PhD-FSTM-2020-09
The Faculty of Sciences, Technology and Medicine

DISSERTATION

Defence held on 16/03/2020 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN Physique

by

TIM HERPICH
Born on 21 June 1989 in Münchberg, Germany

COLLECTIVE EFFECTS IN STOCHASTIC THERMODYNAMICS

Dissertation defence committee

Dr Massimiliano Esposito, dissertation supervisor
Professor, Université du Luxembourg

Dr Alexandre Tkatchenko, Chairman
Professor, Université du Luxembourg

Dr Thomas Schmidt, Vice Chairman
Professor, Université du Luxembourg

Dr Juan Garrahan
Professor, University of Nottingham

Dr Alberto Imparato
Assistant Professor, Aarhus University

Collective Effects In Stochastic Thermodynamics

by

TIM HERPICH

It is not that we have a short space of time, but that we waste much of it.

→←

Abstract

Collective effects are ubiquitous in both living and nonliving nature. Prominent examples that can be observed outside laboratories are given by synchronization processes in firefly flashing and cricket chirping or collective dynamics in swarms of fish and birds. These cooperative phenomena are also present at the nanoscale, *e.g.* many molecular motors (kinesin) can propagate along the same microtubule which occasionally leads to traffic jam but can also trigger synergy effects when performing mechanical work. On the other hand, any functional biological system is intrinsically out-of-equilibrium, *i.e.* there is energy and matter exchange with their environment. These processes are characterized by thermodynamics, the theory of energy conversion, and thus constrained by its fundamental laws: energy can not be created or destroyed and energy conversion entails dissipative losses that limit the thermodynamic efficiency.

In its first part, this doctoral thesis addresses the naturally arising question of the relationship between thermodynamics and collective effects. For this purpose, I formulate a whole class of thermodynamically consistent interacting many-body clock models that display synchronization. My first major achievement is to demonstrate that synchronization, *i.e.* cooperative phenomena, can enhance the performance in a macroscopic assembly of interacting microscopic machines. As another example of how collective effects can positively affect the operation of a thermodynamic many-body system, I compare the energetic costs of information erasure in an array of binary units with and without majority-logic decoding, a scheme to decode repetition codes.

In order to determine the thermodynamic properties of macroscopic ensembles made up of interacting microscopic machines, the development of methodologies to study fluctuating thermodynamic quantities across different length scales is required. The second major achievement of this thesis is to provide a consistent nonequilibrium thermodynamic description across three different scales - microscopic, mesoscopic and macroscopic - for systems made up of many identical bodies with all-to-all interactions that can be coupled to multiple heat reservoirs and are subjected to (non-)autonomous external forces. The thermodynamic consistency of the different representations of the stochastic dynamics is shown to be encoded in a detailed fluctuation theorem that exists on each of these levels. Moreover, as a complement, I will assess the validity of various coarse-graining methods established for master-equation systems in the context of an underdamped system made up of interacting Brownian particles.

Table of Contents

Page

Abstract **v**

Table of Contents **ix**

List of Publications **xi**

List of Figures **xiv**

1 Introduction and Motivation **1**

2 Preliminaries **7**
2.1 Traditional Thermodynamics 7
2.2 Stochastic Processes 10
2.2.1 Probability Theory 10
2.2.2 Markov Processes 12
2.3 Stochastic Thermodynamics 14
2.3.1 Ensemble Thermodynamics 14
2.3.2 Trajectory Thermodynamics 18
2.3.2.1 Stochastic First Law and Entropy Balance 18
2.3.2.2 Generating Function Techniques 22
2.3.2.3 Fluctuation Theorems 25

3 Enhancing The Thermodynamic Performance via Collective Effects **31**
3.1 Introduction 31
3.2 Thermodynamics of Nonequilibrium Phase Transitions in Clock Models . . 34
3.2.1 Driven Three-State Model 34
3.2.1.1 Setup 34
3.2.1.2 Master Equation 35
3.2.1.3 Coarse-Graining 36
3.2.1.4 Mean-Field Dynamics 39
3.2.1.5 Stochastic Dynamics 45
3.2.1.6 Stochastic Thermodynamics 55
3.2.1.7 Efficiency at Maximum Power 65

3.2.2 Class of Driven Potts Models . . . 68
3.2.2.1 Setup . . . 68
3.2.2.2 Mean-Field Dynamics . . . 69
3.2.2.3 Mean-Field Thermodynamics . . . 74
3.2.2.4 Efficiency at Maximum Power . . . 77
3.3 Thermodynamics of Majority-Logic Decoding in Information Erasure . . . 81
3.3.1 Majority-Logic Decoding . . . 81
3.3.2 Information Thermodynamics . . . 84
3.3.2.1 Definitions . . . 84
3.3.2.2 Reversible Erasure Protocols . . . 86
3.3.2.3 Finite-Time Erasure Protocol under Majority-Logic Decoding 88
3.3.2.4 Fixed Erasure Duration . . . 93
3.3.2.5 Optimal Erasure Protocol under Majority-Logic Decoding 93

4 Coarse-Graining in Stochastic Thermodynamics 99
4.1 Introduction . . . 99
4.2 Many-Body Systems with All-To-All Interactions . . . 101
4.2.1 Stochastic Dynamics . . . 101
4.2.1.1 Microscopic Description . . . 101
4.2.1.2 Mesoscopic Description . . . 103
4.2.2 Stochastic Thermodynamics . . . 106
4.2.2.1 Trajectory Definitions Revisited . . . 106
4.2.2.2 Generating Function Techniques Revisited . . . 108
4.2.2.3 Detailed Fluctuation Theorems Across Scales . . . 112
4.2.2.4 Microscopic And Mesoscopic First And Second Law . . . 113
4.2.3 Macroscopic Theory . . . 115
4.2.3.1 Macroscopic Fluctuations . . . 115
4.2.3.2 Mean-Field Description . . . 117
4.2.4 Example . . . 119
4.3 Two Interacting Underdamped Particles . . . 125
4.3.1 Stochastic Thermodynamics in Underdamped Systems . . . 125
4.3.1.1 Single Underdamped Particle . . . 125
4.3.1.2 Special Cases . . . 127
4.3.1.3 Two Coupled Underdamped Particles . . . 128
4.3.2 Coarse graining . . . 131
4.3.2.1 Effective Dynamics . . . 131
4.3.2.2 Effective Thermodynamics . . . 132
4.3.2.3 Limiting Cases . . . 139
4.3.3 Two Linearly Coupled Harmonic Oscillators . . . 145
4.3.3.1 Full Solution . . . 145
4.3.3.2 Fast-Dynamics Limit . . . 148
4.3.3.3 Large-Mass Limit . . . 150

5 Conclusion and Perspectives **153**

A Appendices **161**
A.1 Derivation of the Fluctuation Theorem (2.105) 161
A.2 Proof of Supercritical Hopf Bifurcation (3.38) 165
A.3 Simulation of a Markovian Jump Process 167
A.4 The Monotonicity of the Entropy Function (3.140) 171
A.5 Derivation of the Optimal Erasure Protocol 173
A.6 Martin-Siggia Rose Formalism in a Nutshell 174
A.7 Derivation of the Fluctuation Theorem (4.87) 177
A.8 Generalization to Many-Body Systems 180

Bibliography **192**

Author's Contribution **206**

Acknowledgments **208**

List of publications

The following publications constitute the core of my doctorate and have been published in or submitted to the following peer-reviewed journals:

- Tim Herpich, Juzar Thingna, and Massimiliano Esposito
  Collective Power: Minimal Model for Thermodynamics of Nonequilibrium Phase Transitions
  Phys. Rev. X **8**, 031056 (2018), (Ref. [1]).
  Section 3.2 is in words and content partially equivalent to this publication.

- Tim Herpich and Massimiliano Esposito
  Universality in driven Potts models
  Phys. Rev. E **99**, 022135 (2019), (Ref. [2]).
  Section 3.2 is in words and content partially equivalent to this publication.

- Shiqi Sheng, Tim Herpich, Giovanni Diana, and Massimiliano Esposito
  Thermodynamics of Majority-Logic Decoding in Information Erasure
  Entropy **21**, 284 (2019), (Ref. [3]).
  Section 3.3 is in words and content partially equivalent to publication.

- Tim Herpich, Kamran Shayanfard, and Massimiliano Esposito
  Effective Thermodynamics of Two Interacting Underdamped Particles
  Phys. Rev. E **101**, 022116 (2020), (Ref. [4]).
  Section 4.2 is in words and content partially equivalent to this publication.

- Tim Herpich, Tommaso Cossetto, Gianmaria Falasco and Massimiliano Esposito
  Stochastic thermodynamics of all-to-all interacting and identical many-body systems.
  Submitted to New Journal of Physics. Preprint on arXiv:2001.09744, (Ref. [5]).
  Section 4.3 is in words and content partially equivalent to this publication.

A publication in preparation *not* included in this thesis is:

- Tim Herpich and Massimiliano Esposito
  The molecular zipper: A nonequilibrium renarration
  In preparation.

List of Figures

Page

2.1 Classification of equilibrium phase transitions . . . 9
2.2 Illustration of a discrete system weakly coupled to a heat bath . . . 15
2.3 Schematic representation of a single trajectory and its energetics . . . 19
2.4 Schematic depiction of the backward and forward process . . . 27

3.1 Depiction of all-to-all interacting three-state units . . . 34
3.2 Mean-field phase space of the three-state model . . . 41
3.3 Density plot of the occupation densities at the critical temperature . . . 42
3.4 Density plot of the equilibrium occupation density . . . 43
3.5 Density plot of the non-equilibrium occupation density . . . 44
3.6 Dominant part of Markov generator's spectrum for different temperatures 47
3.7 Dominant part of Markov generator's spectrum for different system sizes in the synchronization phase . . . 49
3.8 Dominant part of Markov generator's spectrum for different system sizes in the multistability phase . . . 50
3.9 3D-plot of the Markov generator's spectrum for different temperatures . . . 51
3.10 Many eigenvalues of the Markov generator for different system sizes . . . 52
3.11 Stochastic dynamics of the three-state model . . . 54
3.12 Role of initial condition for the metastability in the three-state model . . . 56
3.13 Dissipated work as a function of temperature for interacting and noninteracting systems . . . 62
3.14 Dissipated work as a function of time for different system sizes . . . 64
3.15 Comparison of dissipated work between interacting and noninteracting systems for different forces . . . 65
3.16 Power-Efficiency trade-off in the three-state model . . . 67
3.17 Depiction of all-to-all interacting q-state units . . . 69
3.18 Stochastic dynamics of the q-state model . . . 74
3.19 Comparison of dissipated work between interacting and noninteracting systems for different forces and different number of unit states . . . 76
3.20 Power landscape for different temperatures and forces of systems with an odd number of unit states . . . 77
3.21 Efficiency at maximum power different temperatures of systems with an odd number of unit states . . . 78

3.22 Power landscape for different temperatures and forces of systems with an even number of unit states . . . 79
3.23 Efficiency at maximum power for different temperatures of systems with an odd number of unit states . . . 80
3.24 Schematics of a single-unit and a majority-logic decoding bit . . . 81
3.25 Majority-logic decoding for different array sizes and threshold values. . . . 83
3.26 Heat dissipation and efficiency for reversible erasure processes in different macroscopic bits . . . 86
3.27 Heat dissipation, erasure power and efficiency for erasure processes in different macroscopic Fermi bits . . . 92
3.28 Efficiency for fixed-time erasure processes in different macroscopic Fermi bits 94
3.29 Efficiency for erasure processes following an optimal protocol in different macroscopic Arrhenius bits . . . 95
3.30 Heat dissipation and efficiency for erasure processes following an optimal protocol for different erasure times and macroscopic Arrhenius bits . . . 97

4.1 Density plot of the stationary solution for different temperatures and all physical initial conditions . . . 120
4.2 The equilibrium free energy in the vicinity of the critical point β_c . . . 121
4.3 The scaled cumulant generating function and the rate function . . . 123
4.4 The scaled cumulant generating function and the rate function in the presence of a nonequilibrium phase transition . . . 124
4.5 Schematic depiction of two-particle system and the coarse-grained one-particle system . . . 129
4.6 Comparison between full and effective thermodynamics in the limit of time-scale separation . . . 149
4.7 Variance of the second particles' spatial coordinates with increasing particle mass . . . 151
4.8 Comparison between full and effective thermodynamics in the deterministic limit . . . 151

A.1 Exemplary system consisting of five interacting particles . . . 181

Chapter 1

Introduction and Motivation

The starting point of thermodynamics as a modern scientific discipline is commonly attributed to Sadi Carnot's 1824 book *Reflections on the Motive Power of Fire and on Machines Fitted to Develop that Power* [6]. It is not a mere coincidence that thermodynamics was established during the 19^{th} century, since the Industrial Revolution dating back to the end of the 18^{th} century and the emergence of heat machines that convert thermal energy into mechanical work, had prompted the question about the underlying principles that govern these energy-converting systems [7]. The traditional theory of thermodynamics succeeded in formalizing the energetic processes inside heat engines. At the core of this phenomenological theory are the first and second law of thermodynamics, which prescribe energy conservation and that energy exchange comes at the cost of dissipation, respectively. The entropy production quantifies the amount of dissipation and is therefore non-negative during any macroscopic thermodynamic process [8].

