Martínez, L. Dinis and R.A. Rica, *Appl. Phys. Lett.* **104**(23), 234103. *arXiv:1403.2969* (2014).
7. *Quantifying irreversibility in biological time series as a tool to estimate dissipation: the case of oscillations in hair cell bundles*. É. Roldán, F. Jülicher, P. Martin and J.M.R. Parrondo (in preparation).
8. *Implementing nonequilibrium processes with external random forces*. P. Mestres, I.A. Martínez, A. Ortiz-Ambriz, R. Rica and É. Roldán (in preparation).