The range of applicability of traditional thermodynamics was limited to macroscopic systems kept at equilibrium. These limitations were gradually overcome in the past and current century by generalizing the phenomenological theory towards a systematic treatment of nonequilibrium stochastic processes: Notably, in the first half of the last century Lars Onsager established a first formulation of irreversible thermodynamics for systems in the linear-response regime, *i.e.* systems subjected to small perturbations and thus close to equilibrium. For this class of systems he derived universal symmetries, the celebrated reciprocal relations [9, 10], of the coefficients that linearly couple currents to the thermodynamic forces from which they ensue. The theory of irreversible thermodynamics was subsequently extended [11], most importantly by the works of Ilya Prigogine in which he introduced the notion of *local* equilibrium in order to formulate irreversible processes in terms of equilibrium quantities [12].

Over the past two decades the field of nonequilibrium thermodynamics has witnessed

spectacular progress due to the development of stochastic thermodynamics [13–18]. This novel formalism rigorously builds the thermodynamic structure on top of stochastic systems obeying Markovian dynamics. Stochastic thermodynamics has proven instrumental to systematically infer the thermodynamics in small fluctuating systems that can be driven arbitrarily far away from equilibrium. This framework allows to formulate the thermodynamics at the level of a single trajectory, that is for a single realization of the Markov process. Remarkably, it was discovered that the trajectory entropy production needs not to be always non-negative [19]. This is in agreement with the so-called fluctuation theorems, which are universal symmetries exhibited by various fluctuating thermodynamic observables for systems arbitrarily far away from equilibrium. A prominent example is the Seifert fluctuation theorem [20]

$$\frac{P(\Sigma)}{\tilde{P}(-\Sigma)} = \mathrm{e}^{\Sigma}, \tag{1.1}$$

which in words states that the probability to observe an entropy production Σ along a given process is exponentially more probable than to observe a corresponding decrease $-\Sigma$ along the time-reversed process. Thus, processes with negative entropy production exist, but they are exponentially rare. The known results from traditional thermodynamics are however reproduced if an ensemble of trajectories is considered. In particular, the detailed fluctuation theorem (1.1) implies the non-negativity of the average entropy production.

Stochastic thermodynamics has been successfully applied in several fields such as information processing [21], chemical reaction networks [22] and active matter [23, 24]. Its predictions have been experimentally validated in various fields ranging from electronics to single molecules and Brownian particles [25, 26]. It has been particularly successful in studying the performance of small energy converters operating far-from-equilibrium and their power-efficiency trade-off [13, 27–31]. Until now, most of the focus has been on systems with finite phase space or few particle systems. Of particular interest hereby has been the study of the performance of small energy converters operating far-from-equilibrium [13, 26, 32] (e.g. thermoelectric quantum dots [33, 34], photoelectric nanocells [35], molecular motors [17, 36–39]) and their power-efficiency trade-off [27, 30, 31, 40–44]. The efficiency fluctuations in generic stochastic processes were first studied in [28, 45]. It was found that the upper efficiency bound derived for reversible macroscopic cyclic processes, the Carnot efficiency, can also be reached in irreversible stochastic processes, although the associated probability is the smallest of the entire probability distribution of possible efficiency values. While these microscopic systems have thus been shown to make very efficient energy converters, the main drawback, on absolute terms, remains their low power output rendering them impractical for daily-life applications.

A natural and trivial way to overcome this limitation is to assemble large numbers of these efficiently operating microscopic machines and to collect the cumulated power output. The nontrivial approach we will pursue is to allow for interactions between the microscopic machines. Interacting many-body systems can give rise to a very rich variety of emergent behaviors such as phase transitions corresponding to a dramatic change of the system's

qualitative behavior. The thermodynamics of equilibrium phase transitions in interacting many-body systems has a long history and is well understood [46–50]. When driven out-of-equilibrium, these systems are known to give rise to complex dynamical behaviors [51–57]. After many works that focused on the ensemble average description, progress was also made in characterizing the fluctuations of these dynamical systems [58–62]. Yet, the thermodynamics of nonequilibrium phase transitions has been started to be explored only recently [63–72]. Part of the reason is that before the development of stochastic thermodynamics there was no formalism to systematically address the thermodynamics properties in nonequilibrium systems. Besides the general interest in exploring the thermodynamic signatures in nonequilibrium phase transitions, it has also been observed that they can play a crucial role in how an ensemble of microscopic devices performs as a team: For instance, many (kinesin) molecular motors can operate on the same microtubule, which in turn can lead to the formation of traffic jam in some cases [73]. On the other hand, the kinesin motors can carry the same cargo which results in a strong synergy [74].

Ironically, although it took about two hundred years to change the scale on which the thermodynamic properties of a system could be consistently accessed from a macroscopic to a microscopic one, there is now a practical motivation to reverse that direction of progress, *i.e.* moving back from microscopic to macroscopic scales, *while* exploiting the tools of stochastic thermodynamics. Hence it is the declared goal of this thesis, starting from a system of interacting microscopic machines, to consistently formulate the stochastic thermodynamics thereof across different length scales and to study the collective effects, *e.g.* synchronization, that may emerge at the macroscopic level. Hence there are two main research questions we will address in the following.

Enhancing the Thermodynamic Performance by Collective Effects

We are particularly interested in identifying interactions from which phase transitions ensue that in turn induce synergy effects in a macroscopic ensemble of coupled microscopic machines. As an educated guess, we want to consider synchronization as the paradigmatic phase transition: That is the phenomenon that coupled units with different eigenfrequencies exhibit a spontaneous phase-locking to a global frequency [75]. This collective phenomenon was famously described by Huygens who experimentally observed that two pendulum clocks attached to a common support display an “odd kind of sympathy” [76], that is they synchronize in anti-phase. It was later found to be ubiquitous in nature, *e.g.* the synchronous flashing of fireflies or the cricket chirping [77]. Synchronization is typically modeled by coupled phase oscillators which exhibit phase-locking when the coupling strength exceeds a critical value [78]. The commonly used noisy Kuramoto (also Sakaguchi) model [56, 79, 80] is well understood for an infinite population of oscillators at the mean-field level. Some works also considered few locally coupled oscillators [81–83] and even the relationship between the dissipation and the synchronization of two coupled oscillators [84]. However, owing to their complexity, little is known about the thermodynamic features of systems consisting of a large ensemble of stochastic oscillators.

To this end, we will consider a minimal stochastic model made of interacting three-state units which exhibit phase synchronization [85–87]. By doing so, we will be analyzing how the performance of the network changes across the different scales and from the desynchronized to the synchronized phase. To further corroborate the findings made within the three-state model, the latter will be generalized to a class of q-state clock models as a subclass of these models also display a transition from an asynchronous to a synchronous phase. The first main achievement of this thesis is to demonstrate that, within that class of clock models, the overall performance of the macroscopic system made up of many interacting microscopic machines is enhanced in the synchronization regime. Another important aspect of this study will be to analyze in detail how a linear and irreducible Markovian stochastic dynamics can give rise to a nonlinear mean-field dynamics which exhibits a nonequilibrium phase transition with increasing system size. This question is particularly intriguing since the Perron-Frobenius theorem ensures that the former dynamics has a unique stationary solution while the latter can exhibit complex non-unique solutions [88]. We will see that this question is closely related to the presence of metastability that can be characterized via a gap in the spectrum of the generator of the Markovian stochastic dynamics [89–92].

The computation of thermodynamic efficiencies is not only meaningful for energy conversion processes but also in the context of information processing [21, 93]. For quasistatic processes, the so-called Landauer's principle [94] provides an explicit lower bound for the dissipative costs incurred when one bit of information is erased. In practice only finite-time processes are relevant. With the development of stochastic thermodynamics tools to study finite-time erasure processes became available and initiated several works, *e.g.* [95–98]. Of particular interest is the thermodynamics of finite-time information erasure in macroscopic bits which normally are composed of arrays of microscopic binary units. Decoding schemes prescribe how the collective physical information stored in an array of microscopic units is translated into logical information stored by the macroscopic bit. We will determine the thermodynamic performance of the information erasure in a logical macroscopic bit with and without employing a specific decoding scheme. Hence we will gather additional insights if collective effects, corresponding to the decoding procedure, are overall energetically beneficial for a thermodynamic operation on a macroscopic system made up of many microscopic units. This frames the scope of the first part of this thesis.

Thermodynamically Consistent Coarse Graining in Many-Body Systems

In order to compute the stochastic thermodynamic properties of non-equilibrium systems exhibiting complex emergent behavior, there is a need for methodologies to study the fluctuating quantities from microscopic to macroscopic scales in a thermodynamically consistent way. While many coarse-graining schemes preserving thermodynamic consistency have been proposed in the literature [99–111], the aforementioned issue has yet not been addressed. The second main achievement of this thesis is to identify all-to-all interactions as a physical limit for which the stochastic thermodynamics of many-body systems can

be equivalently formulated across microscopic, mesoscopic and macroscopic scales and to develop the methodology to do so: To this end, we will generalize the setup of the three- and q-state clock models towards a system of N identical and all-to-all interacting q-state units that are subjected to both autonomous and non-autonomous forces and that are coupled with multiple heat reservoirs.

Then, the *microscopic* stochastic dynamics, where the system is unambiguously characterized by many-body states, can be exactly coarse-grained to a *mesoscopic* stochastic dynamics, where the system is now fully characterized by the numbers of the units which occupy the different unit states. In the *macroscopic limit* ($N \to \infty$) the deterministic dynamics is now governed by a nonlinear mean-field rate equation for the most likely values of the occupation of each unit state. We will furthermore identify the rather generic conditions under which the stochastic thermodynamics is invariant under the dynamically exact coarse-graining. Using a path-integral representation of the mesoscopic stochastic dynamics, a macroscopic fluctuating thermodynamic theory is formulated. Its thermodynamic consistency is encoded in the detailed fluctuation theorem (1.1) for the fluctuations which scale exponentially with the number of units.

As a complement to the study of the all-to-all interacting q-state units, we will furthermore consider two interacting underdamped (inertia) Brownian particles and apply three different coarse-graining schemes that haven proven instrumental for overdamped (inertialess) systems: Marginalization over one particle [112, 113], bipartite structure with information flows [114–117] and the Hamiltonian of mean force formalism [118–120]. In particular, we will study the validity of these three coarse-graining methods in two distinct physical limits: First, in the limit of time-scale separation, where the second particle with a fast relaxation time scale locally equilibrates with respect to the coordinates of the first, slowly relaxing particle. Secondly, in the limit where the second particle becomes heavy and thus deterministic. These physical limits are shown to be compatible with an exact thermodynamics for a reduced stochastic dynamics if the appropriate coarse-graining method is chosen. In these cases, the coarse-grained degrees of freedom are shown to constitute a heat reservoir and a work source coupled with the effective system in the first and second limit, respectively. This frames the scope of the second part of this thesis.

Plan of this thesis

The thesis is structured as follows. In chapter 2 some theoretical background pertinent to the subsequent chapters is given. Chapter 3 studies how collective effects can affect the thermodynamic performance of a system. First, the results related to the thermodynamic performance of macroscopic assemblies of a large number of interacting microscopic machines exhibiting synchronization is presented. Secondly, as a complement to energy-conversion processes, the role of collective effects in information-erasure processes is discussed. Furthermore, chapter 4 provides a stochastic thermodynamic description from microscopic to macroscopic scales for many-body systems with all-to-all interactions in contact with multiple heat reservoirs and subjected to external forces. Moreover, the validity of several

coarse-graining methods, which have been established for jump processes in underdamped systems is assessed using a minimal model. This thesis is concluded by a summary and an outlook to potential follow-up projects in chapter 5.

Chapter 2

Preliminaries

It is the declared goal of this dissertation to be mostly self-contained. This chapter is therefore devoted to briefly introducing the concepts on which the following chapters, which contain the actual results of the research work done during the doctorate, are based on.

2.1 Traditional Thermodynamics

We start by reviewing the basis elements of traditional thermodynamics that were developed over the past two centuries by the initial works of for instance Carnot and Clausius as well as more recent works by Prigogine and others [8, 11, 12]. First, we make the crucial observation that a closed macroscopic system will reach a so-called equilibrium state that is characterized by a finite number of macroscopic state functions. Which state functions are relevant depends on the model under consideration. In the following, we restrict to the state functions energy E, volume $\mathbb{V}$ and the number of particles N. Secondly, the energy of the system can be varied via an external perturbation. Here, we distinguish between work W (corresponding to controllable energy exchanges) and heat contributions Q (corresponding to uncontrollable energy exchanges). From energy conservation follows the first law of thermodynamics

$$\mathrm{d}E = đW + đQ, \tag{2.1}$$

where we indicate via the notation d and $đ$ that energy is a state function, while heat and work depend on how the energy of the system is changed by the external perturbation.

Next, we introduce another state function called entropy S that depends on the other state

functions, *e.g.* $S = S(E, \mathbb{V}, N)$. This relation can be inverted to obtain the equivalent relation $E(S, \mathbb{V}, N)$. If the changes in the system are quasi-static, *i.e.* they are slowly enough for the system to remain at equilibrium throughout the transformation, the relation $E(S, \mathbb{V}, N)$ holds and one arrives via differentiation at the Gibbs relation

$$\mathrm{d}E = \frac{1}{\beta}\,\mathrm{d}S - \mathfrak{p}\,\mathrm{d}\mathbb{V} + \mu\,\mathrm{d}N. \tag{2.2}$$

Hence partial differentiation of the energy with respect to the state functions, S, $\mathbb{V}$ and N, yields the inverse temperature $1/T \equiv \beta$, minus the pressure $\mathfrak{p}$ and the chemical potential μ, respectively. Here, and throughout this thesis we set the Boltzmann constant $k_b \equiv 1$. A comparison with the first law (2.1) suggests the following identifications: $đW = -\mathfrak{p}\,\mathrm{d}\mathbb{V} + \mu\,\mathrm{d}N$ and $đQ = \mathrm{d}S/\beta$. If a constraint on the system is released, traditional thermodynamics can not characterize the relaxation process of the system as more variables and a microscopic description are needed therefor. Though, the final equilibrium state is characterized solely by state variables and thus described by traditional thermodynamics.

Since the thermodynamics of *nonequilibrium* phase transitions represents a central topic of this dissertation, we briefly introduce the concepts and terminology of the thermodynamics of *equilibrium* phase transitions [48, 49]. The equilibrium state of a macroscopic system, unambiguously characterized by a finite set of macroscopic state functions, is referred to as a phase. It can be observed that thermodynamic systems may exist in different phases whose macroscopic behavior can differ significantly under variation of external conditions. The phase that is realized in nature is the one that minimizes the Gibbs free energy

$$G = E + \mathfrak{p}\mathbb{V} - \frac{1}{\beta}\,\mathrm{d}S = \mu\,N, \tag{2.3}$$

at constant pressure and temperature. As the intensive variables of the system, *e.g.* the temperature β, are changed, they take values for which phase transitions can occur. At such critical points the chemical potentials of the different coexisting phases must be equal.

Hence at a phase transition the chemical potentials of the phases, and therefore the Gibbs free energy must change continuously. However, phase transitions are classified into two classes according to the behavior of the derivatives of the Gibbs free energy with respect to the temperature at its critical value, as illustrated in Fig. 2.1. First-order equilibrium phase transitions are accompanied by discontinuous first-order derivatives of the Gibbs free energy with respect to temperature, while second-order (also continuous) phase transitions display a continuous first- but possibly discontinuous second-and/or-higher-order derivatives of the Gibbs free energy.

Idealized systems, such as for instance heat reservoirs, are of particular interest. An ideal heat reservoir is defined as a thermodynamic system with an intrinsic time scale that is much faster than the typical time scale of the external perturbation such that it remains at equilibrium during thermodynamic transformations. We furthermore assume that these reservoirs have an infinite capacity, so that their inverse temperature β and chemical potential μ remain constant during any thermodynamic process. Let us now

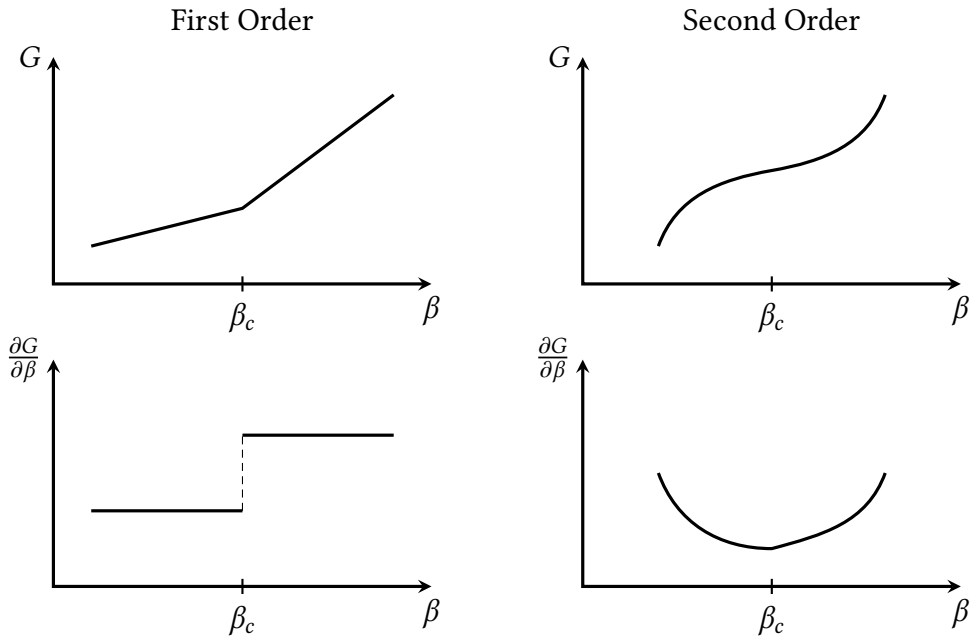

Figure 2.1: *Illustration of the typical functional behavior of the Gibbs free energy G in the vicinity of the critical temperature β_c for first- [left-hand plots] and second-order equilibrium phase transitions [right-hand plots].*

consider a system with entropy S weakly coupled to an ideal heat reservoir at inverse temperature β and chemical potential μ. The second law of thermodynamics reads

$$đ\Sigma = \mathrm{d}S - \beta\, đQ = \mathrm{d}S - đS_e \geq 0, \tag{2.4}$$

and states that the difference between the entropies of final and initial *global* equilibrium states is non-negative. The interpretation of Eq. (2.4) due to Prigogine has proven instrumental in the modern formulation of thermodynamics. Hereby central are the concepts of entropy production rate $đ\Sigma$ and entropy flow $đS_e$. The former is an entropy increase due to irreversible processes, while the latter is an entropic contribution from the reversible exchange with the ideal heat reservoir. Consequently, for a reversible processes the equality sign holds in Eq. (2.4), $\mathrm{d}S = đS_e$. There is an alternative formulation of the second law. For this purpose, we introduce the state function free energy

$$A = E - \frac{1}{\beta}S, \tag{2.5}$$

which along with Eq. (2.1) allows to rewrite Eq. (2.4) as follows

$$đ\Sigma = \beta\,(đW - \Delta A) \geq 0. \tag{2.6}$$

This alternative representation of the second law states that the amount of work required for a thermodynamic transformation of the system is at least equal to the change of its

free energy, where the equality sign holds for reversible processes. If $\Delta A < 0$, work can be extracted from the system, at most the maximum amount $-\Delta A$.

We remark that the second law (2.4) was originally derived using arguments against the existence of a perpetuum mobile of second kind. There are alternative formulations, *e.g.* that work can not be extracted solely from a single heat reservoir (Kelvin) or that heat does not spontaneously flow from a cold to a hot reservoir (Clausius). Based on these arguments, Carnot derived a universal bound for the efficiency of thermal machines. For this purpose, let us consider a quasi-static cyclic process of a system (the thermal machine) weakly coupled to a hot and a cold ideal heat reservoir at two inverse temperatures $\beta_h < \beta_c$, respectively. The thermodynamic cycle consists an uptake of heat Q_h from the hot reservoir and a heat release $-Q_c$ into the cold reservoir. The rest of the energy is converted into work done by the engine, $-W = Q_h - Q_c$. Since entropy is a state function, one has for the entropy production along a full cycle $\Sigma = \beta_h Q_h - \beta_c Q_c$, which in turn implies that the efficiency of the thermal engine is bounded by the Carnot efficiency

$$\eta_c = -\frac{W}{Q_h} \leq 1 - \frac{\beta_h}{\beta_c}. \tag{2.7}$$

When operating small engines fluctuations need to be taken into account. This, however, is not possible within the framework of traditional thermodynamics. To address this limitation the novel theory called stochastic thermodynamics has been developed over the past two decades. The central idea of this formalism is to consistently build a thermodynamic description on top of a Markov process. An introduction to the most fundamental concepts of stochastic thermodynamics is provided in the remainder of this chapter.

2.2 Stochastic Processes

2.2.1 Probability Theory

Before proceeding with the theory of stochastic thermodynamics, elementary stochastic methods as discussed in Refs. [88, 121, 122] are reviewed. These naturally rely on probabilistic concepts which are formulated hereafter. We introduce the probability $p(A)$ of a set A of events $x \in A$, that acts as a function of A. Moreover, we denote by $\oslash$ and Ω the set of no and all events, respectively. Then, the probability $p(A)$ satisfies the following axioms:

$$p(A) \geq 0 \ \forall A \tag{2.8}$$

$$p(\Omega) = 1 \tag{2.9}$$

$$p\Big(\bigcup_i A_i\Big) = \sum_i p(A_i), \quad \text{if } A_i \cap A_j = \oslash \ \ \forall i, j \tag{2.10}$$

$$p(\overline{A}) = 1 - p(A) \tag{2.11}$$
$$p(\oslash) = 0, \tag{2.12}$$

where $\overline{A}$ is the complement of A, *i.e.* the set of all events that are not contained in A. We note that, in fact, the last two equations follow from the first three equations. Hence, there are only three axioms required to formulate a consistent probabilistic theory.

Furthermore, the joint probability $p(\{A, B\})$ corresponds to the probability that the event x is contained in the set A and the set B, thus

$$p(A \cap B) = p(\{(x \in A) \text{ and } (x \in B)\}). \tag{2.13}$$

It is often of interest to consider a subset of the events in A given that they these events are also in the set B. We define the conditional probability as

$$p(A|B) = \frac{p(A \cap B)}{p(B)}, \tag{2.14}$$

and since one can define the conditioning in both directions, the following relation

$$p(\{A, B\}) = p(A|B)p(B) = p(B|A)p(A) \Leftrightarrow p(A|B) = \frac{p(B|A)p(A)}{p(B)}, \tag{2.15}$$

referred to as Baye's theorem, holds.

Central to stochastic processes is the concept of a random variable $X(x)$. The mean value of a random variable $X(x)$ is given by

$$\langle X \rangle = \sum_x p_x X(x), \tag{2.16}$$

where p_x is the probability of the set $A = \{x\}$ that only contains the single event x. Higher order moments, $\langle X^n \rangle$, are also of interest. In particular, the variance

$$\sigma(X) \equiv \langle \big(X - \langle X \rangle\big)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2, \tag{2.17}$$

measures the degree of which the values of X deviate from their mean value $\langle X \rangle$. For multivariate stochastic processes, one defines the covariance as

$$\langle (X_i, X_j) \rangle = \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle. \tag{2.18}$$

The moment generating function is defined as follows

$$g(\gamma) \equiv \left\langle \mathrm{e}^{-\gamma X} \right\rangle, \tag{2.19}$$

and its derivatives with respect to the so-called counting field γ and subsequent evaluation at $\gamma = 0$ yield the moments of the probability distribution

$$\langle X^n \rangle = (-1)^n \frac{\partial^n}{\partial \gamma^n} g(\gamma) \bigg|_{\gamma=0}, \quad n \in \mathbb{N}. \tag{2.20}$$

2.2.2 Markov Processes

Stochastic processes describe systems that evolve probabilistically in time, that is they are characterized by a time-dependent random variable $X(t)$. Suppose the random variable takes the values $x_n, \ldots, x_2, x_1$ at times $t_n, \ldots, t_2, t_1$,. Then, the joint probability density

$$p\left(x_n(t_n); \ldots; x_2(t_2); x_1(t_1)\right) \tag{2.21}$$

unambiguously describes the time-evolution of the system. According to Eq. (2.14), the conditional probability distribution can be written in terms of joint probabilities as follows

$$p(x_n(t_n), \ldots | x_2(t_2), x_1(t_1)) = \frac{p(x_n(t_n), \ldots | x_2(t_2), x_1(t_1))}{p(x_2(t_2), x_1(t_1))}, \tag{2.22}$$

where we use the convention that the time increases from right to left, $t_n \geq \ldots \geq t_2 \geq t_1$. Throughout this work we will only consider Markov processes. These are processes for which the conditional probability is determined entirely by the most recent condition, *i.e.*

$$p(x_n(t_n), \ldots | x_2(t_2), x_1(t_1)) = p(x_n(t_n), \ldots | x_2(t_2)). \tag{2.23}$$

The Markov assumption is a strong property as it physically states that the stochastic system has no memory about its past and as it furthermore mathematically allows us to write any joint probability simply as a product of two-point conditional probabilities and the initial condition

$$p(x_n(t_n), \ldots | x_2(t_2), x_1(t_1)) = p(x_n(t_n) | x_{n-1}(t_{n-1})) \cdot \ldots \cdot p(x_2(t_2) | x_1(t_1)) \cdot p_{x_1}(t_1)\,. \tag{2.24}$$

The Markov property, in turn, implies the Chapman-Kolmogorov equation,

$$p(x_3(t_3), x_1(t_1)) = \sum_{x_2} p(x_3(t_3), x_2(t_2), x_1(t_1)) = \sum_{x_2} p(x_3(t_3) | x_2(t_2)) \cdot p(x_2(t_2) | x_1(t_1)). \tag{2.25}$$

While the Chapman-Kolmogorov Eq. (2.25) is intuitive, it is often more practical to consider its differential formulation. Since there is no need of carefully ordering the time instances, we can ease the notation and write $p(i, t + \mathrm{d}t | j, t)$, for the conditional probability to find the system in state i at time $t + \mathrm{d}t$ given that it was in state j at time t. For an infinitesimal change in time $\mathrm{d}t$, we can linearize the conditional probability,

$$p(i, t + \mathrm{d}t | j, t) = V_{ij}\, \mathrm{d}t + \mathbb{O}(\mathrm{d}t^2), \quad i \neq j, \tag{2.26}$$

Here, the quantity

$$V_{ij} = \lim_{\mathrm{d}t \to 0} \frac{p(i, t + \mathrm{d}t | j, t)}{\mathrm{d}t}, \quad i \neq j, \tag{2.27}$$

denotes the probability per unit time to observe a transition from j to i. Thus, $V_{ij}\,\mathrm{d}t$ gives the probability that a transition from j to i occurs during the time interval $\mathrm{d}t$ and since $1 - \mathrm{d}t \sum_{i\neq j} V_{ij}$ is the probability that no transition is made at all during $\mathrm{d}t$, we have

$$p(i, t + \mathrm{d}t|j, t) = V_{ij}\,\mathrm{d}t + \Big[1 - \mathrm{d}t \sum_{i\neq j} V_{ij}\Big]\,\delta_{ij} + \mathbb{O}(\mathrm{d}t^2). \tag{2.28}$$

Using the Chapman-Kolmogorov Eq. (2.25), we obtain

$$p(i, t + \mathrm{d}t) = \sum_j p(i, t + \mathrm{d}t|j, t)\,p(j, t) = \mathrm{d}t \sum_j V_{ij}\,p(j, t) + \Big[1 - \mathrm{d}t \sum_{j\neq i} V_{ij}\Big]\,p(i, t) + \mathbb{O}(\mathrm{d}t^2), \tag{2.29}$$

so that in the limit $\mathrm{d}t \to 0$, we arrive at the differential Chapman-Kolmogorov equation

$$\partial_t p_i(t) = \sum_j \big[V_{ij}\,p_j(t) - V_{ji}\,p_i(t)\big], \tag{2.30}$$

that is referred to as Markovian master equation in the following. To further ease notation, we write the state as a subscript of the probability in the master equation. The interpretation of Eq. (2.30) as a probability-current balance equation is obvious. Conversely, the latter can be rewritten in a more compact form,

$$\partial_t p_i(t) = \sum_j W_{ij}\,p_j(t)\,, \tag{2.31}$$

with the transition rate matrix

$$W_{ij} = V_{ij} - \delta_{ij} \sum_k V_{kj}. \tag{2.32}$$

The transition rate matrix W_{ij} is stochastic[1], *i.e.* $\sum_i W_{ij} = 0$. The system is in a stationary state if $\sum_j W_{ij}\,p_j^s = 0$, and if there are furthermore no net-currents the detailed balance condition is satisfied,

$$W_{ij}\,p_j^{eq} = W_{ji}\,p_i^{eq}, \tag{2.33}$$

which corresponds to the system being at equilibrium characterized by the distribution p^{eq}.

So far, it was assumed that the state space is discrete. In case of stochastic processes that are characterized by random variables with a continuous range, the derivation of the differential Chapman-Kolmogorov Equation (2.30) remains conceptionally the same as above and is thus omitted. Within the scope of this thesis, we will restrict the statistics

[1]The author is indeed aware that, strictly speaking, this term actually refers to the sum of all entries along a row being equal to unity. In the following, we will however use this term synonymously to the property of each column summing up to identical zero.

of continuous random variables to the first two moments. In this case, the differential continuous-state Chapman-Kolmogorov equation can be truncated in order to obtain the so-called Fokker-Planck equation

$$\dot{\rho}(x,t) = -\partial_x \left[\mu(x)\rho(x,t)) + \partial_x\{D(x)\,\partial_x\rho(x,t)\}\right] = -\partial_x J(x,t), \tag{2.34}$$

that governs the time evolution of the probability density $\rho(x,t)$. Here, μ is the drift, D is the diffusion coefficient and J denotes the probability current. Note that, for simplicity, we assumed that there is only one random variable. For multiple random variables, the Fokker-Planck Eq. (2.34) becomes a multivariate equation. An equivalent description of the stochastic process is given by the Langevin equation

$$\dot{x}(t) = \mu(x,t) + \xi(t), \tag{2.35}$$

where ξ denotes Gaussian white noise[2],

$$\langle\xi(t)\rangle = 0, \quad \langle\xi(t)\xi(t')\rangle = 2D\delta(t-t'), \tag{2.36}$$

If for instance x represents the velocity of a Brownian particle, the stochastic dynamics (2.35) is called overdamped (intertialess), whereas for $\boldsymbol{x} = (x,v)^\top$ being an array that contains position x and velocity v of the particle, the stochastic dynamics is called underdamped.

2.3 Stochastic Thermodynamics

2.3.1 Ensemble Thermodynamics

The key idea of stochastic thermodynamics [13–18] is to consistently build the thermodynamic description on top of a Markov process, turning the thermodynamic observables known from traditional thermodynamics (cf. Sec. 2.1) into stochastic processes. First, we focus on the ensemble level before proceeding with fluctuations. To this end, we consider an open system in contact with several ideal heat reservoirs $\nu = 1, 2, \ldots, L$ at inverse temperature $\beta^{(\nu)}$. The system has discrete states $i = 1, 2, 3$ of energy $e_i(\lambda_t)$ that are externally driven according to a known protocol λ_t. In addition, there are nonconservative forces $f_{ij}^{(\nu)}$ that can induce or oppress transitions from state j to state i. For generality, we assume that the force can differ for a transition between to given states depending on which bath ν exchanges energy with the system during that transition. An example system is depicted in Fig. 2.2. Next, we make the crucial assumption that the heat bath is weakly coupled to the system. In this case, the heat reservoir remains always in equilibrium regardless how far the system

[2] If not truncated, the continuous Master equation would be equivalent to a Langevin equation with Poisson noise.

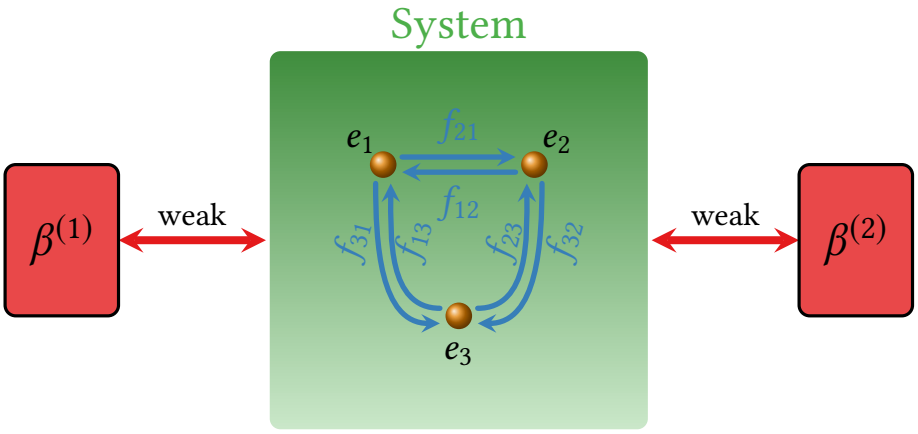

Figure 2.2: *Illustrative example of a driven three-state system weakly coupled to two ideal heat reservoirs at inverse temperature $\beta^{(1)}$ and $\beta^{(2)}$. For better visualization, the distinction of the nonconservative forces $f^{(\nu)}$ in terms of the heat resrevoirs is omitted, here.*

is driven out of equilibrium [123, 124]. As a consequence, the stochastic dynamics of the system is governed by a Markovian master equation (2.31), where we denote the probability to be in state i by p_i and write $W_{ij}^{(\nu)}(\lambda_t)$ for the transition rate from j to i. Furthermore, the transition rates satisfy the so-called *local* detailed balance condition

$$\frac{W_{ij}^{(\nu)}(\lambda_t)}{W_{ji}^{(\nu)}(\lambda_t)} = \mathrm{e}^{-\beta^{(\nu)}\left[e_i(\lambda_t)-e_j(\lambda_t)-f_{ij}^{(\nu)}\right]}, \tag{2.37}$$

ensuring thermodynamic consistency. If the driving is switched off, $\lambda_t = \lambda$, the system will inevitably relax towards a unique stationary state, $p^s(\lambda)$, in the long-time limit. If furthermore the heat reservoirs are at equal temperature, $\beta^{(\nu)} = \beta$, and all nonconservative forces vanish, $f_{ij}^{(\nu)} = 0$, stationary distribution coincides with the equilibrium one. The latter satisfies the detailed balance condition (2.33), and because of Eq. (2.37) in turn assumes canonical form,

$$p_i^{eq}(\lambda) = \mathrm{e}^{-\beta\,[e_i(\lambda)-A^{eq}(\lambda)]}, \tag{2.38}$$

with the equilibrium free energy

$$A^{eq}(\lambda) = -\frac{1}{\beta}\ln\sum_i \mathrm{e}^{-\beta e_i(\lambda)}. \tag{2.39}$$

For completeness, we add that if the system is also open for particle exchange, the equilibrium distribution assumes a grand canonical form,

$$p_i^{eq}(\lambda) = \mathrm{e}^{-\beta[e_i(\lambda)-\mu N_i-\Phi^{eq}(\lambda)]}, \tag{2.40}$$

with the equilibrium landau potential

$$\Phi^{eq}(\lambda) = -\frac{1}{\beta} \ln \sum_i \mathrm{e}^{-\beta[e_i(\lambda)-\mu N_i]}. \tag{2.41}$$

The average energy of the system is naturally defined as the ensemble average over state energies

$$\langle E\rangle(t) = \sum_i e_i(\lambda_t)\, p_i(t), \tag{2.42}$$

and its time-derivative stipulates the first law of thermodynamics

$$\mathrm{d}_t\langle E\rangle(t) = \langle \dot{Q}\rangle(t) + \langle \dot{W}\rangle(t), \tag{2.43}$$

with the average heat current from the reservoir to the system

$$\begin{aligned}\langle \dot{Q}\rangle(t) &= \sum_{\nu,i,j} (e_i(\lambda_t) - e_j(\lambda_t) - f_{ij}^{(\nu)})\; W_{ij}^{(\nu)}(\lambda_t)\, p_j(t) \\ &= -\frac{1}{\beta} \sum_{\nu,i,j} W_{ij}^{(\nu)}(\lambda_t)\, p_j(t)\;\; \ln \frac{W_{ij}^{(\nu)}(\lambda_t)}{W_{ji}^{(\nu)}(\lambda_t)} = \sum_\nu \langle \dot{Q}^{(\nu)}\rangle(t),\end{aligned} \tag{2.44}$$

and the average work current

$$\langle \dot{W}\rangle(t) = \sum_i \Big[\dot{\lambda}_t\, \partial_{\lambda_t}\, e_i(\lambda_t)\, p_i(t) + \sum_{\nu,,j} f_{ij}^{(\nu)}\, W_{ij}^{(\nu)}(\lambda_t)\, p_j(t)\Big]. \tag{2.45}$$

We identify the Shannon entropy

$$\langle S\rangle(t) = -\sum_i p_i(t) \ln p_i(t), \tag{2.46}$$

as a physically meaningful choice for the average nonequilibrium system entropy. Its rate of change

$$\mathrm{d}_t\langle S\rangle(t) = -\sum_{\nu,i,j} \left(\ln \frac{W_{ji}^{(\nu)}(\lambda_t) p_i(t)}{W_{ij}^{(\nu)}(\lambda_t) p_j(t)} + \ln \frac{W_{ij}^{(\nu)}(\lambda_t)}{W_{ji}^{(\nu)}(\lambda_t)} \right) W_{ij}^{(\nu)}(\lambda_t)\, p_j(t) = \langle \dot{S}_e\rangle(t) + \langle \dot{\Sigma}\rangle(t), \tag{2.47}$$

can be decomposed into the average entropy flow from the bath to the system

$$\langle \dot{S}_e\rangle(t) = -\sum_{\nu,i,j} \ln \frac{W_{ij}^{(\nu)}(\lambda_t)}{W_{ji}^{(\nu)}(\lambda_t)}\, W_{ij}^{(\nu)}(\lambda_t)\, p_j(t) = \sum_\nu \beta^{(\nu)} \langle \dot{Q}^{(\nu)}\rangle(t), \tag{2.48}$$

and the average entropy production rate

$$\langle \dot{\Sigma} \rangle(t) = \sum_{\nu,i,j} \ln \frac{W_{ij}^{(\nu)}(\lambda_t) p_j(t)}{W_{ji}^{(\nu)}(\lambda_t) p_i(t)} W_{ij}^{(\nu)}(\lambda_t)\, p_j(t) \geq 0, \tag{2.49}$$

whose nonnegativity is proved using $\ln x \leq 1-x$. Hence Eq. (2.47) constitutes the second law of thermodynamics. If the system in equilibrium, the entropy production rate is identically zero according to Eq. (2.33) and is therefore a measure of irreversibility.

Combining the first two laws of thermodynamics for an isothermal system, $\beta^{(\nu)} = \beta$, one obtains for the average non-equilibrium free energy

$$\mathrm{d}_t \langle A \rangle(t) = \mathrm{d}_t \langle E \rangle(t) - \frac{1}{\beta} \mathrm{d}_t \langle S \rangle(t) = \langle \dot{W} \rangle(t) - \frac{1}{\beta} \langle \dot{\Sigma} \rangle(t). \tag{2.50}$$

It follows from the non-negativity of the entropy production rate that the maximum average output work, $-\langle W \rangle$, that can be extracted during a thermodynamic process is bounded by the average change in free energy, $\Delta \langle A \rangle$ during this transformation. It can be shown that Eq. (2.50) reproduces at equilibrium the well-known properties of the equilibrium state. Setting $p_i = p_i^{eq}$, we indeed obtain that

$$A^{eq}(\lambda) = E^{eq}(\lambda) - \frac{1}{\beta} S^{eq}(\lambda). \tag{2.51}$$

An interesting property of the average non-equilibrium free energy is that it is always larger than its equilibrium value,

$$\langle A \rangle(t) - A^{eq}(\lambda) = \frac{1}{\beta} \sum_i p_i(t) \ln \frac{p_i(t)}{p_i^{eq}(\lambda)} \equiv \frac{1}{\beta} \langle I \rangle(t) \geq 0. \tag{2.52}$$

Here, we defined the average mutual information $\langle I \rangle$ to be equal to the nonnegative relative entropy (also Kullback-Leibler divergence) defined as

$$\mathrm{D}(x||y) \equiv \sum_i x_i \ln \frac{x_i}{y_i} \geq 0, \tag{2.53}$$

measuring the distance between the distributions $p(t)$ and $p^{eq}(\lambda)$. The nonnegativity of the relative entropy is again proved using $\ln x \leq 1 - x$.

We conclude this section with a remarkable result found for the efficiency of energy conversion processes: Let us assume that there are two thermodynamics forces F_1 and F_2 present that induce the probability currents J_1 and J_2 such that the mean stationary entropy production rate reads

$$\langle \dot{\Sigma}^s \rangle(\lambda) = \langle \dot{\Sigma}_1^s \rangle(\lambda) + \langle \dot{\Sigma}_2^s \rangle(\lambda) = F_1^s(\lambda) \langle J_1^s \rangle(\lambda) + F_2^s(\lambda) \langle J_2^s \rangle(\lambda). \tag{2.54}$$

A thermodynamic system that on average operates as a machine uses a fueling process flowing in the direction of its associated force $\langle \dot{\Sigma}_1^s \rangle(\lambda) > 0$ (e.g. heat flowing down a

temperature gradient or a generic driving force) in order to power a second process that flows against the direction of its associated force $\langle \dot{\Sigma}_2^s \rangle(\lambda) < 0$ (e.g. coordinate moving against the direction of a mechanical force).

Owing to the nonnegativity of the entropy production rate, a meaningful definition for the efficiency of that stationary energy conversion process reads

$$\eta^s = -\frac{\langle \dot{\Sigma}_1^s \rangle(\lambda)}{\langle \dot{\Sigma}_2^s \rangle(\lambda)} \leq \eta_c, \tag{2.55}$$

where the Carnot efficiency η_c is achieved in the reversible limit. Since the latter corresponds to zero power generation and is thus not of any practical use, one considers instead the efficiency at maximum power η^*. As an example, the Curzon-Ahlborn efficiency that holds for *macroscopic* semi-ideal heat engines, where the heat transfer is the only irreversible step during a thermodynamic cycle, has the appealing form [125]

$$\eta^* = 1 - \sqrt{1 - \eta_c} = \frac{\eta_c}{2} + \frac{\eta_c^2}{8} + \mathbb{O}(\eta_c^3). \tag{2.56}$$

Using the formalism of stochastic thermodynamics, it can be shown that the term $\eta_c/2$ is universal for tightly-coupled *microscopic* systems, *i.e.* systems that only have one net current, $J_1 \propto J_2$, and that respond linearly [43]. Strikingly, the second term, $\eta_c^2/8$ is shown to be also universal for systems exhibiting a left-right symmetry [27].

2.3.2 Trajectory Thermodynamics

2.3.2.1 Stochastic First Law and Entropy Balance

We will now proceed with developing a framework to describe the thermodynamics for a single realization of the Markov process. To this end, the thermodynamic laws need to be formulated for a single trajectory. Generically, a trajectory is denoted by $\boldsymbol{m}_{(\tau)}(t)$. This notation corresponds to the specification of the actual state in the time interval under consideration, $\boldsymbol{m}_{(\tau)}(t)$, $t \in [t_0, t_f]$. Here, τ is a parametrization of the trajectory specifying the initial state $\boldsymbol{m}_{(\tau)}(t_0) = \alpha_0$, the subsequent jumps from α_{j-1} to α_j as well as the heat reservoir ν_j involved at the instances of time, $t = \tau_j$, $j = 1, \dots M$, and the final state, $\boldsymbol{m}_{(\tau)}(t_f) = \alpha_M$, where M is the total number of jumps. More explicitly, we write

$$\boldsymbol{m}_{(\tau)} = \left\{ \boldsymbol{m}_0 \xrightarrow{\nu_1,\tau_1} \boldsymbol{m}_1 \xrightarrow{\nu_2,\tau_2} \boldsymbol{m}_2 \xrightarrow{\nu_3,\tau_3} \dots \xrightarrow{\nu_j,\tau_j} \boldsymbol{m}_j \xrightarrow{\nu_{j+1},\tau_{j+1}} \dots \boldsymbol{m}_{M-1} \xrightarrow{\nu_M,\tau_M} \boldsymbol{m}_M \right\}, \tag{2.57}$$

and refer to Fig. 2.3 a) for an illustrative example of such a stochastic trajectory.
In the following, we will use lower scripts to label trajectory-dependent quantities and write $o[\boldsymbol{m}_{(\tau)}, t]$ for the value the observable o takes at time t for the trajectory $\boldsymbol{m}_{(\tau)}$. We define the

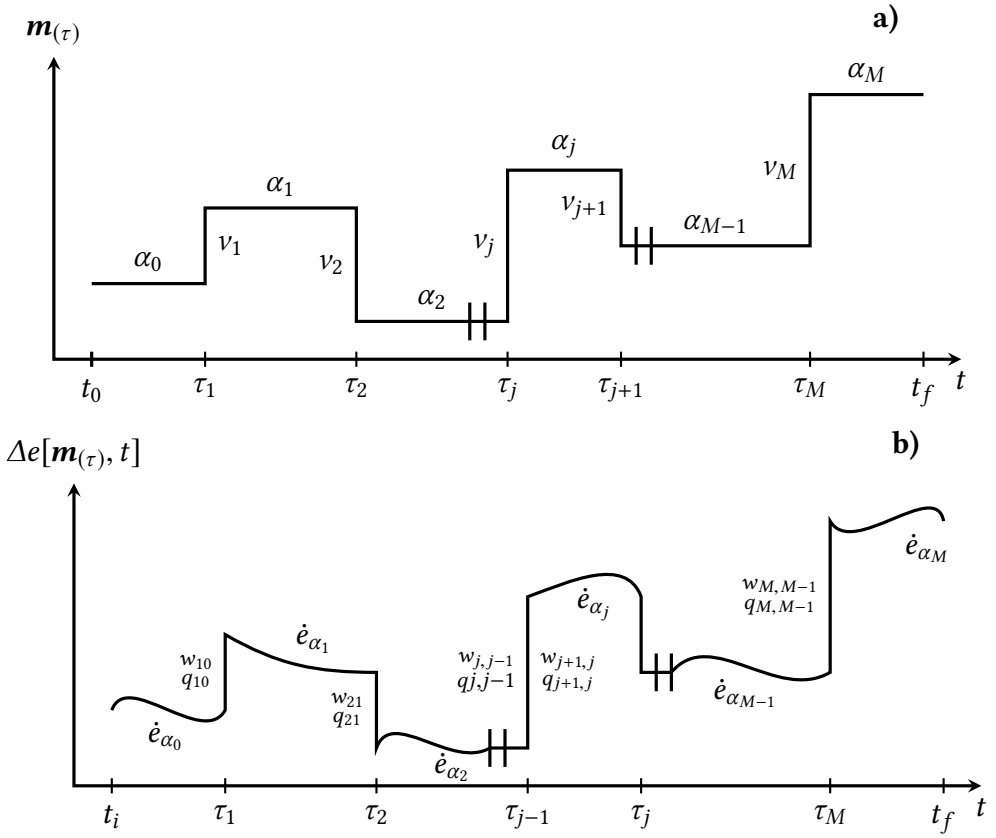

Figure 2.3: *Schematic representation of a single trajectory $\boldsymbol{m}_{(\tau)}$ during the time $[t_0, t_f]$ in a) and the corresponding time-integrated first law of thermodynamics for the same trajectory in b).*

energy associated with the trajectory at time t to be given by the energy of the particular state m_j the system is in for the trajectory under consideration, *i.e.*

$$e[\boldsymbol{m}_{(\tau)}, t] = \sum_m e_m(\lambda_t)\, \delta_{m,\boldsymbol{m}_{(\tau)}(t)}, \tag{2.58}$$

where the Kronecker delta $\delta_{m,\boldsymbol{m}_{(\tau)}(t)}$ selects the state α in which the trajectory is at time t. The stochastic energy is a state function,

$$\Delta e[\boldsymbol{m}_{(\tau)}, t] = \sum_m \left[e_m(\lambda_t)\, \delta_{m,\boldsymbol{m}_{(\tau)}(t)} - e_m(\lambda_0)\, \delta_{m,\boldsymbol{m}_{(\tau)}(0)} \right], \tag{2.59}$$

as indicated by the notation Δe, and its time-derivative[3] can be decomposed as follows,

$$\mathrm{d}_t e[\boldsymbol{m}_{(\tau)}, t] = \dot{q}[\boldsymbol{m}_{(\tau)}, t] + \dot{w}[\boldsymbol{m}_{(\tau)}, t], \tag{2.60}$$

with the stochastic heat

$$\begin{aligned}\dot{q}[\boldsymbol{m}_{(\tau)}, t] &= \sum_{\nu=1}^{L}\sum_{j=1}^{M} \delta(\nu-\nu_j)\delta(t-\tau_j)\left[e_{m_j}(\lambda_{\tau_j}) - e_{m_{j-1}}(\lambda_{\tau_j}) - f^{(\nu_j)}_{m_j,m_{j-1}}\right] \\ &= \sum_{\nu=1}^{L} -\sum_{j=1}^{M} \delta(\nu-\nu_j)\delta(t-\tau_j)\,\frac{1}{\beta^{(\nu_j)}}\,\ln\frac{W^{(\nu_j)}_{m_j,m_{j-1}}(\lambda_{\tau_j})}{W^{(\nu_j)}_{m_{j-1},m_j}(\lambda_{\tau_j})} = \sum_{\nu=1}^{L}\dot{q}^{(\nu)}[\boldsymbol{m}_{(\tau)}, t],\end{aligned} \tag{2.61}$$

and the stochastic work currents

$$\begin{aligned}\dot{w}[\boldsymbol{m}_{(\tau)}, t] &= \sum_m \left[\dot{\lambda}_t\,\partial_{\lambda_t}\,e_m(\lambda_t)\right]\delta_{m,\boldsymbol{m}_{(\tau)}(t)}\Big|_{\boldsymbol{m}_{(\tau)}(t)} + \sum_{\nu=1}^{L}\sum_{j=1}^{M}\delta(\nu-\nu_j)\,\delta(t-\tau_j)f^{(\nu_j)}_{m_j,m_{j-1}} \\ &= \dot{w}_\lambda[\boldsymbol{m}_{(\tau)}, t] + \sum_{\nu=1}^{L}\dot{w}_f^{(\nu)}[\boldsymbol{m}_{(\tau)}, t].\end{aligned} \tag{2.62}$$

The notation $\dot{x}|_{\boldsymbol{m}_{(\tau)}(t)}$ corresponds to the instantaneous and smooth changes of x along the horizontal segments of the trajectory $\boldsymbol{m}_{(\tau)(t)}$ in Fig. 2.3a). Since it will be instrumental further below, we split the fluctuating work current into the contribution $\dot{w}_\lambda[\boldsymbol{m}_{(\tau)}, t]$ from the nonautonomous driving and the dissipative contribution $\sum_{\nu=1}^{L}\dot{w}_f^{(\nu)}[\boldsymbol{m}_{(\tau)}, t]$ due to the nonconservative forces. Eq. (2.60) is the stochastic first law and ensures energy conservation at the trajectory level [126]. Fig. 2.3 b) illustrates the time-integrated stochastic first law for the trajectory in a).

We now want to formulate the second law of thermodynamics for a single trajectory. To do so, we need a consistent definition of entropy trajectorywise, which appears to be an oxymoron, since entropy is considered as a property related to an ensemble. To this end, we will use a common idea in information theory [127], where one assigns an entropy to each event taking place quantifying the amount of surprise as $-\ln p_m$, upon observing the outcome m when its ensemble probability is p_m. There is no surprise when $p_m = 1$ and the surprise becomes infinite as $p_m \to 0$. After many repetitions of the experiment, the average surprise is equal to the Shannon entropy defined as the system entropy (2.46). The stochastic system entropy is thus defined as follows [20]

$$s[\boldsymbol{m}_{(\tau)}, t] = -\sum_m \ln p_m(t)\;\delta_{m,\boldsymbol{m}_{(\tau)}(t)}, \tag{2.63}$$

[3] To determine the time-derivative of the Kronecker delta, we realize that $\dot{\delta}_{m,\boldsymbol{m}_{(\tau)}(t)}$ goes from 0 to 1, and from 1 to 0, when $\boldsymbol{m}_{(\tau)}$ jumps into, or out of the microstate m, respectively, at time t. Hence the time-derivative consists of a sum of delta functions, with weights 1 and -1, respectively, centered at the times, τ_j, of the jumps.

and is therefore also a state-function,

$$\Delta s[\boldsymbol{m}_{(\tau)}, t] = -\sum_m \left[\ln p_m(t)\, \delta_{m,\boldsymbol{m}_{(\tau)}(t)} - \ln p_m(0)\, \delta_{m,\boldsymbol{m}_{(\tau)}(0)} \right]. \tag{2.64}$$

Its time-derivative

$$\mathrm{d}_t\, s[\boldsymbol{m}_{(\tau)}, t] = \dot{s}_e[\boldsymbol{m}_{(\tau)}, t] + \dot{\sigma}[\boldsymbol{m}_{(\tau)}, t], \tag{2.65}$$

can be split into the stochastic entropy flow

$$\dot{s}_e[\boldsymbol{m}_{(\tau)}, t] = \sum_{\nu=1}^{L} - \sum_{j=1}^{M} \delta(\nu - \nu_j)\, \delta(t - \tau_j)\, \ln \frac{W^{(\nu_j)}_{m_j,m_{j-1}}(\lambda_{\tau_j})}{W^{(\nu_j)}_{m_{j-1},m_j}(\lambda_{\tau_j})} = \sum_{\nu=1}^{L} \beta^{(\nu)}\, \dot{q}^{(\nu)}[\boldsymbol{m}_{(\tau)}, t], \tag{2.66}$$

and the stochastic entropy production rate

$$\dot{\sigma}[\boldsymbol{m}_{(\tau)}, t] = -\sum_m \frac{\partial_t p_m(t)}{p_m(t)} \delta_{m,\boldsymbol{m}_{(\tau)}(t)}\Big|_{\boldsymbol{m}_{(\tau)}(t)} + \sum_{\nu=1}^{L}\sum_{j=1}^{M} \delta(\nu - \nu_j)\delta(t - \tau_j) \ln \frac{W^{(\nu_j)}_{m_j,m_{j-1}}(\lambda_{\tau_j}) p_{m_{j-1}}(t)}{W^{(\nu_j)}_{m_{j-1},m_j}(\lambda_{\tau_j}) p_{m_j}(t)}. \tag{2.67}$$

We note that Eq. (2.65) corresponds to the entropy balance at the trajectory level.

It will also prove useful to also consider the time-integrated stochastic first law

$$\Delta e[\boldsymbol{m}_{(\tau)}, t] \equiv \sum_{\nu=1}^{L} \delta e^{(\nu)}[\boldsymbol{m}_{(\tau)}, t] = \delta q[\boldsymbol{m}_{(\tau)}, t] + \delta w[\boldsymbol{m}_{(\tau)}, t], \tag{2.68}$$

with the time-integrated fluctuating energy current

$$\delta e^{(\nu)}[\boldsymbol{m}_{(\tau)}, t] = \int_0^t \mathrm{d}t' \sum_{j=1}^{M} \delta(\nu - \nu_j)\delta(t' - \tau_j)[e_{m_j} - e_{m_{j-1}}], \tag{2.69}$$

as well as the fluctuating heat

$$\begin{aligned} \delta q[\boldsymbol{m}_{(\tau)}, t] &= \sum_{\nu=1}^{L} - \int_0^t \mathrm{d}t' \sum_{j=1}^{M} \delta(\nu - \nu_j)\delta(t' - \tau_j) \frac{1}{\beta^{(\nu_j)}} \ln \frac{W^{(\nu_j)}_{m_j,m_{j-1}}(\lambda_{\tau_j})}{m^{(\nu_j)}_{m_{j-1},m_j}(\lambda_{\tau_j})} \\ &= \sum_{\nu=1}^{L} \left(\delta e^{(\nu)}[\boldsymbol{m}_{(\tau)}, t] - \delta w_f^{(\nu)}[\boldsymbol{m}_{(\tau)}, t] \right) = \sum_{\nu=1}^{L} \delta q^{(\nu)}[\boldsymbol{m}_{(\tau)}, t], \end{aligned} \tag{2.70}$$

and the fluctuating work

$$\delta w[\boldsymbol{m}_{(\tau)}, t] = \underbrace{\int_0^t \mathrm{d}t' \sum_m \left[\dot{\lambda}_{t'} \partial_{\lambda_{t'}} e_m(\lambda_{t'}) \right] \delta_{m,\boldsymbol{m}_{(\tau)}(t')}\Big|_{\boldsymbol{m}_{(\tau)}(t')}}_{\delta w_\lambda[\boldsymbol{m}_{(\tau)}, t]} + \underbrace{\sum_{\nu=1}^{L} \int_0^t \mathrm{d}t' \sum_{j=1}^{M} \delta(\nu - \nu_j)\, \delta(t' - \tau_j) f^{(\nu_j)}_{m_j,m_{j-1}}}_{\delta w_f^{(\nu)}[\boldsymbol{m}_{(\tau)}, t]}. \tag{2.71}$$

Using Eqs. (2.67) and (2.70), the entropy production can be expressed as

$$\begin{aligned}\delta\sigma[\boldsymbol{m}_{(\tau)},t] &= \ln\frac{p_{m_0}(0)}{p_{m_M}(t)} + \sum_{\nu=1}^{L}\int_0^t \mathrm{d}t' \sum_{j=1}^{M}\delta(\nu-\nu_j)\delta(t'-\tau_j)\,\ln\frac{W^{(\nu_j)}_{m_j,m_{j-1}}(\lambda_{\tau_j})}{W^{(\nu_j)}_{m_{j-1},m_j}(\lambda_{\tau_j})} \\ &= \ln\frac{p_{m_0}(0)}{p_{m_M}(t)} - \sum_{\nu=1}^{L}\beta^{(\nu)}\delta q^{(\nu)}[\boldsymbol{m}_{(\tau)},t].\end{aligned} \tag{2.72}$$

It is easy to verify that the ensemble average of both the differential or time-integrated stochastic observables at the trajectory level coincides the corresponding definitions at the ensemble level, see Sec. 2.3.1 It is important to note that the stochastic entropy production rate is not always non-negative. Further below it will become clear that it is in fact a necessity to observe trajectories with a negative entropy production assigned to them.

2.3.2.2 Generating Function Techniques

The generating functions encode the entire statistics of the different stochastic observables according to Eq. (2.20). The generating function related to the change $\delta o[\boldsymbol{m}_{(\tau)},t]$ of the observable o along a trajectory $\boldsymbol{m}_{(\tau)}$ up to time t is defined as

$$g(\gamma_o,t) \equiv \left\langle \mathrm{e}^{-\gamma_o\,\delta o[\boldsymbol{m}_{(\tau)},t]}\right\rangle = \int_{\boldsymbol{m}_{(\tau)}} \mathcal{D}[\boldsymbol{m}_{(\tau)}]\,\mathcal{P}[\boldsymbol{m}_{(\tau)}]\,\mathrm{e}^{-\gamma_o\,\delta o[\boldsymbol{m}_{(\tau)},t]}. \tag{2.73}$$

where γ_o is the counting field (also bias) for the observable o, $\mathcal{P}[\boldsymbol{m}_{(\tau)}]$ denotes the path probability of the trajectory $\boldsymbol{m}_{(\tau)}$ and $\mathcal{D}[\boldsymbol{m}_{(\tau)}]$ refers to the associated measure for the trajectory contributing to the path integral. The probability to observe a change $\delta o[\boldsymbol{m}_{(\tau)},t] = \delta o$ along a trajectory $\boldsymbol{m}_{(\tau)}$ at time t, can be determined from the generating function using Eq. (2.73) as follows

$$p(\delta o,t) = \langle\delta(\delta o - \delta o[\boldsymbol{m}_{(\tau)},t])\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}\gamma_o\,\mathrm{e}^{\mathrm{i}\gamma_o\,\delta o}\left\langle\mathrm{e}^{\mathrm{i}\gamma_o\,\delta o[\boldsymbol{m}_{(\tau)},t]}\right\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}\gamma_o\,\mathrm{e}^{\mathrm{i}\gamma_o\delta o}\,g(\mathrm{i}\gamma_o,t). \tag{2.74}$$

Inversion of the last equation yields

$$g(\mathrm{i}\gamma_o,t) = \int_{-\infty}^{\infty}\mathrm{d}\,(\delta o)\;\mathrm{e}^{-\mathrm{i}\,\gamma_o\,\delta o}\,p(\delta o,t) \tag{2.75}$$

The moments of the distribution are obtained as in Eq. (2.20)

$$\left\langle\delta o^l[\boldsymbol{m}_{(\tau)},t]\right\rangle = (-1)^l\;\frac{\partial^l g^{(o)}(\gamma_o,t)}{\partial\gamma_o^l}\bigg|_{\gamma_o=0},\quad l\in\mathbb{N}. \tag{2.76}$$

It will prove useful to rewrite the GF as follows

$$g(\gamma_o, t) = \sum_m g_m(\gamma_o, t) \tag{2.77}$$

with

$$g_m(\gamma_o, t) = p_m(t)\left\langle \mathrm{e}^{-\gamma_o\, \delta o[\boldsymbol{m}_{(\tau)}, t]}\right\rangle_m = p_m(t) \int_{\boldsymbol{m}_{(\tau)}} \mathcal{D}[\boldsymbol{m}_{(\tau)}]\, \mathcal{P}[\boldsymbol{m}_{(\tau)}|m, t]\, \mathrm{e}^{-\gamma_o\, \delta o[\boldsymbol{m}_{(\tau)}, t]}, \tag{2.78}$$

where $\langle\cdot\rangle_m$ denotes an ensemble average over all trajectories $\boldsymbol{m}_{(\tau)}$ that are in state m at time t so that $\mathcal{P}[\boldsymbol{m}_{(\tau)}|m, t] = \mathcal{P}[\boldsymbol{m}_{(\tau)}]\, \delta_{\boldsymbol{m}_{(\tau)}, m}(t)$ is the conditional path probability.

Comparison with Eqs. (2.73) and (2.75) yields

$$g_m(\gamma_o, t) = \int_{\delta o} \mathrm{d}\,(\delta o)\; \mathrm{e}^{-\gamma_o\, \delta o}\, p(\delta o, m, t) \tag{2.79}$$

where $p(\delta o, m, t)$ denotes the joint probability to observe a variation δo up to time t while the system is in state m at that time. From Eq. (2.78) we easily verify that $g_m(\gamma_o, 0) = p_m(0)$. Furthermore, we have $g(0, t) = 1$ and $g_m(0, t) = p_m(t)$. With Eq. (2.77) we can write the time-derivative of the generating function as

$$\partial_t g(\gamma_o, t) = \sum_m \partial_t g_m(\gamma_o, t). \tag{2.80}$$

In order to derive the equations of motions for $\partial_t g_m(\gamma_o, t)$, we consider the joint probability $p_m(\delta o, t + \mathrm{d}t)$ for observing a change in the observable $\delta o = \delta o[\boldsymbol{m}_{(\tau)}, t]$ along the trajectory $\boldsymbol{m}_{(\tau)}$ and to find the trajectory in state m at time $t + \mathrm{d}t$. This probability, to its leading order in $\mathrm{d}t$, is given by

$$\begin{aligned} p_m(\delta o, t + \mathrm{d}t) = {} & p_m\big(\delta o - \tfrac{\partial o_m}{\partial t}\mathrm{d}t, t\big)(1 + W_{m,m}(\lambda_t)\, \mathrm{d}t) + \\ & + \sum_{\nu=1}^{L} \sum_{m' \neq m} p_{m'}\Big(\delta o - o^{(\nu)}_{m,m'}, t\Big) W^{(\nu)}_{m,m'}(\lambda_t)\, \mathrm{d}t + \mathbb{O}(\mathrm{d}t^2), \end{aligned} \tag{2.81}$$

where $W_{m,m}(\lambda_t) = -\sum_{m' \neq m} W_{m',m}(\lambda_t)$ as prescribed by the definition (2.32). The first term on the right hand side of the last equation corresponds to the scenario that the system remains in state m during the time interval $[t, t + \mathrm{d}t]$ while the observable changes by an amount $\frac{\partial o_m}{\partial t}\, \mathrm{d}t$. In contrast, the second term accounts for all possible transitions from state m' into the state m during the time interval $[t, t + \mathrm{d}t]$ such that the observable is changed by an amount $o^{(\nu)}_{m,m'}$. Using Eq. (2.79), we have

$$\begin{aligned} & g_m(\gamma_o, t + \mathrm{d}t) = \\ & = \int_{\delta o} \mathrm{d}(\delta o)\Big[\mathrm{e}^{-\gamma_o \delta o} p_m\big(\delta o - \tfrac{\partial o_m}{\partial t}\mathrm{d}t, t\big)\, \big(1 + W_{m,m}\mathrm{d}t\big) \; + \end{aligned}$$

$$
\begin{aligned}
&+ \sum_{\nu=1}^{L} \sum_{m' \neq m} \mathrm{e}^{-\gamma_o \delta o} \delta(\nu - \nu_j)\, p_{m'} \Big(\delta o - o^{(\nu_j)}_{m,m'}, t \Big) W^{(\nu_j)}_{m,m'}(\lambda_t) \mathrm{d}t \Big] + \mathbb{O}(\mathrm{d}t^2) \\
&= \int_{\delta o} \mathrm{d}(\delta o) \Big[\mathrm{e}^{-\gamma_o \left(\delta o + \frac{\partial o_m}{\partial t} \mathrm{d}t \right)} p_m(\delta o, t) \left(1 + W_{m,m} \mathrm{d}t \right) + \\
&+ \sum_{\nu=1}^{L} \sum_{m' \neq m} \mathrm{e}^{-\gamma_o \left(\delta o + \delta o^{(\nu_j)}_{m,m'} \right)} \delta(\nu - \nu_j) p_{m'}(\delta o, t) W_{m,m'}(\lambda_t)\, \mathrm{d}t \Big] + \mathbb{O}(\mathrm{d}t^2) \\
&= \mathrm{e}^{-\gamma_o \frac{\partial o_m}{\partial t} \mathrm{d}t} \left(1 + W_{m,m}\, \mathrm{d}t \right) g_m(\gamma_o, t) + \\
&+ \sum_{\nu=1}^{L} \sum_{m' \neq m} \mathrm{e}^{-\gamma_o\, \delta o^{(\nu_j)}_{m,m'}}\, W^{(\nu_j)}_{m,m'}(\lambda_t)\, \mathrm{d}t\; g_{m'}(\gamma_o, t) + \mathbb{O}(\mathrm{d}t^2) \\
&= \left(1 - \gamma_o \tfrac{\partial o_m}{\partial t}\, \mathrm{d}t \right) g_m(\gamma_o, t) + \sum_{\nu=1}^{L} \sum_{m'} \mathrm{e}^{-\gamma_o\, \delta o^{(\nu_j)}_{m,m'}}\, W^{(\nu_j)}_{m,m'}(\lambda_t)\, \mathrm{d}t\; g_{m'}(\gamma_o, t) + \mathbb{O}(\mathrm{d}t^2). \qquad (2.82)
\end{aligned}
$$

The last equation can be rearranged as follows,

$$
\frac{g_m(\gamma_o, t + \mathrm{d}t) - g_m(\gamma_o, t)}{\mathrm{d}t} = -\gamma_o \tfrac{\partial o_m}{\partial t}\, g_m(\gamma_o, t) + \sum_{\nu=1}^{L} \sum_{m'} \mathrm{e}^{-\gamma_o\, \delta o^{(\nu)}_{m,m'}}\, W^{(\nu)}_{m,m'}\, g_{m'}(\gamma_o, t) + \mathbb{O}(\mathrm{d}t), \qquad (2.83)
$$

so that in the limit $\mathrm{d}t \to 0$ we arrive at an equation of motion which takes the form of a biased microscopic master equation [128],

$$
\partial_t g_m(\gamma_o, t) = \sum_{m'} W_{mm'}(\gamma_o, \lambda_t)\, g^{(o)}_{m'}(\gamma, t), \qquad (2.84)
$$

with the generator of the biased stochastic dynamics

$$
W_{mm'}(\gamma_o, \lambda_t) = -\gamma_o\, \dot{o}_m(\lambda_t)\, \delta_{m,m'} + \sum_{\nu=1}^{L} \mathrm{e}^{-\gamma_o\, o^{(\nu)}_{mm'}(\lambda_t)}\, W^{(\nu)}_{mm'}(\lambda_t). \qquad (2.85)
$$

For state functions, $\delta o[\boldsymbol{m}_{(\tau)}, t] = \Delta o[\boldsymbol{m}_{(\tau)}, t] = o[\boldsymbol{m}_{(\tau)}, t] - o[\boldsymbol{m}_{(\tau)}, 0]$, the ensemble average over all trajectories in Eq. (2.73) reduces to an ensemble average with respect to the initial states of the trajectories only. Consequently, the generating function associated with any state function has the simple closed form

$$
g(\gamma_o, t) = \sum_{m,m'} \mathrm{e}^{-\gamma_o\, [o_m(t) - o_{m'}(0)]}\, p_m(t)\, p_{m'}(0). \qquad (2.86)
$$

Using Eqs. (2.58) and (2.63), we have for the generating functions associated with the state functions energy and entropy,

$$
g(\gamma_e, t) = \sum_{m,m'} \mathrm{e}^{-\gamma_e\, [e_m(\lambda_t) - e_{m'}(\lambda_0)]}\, p_m(t)\, p_{m'}(0) \qquad (2.87)
$$

$$g(\gamma_s, t) = \sum_{m,m'} \mathrm{e}^{\gamma_s \ln \frac{p_m(t)}{p_{m'}(0)}}\, p_m(t)\, p_{m'}(0). \tag{2.88}$$

Moreover, substituting Eqs. (2.61), (2.62), (2.66) and (2.67) into Eq. (2.85), we obtain for the generating functions associated with the currents

$$\partial_t g_m(\gamma_q, t) = \sum_{\nu=1}^{L} \sum_{m'} \mathrm{e}^{-\gamma_q \left[e_m(\lambda_t) - e_{m'}(\lambda_t) - f^{(\nu)}_{mm'}\right]}\, W^{(\nu)}_{mm'}(\gamma_q, \lambda_t)\, g_{m'}(\gamma_q, t) \tag{2.89}$$

$$\partial_t g_m(\gamma_w, t) = -\gamma_w\, \dot{\lambda}_t\, \left[\partial_{\lambda_t} e_m(\lambda_t)\right]\, g_m(\gamma_w, t) + \sum_{\nu=1}^{L} \sum_{m'} \mathrm{e}^{-\gamma_w f^{(\nu)}_{mm'}}\, W^{(\nu)}_{mm'}(\gamma_w, \lambda_t)\, g_{m'}(\gamma_w, t) \tag{2.90}$$

$$\partial_t g_m(\gamma_{s_e}, t) = \sum_{\nu=1}^{L} \sum_{m'} \mathrm{e}^{\gamma_{s_e} \ln \frac{W^{(\nu)}_{mm'}(\gamma_{s_e}, \lambda_t)}{W^{(\nu)}_{m'm}(\gamma_{s_e}, \lambda_t)}}\, W^{(\nu)}_{mm'}(\gamma_{s_e}, \lambda_t)\, g_{m'}(\gamma_{s_e}, t) \tag{2.91}$$

$$\partial_t g_m(\gamma_\sigma, t) = \gamma_\sigma\, \frac{\partial_t p_m(t)}{p_m(t)}\, g_m(\gamma_\sigma, t) + \sum_{\nu=1}^{L} \sum_{m'} \mathrm{e}^{\gamma_\sigma \ln \frac{W^{(\nu)}_{mm'}(\gamma_\sigma, \lambda_t) p_{m'}(t)}{W^{(\nu)}_{m'm}(-\gamma_\sigma, \lambda_t) p_m(t)}}\, W^{(\nu)}_{mm'}(\gamma_\sigma, \lambda_t)\, g_{m'}(\gamma_\sigma, t). \tag{2.92}$$

2.3.2.3 Fluctuation Theorems

It is easy to verify that $\partial_t g_m(\gamma_\sigma = 1, t) = p_m(t)$ satisfies the evolution equation (2.92), hence $g(\gamma_\sigma = 1, t) = 1$. From Eq. (2.73) the Seifert integral fluctuation theorem then ensues [20]

$$\left\langle \mathrm{e}^{-\delta\sigma[\boldsymbol{m}_{(\tau)}, t]} \right\rangle = 1, \tag{2.93}$$

which is valid irrespective of any conditions on initial and final state of the stochastic trajectory under consideration. It further follows from Jensen's inequality, $f(\langle x \rangle) \leq \langle f(x) \rangle$, applicable to any convex function $f(x)$, that the macroscopic second law of thermodynamics, $\langle \sigma \rangle \geq 0$, is reproduced.

For a single heat bath ν, in the absence of any nonconservative driving and for initial and final equilibrium distributions the stochastic entropy production (2.72) becomes

$$\delta\sigma[\boldsymbol{m}_{(\tau)}, t] = \beta \big\{ \Delta e[\boldsymbol{m}_{(\tau)}, t] - \delta q[\boldsymbol{m}_{(\tau)}, t] - \Delta A^{eq}(\lambda) \big\} = \beta \big\{ \delta w_\lambda[\boldsymbol{m}_{(\tau)}, t] - \Delta A^{eq}(\lambda) \big\}, \tag{2.94}$$

where $\Delta A^{eq}(\lambda) \equiv A^{eq}(\lambda_t) - A^{eq}(\lambda_0)$ is the change in total equilibrium free energy along the trajectory that only depends on the initial and final value of the driving protocol and thus does not fluctuate. In this case the Seifert fluctuation theorem (2.93) becomes the celebrated Jarzynski integral fluctuation theorem [129, 130]

$$\left\langle \mathrm{e}^{-\beta\, \delta w_\lambda[\boldsymbol{m}_{(\tau)}, t]} \right\rangle = \mathrm{e}^{-\beta\, \Delta A^{eq}(\lambda)}. \tag{2.95}$$

It is important to note that this is a *finite*-time fluctuation theorem because the fluctuating work stops evolving as the parameter of the driving remains at its final value λ_t such that the subsequent relaxation of the system towards the equilibrium distribution characterized by λ_t does not contribute to the work statistics.

Next, we want to derive a so-called detailed fluctuation theorem for the entropy production (2.72). Detailed fluctuation theorems are relations between a trajectory (forward process) and the time-reversed one (backward process). First we define the forward process where the the system is initially ($t' = 0$) in equilibrium with respect to the reference reservoir $\nu = 1$, that is

$$p^{eq}_{m_0}(\lambda_0) = \mathrm{e}^{-\beta^{(1)}[e_m(\lambda_0) - A^{eq}(\lambda_0)]} . \tag{2.96}$$

Such a state can be prepared by disconnecting all other heat reservoirs, switching off any driving with the fixed protocol value λ_0 and letting the system relax. At time $t' = 0$, all other heat reservoirs are simultaneously connected to the system and both the nonconservative driving f and the nonautonomous driving is switched on. As a result, the system evolves under the driven Markov process according to the forward protocol $\lambda_{t'}$, $t' \in [0, t]$ towards a nonequilibrium state $p_m(t)$. During that evolution heat is exchanged between the system and the reservoirs and there is mechanical work $\delta w_\lambda[\boldsymbol{m}_{(\tau)}, t]$ performed on the system by the external driving to change its energy $e_m(\lambda'_t)$. At time t, all heat reservoirs but the reference one $\nu = 1$ are disconnected and the force f is switched off such that system relaxes into an equilibrium state

$$p^{eq}_{m_M}(\lambda_t) = \mathrm{e}^{-\beta^{(1)}[e_m(\lambda_t) - A^{eq}(\lambda_t)]} . \tag{2.97}$$

Further, the backward process is indicated by the notation " ˜ " and constructed as follows. The initial state of the backward process is equal to the equilibrium state of the forward process evaluated at the final value of the forward-process protocol $\tilde{p}^{eq}_{m_0}(\tilde{\lambda}_0) = p^{eq}_{m_M}(\lambda_t)$ and all reservoirs are reconnected and the force f is turned on again. As a result, the system evolves under the time-reversed driven Markov process according to the backward protocol $\tilde{\lambda}_{t'} = \lambda_{t-t'}$, $t' \in [0, t]$ towards a nonequilibrium state $\tilde{p}_m(t)$. At time t all reservoirs except from the reference one $\nu = 1$ are disconnected and the nonconservative driving is turned off, which allows the system to relax towards an equilibrium state that coincides with the initial Gibbs state of the forward process, $\tilde{p}^{eq}_{m_M}(\tilde{\lambda}_t) = p^{eq}_{m_0}(\lambda_0)$. The definition of the forward and backward process is illustrated in Fig. 2.4.

For these processes, we can compute the path probability $\mathcal{P}$ for the trajectory $\boldsymbol{m}_{(\tau)}$ of the forward process and the path probability $\tilde{\mathcal{P}}$ for the trajectory $\tilde{\boldsymbol{m}}_{(\tau)}$ of the backward process. The path probability $\mathcal{P}$, is equal to the probability $p^{eq}_{m_0}(\lambda_0)$ of its initial Gibbs state times the probability to stay in this state until the first jump, times the probability to make this jump, and so on for the other jumps along the trajectory, with a final factor accounting for the trajectory to remain in its final state. The probability per unit time to make a jump from m_j to m_{j+1} at time t' is simply $W_{m_{j+1},m_j}(\lambda'_t)$. Conversely, the probability for not observing a

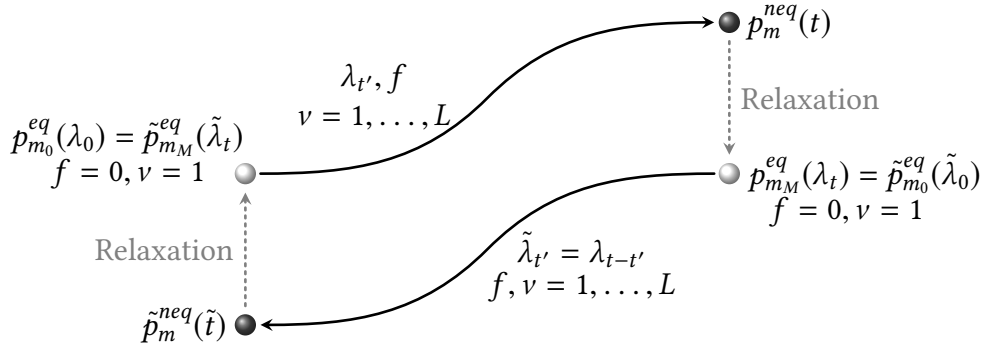

Figure 2.4: *Schematic representation of the backward and forward process.*

jump from state m during the time interval $[t_1, t_2]$

$$\lim_{n\to\infty} \prod_{i=1}^{n} \Big[1 - \sum_{m'\neq m} W_{m'm}(\lambda_{t'})\mathrm{d}t'\Big] = \lim_{n\to\infty} \prod_{i=1}^{n} [1 + W_{mm}(\lambda_{t'})\mathrm{d}t'] = \mathrm{e}^{\int_{t_1}^{t_2} \mathrm{d}t' \, W_{mm}(\lambda_{t'})} \tag{2.98}$$

where we used the discretization $t' = t_1 + i\,\mathrm{d}t'$ with $\mathrm{d}t' = (t_2 - t_1)/n$.

The backward path probability $\tilde{\mathcal{P}}$ is equal to the probability $\tilde{p}^{eq}_{m_0}(\tilde{\lambda}_0)$ of its initial Gibbs state times the probability to stay in this state until the first jump, times the probability to make this jump, and so on for the other jumps along the time-reversed trajectory, with a final factor accounting for the trajectory to remain in its final state. Clearly, the probability per unit time to make the backward jump from m_{j+1} to m_j at time t' is simply $W_{m_j,m_{j+1}}(\tilde{\lambda}'_t)$. We make the crucial observation that the probabilities for staying in the given states for given time-intervals are identical in the forward and backward process. Thus, if we consider the ration between forward and back path probability, these waiting times cancel and we arrive, by comparison with Eq. (2.72), at the following fundamental result

$$\ln \frac{\mathcal{P}(\boldsymbol{m}_{(\tau)})}{\tilde{\mathcal{P}}(\tilde{\boldsymbol{m}}_{(\tau)})} = \ln \frac{p^{eq}_{m_0}(\lambda_0)}{p^{eq}_{m_t}(\lambda_t)} + \sum_{\nu=1}^{L} \int_0^t \mathrm{d}t' \sum_{j=1}^{M} \delta(\nu - \nu_j)\delta(t' - \tau_j) \ln \frac{W^{(\nu_j)}_{m_j,m_{j-1}}(\lambda_{\tau_j})}{W^{(\nu_j)}_{m_{j-1},m_j}(\lambda_{\tau_j})} = \delta\sigma[\boldsymbol{m}_{(\tau)}, t]. \tag{2.99}$$

This relation is arguably the single most important in stochastic thermodynamics. It is fundamental as it implies the Seifert integral fluctuation theorem (2.93) by ensemble averaging over all forward trajectories. More strikingly, so-called detailed fluctuation theorems follow from this relation. Let $p(\delta\sigma)$ be the probability to observe a change in the entropy production by the amount $\delta\sigma$ along the forward process, $\delta\sigma[\boldsymbol{m}_{(\tau)}, t] = \delta\sigma$, and $\tilde{p}(-\delta\sigma)$ the probability to observe a change in the entropy production by the minus that amount along the time-reversed backward process, $\delta\sigma[\tilde{\boldsymbol{m}}_{(\tau)}, t] = -\delta\sigma$. Obviously, the former probability

can be calculated as follows:

$$\begin{aligned} p(\delta\sigma) &= \int\limits_{\boldsymbol{m}_{(\tau)}} \mathrm{d}\boldsymbol{m}_{(\tau)}\, \delta\left(\delta\sigma - \ln\frac{\mathcal{P}(\boldsymbol{m}_{(\tau)})}{\tilde{\mathcal{P}}(\tilde{\boldsymbol{m}}_{(\tau)})}\right)\mathcal{P}(\boldsymbol{m}_{(\tau)}) \\ &= \mathrm{e}^{\delta\sigma}\int\limits_{\tilde{\boldsymbol{m}}_{(\tau)}} \mathrm{d}\boldsymbol{m}_{(\tau)}\, \delta\left(-\delta\sigma - \ln\frac{\tilde{\mathcal{P}}(\tilde{\boldsymbol{m}}_{(\tau)})}{\mathcal{P}(\boldsymbol{m}_{(\tau)})}\right)\mathcal{P}(\boldsymbol{m}_{(\tau)}) = \mathrm{e}^{\delta\sigma}\,\tilde{p}(-\delta\sigma), \end{aligned} \tag{2.100}$$

where in the second equality we used that the Jacobian for the transformation from $\boldsymbol{m}_{(\tau)}$ to $\tilde{\boldsymbol{m}}_{(\tau)}$ is equal to unity. This so-called detailed Seifert fluctuation theorem is commonly written under the following form [20, 131]

$$\frac{p(\delta\sigma)}{\tilde{p}(-\delta\sigma)} = \mathrm{e}^{\delta\sigma}, \tag{2.101}$$

which in words states that the probability to observe a stochastic entropy increase $\delta\sigma$ along the forward process is exponentially more probable than to observe a corresponding decrease $-\delta\sigma$ along the backward process.

We emphasize that in order for minus the stochastic entropy production $-\delta\sigma$ to be the entropy production in the backward path, a class of special initial and final conditions of the trajectories is required. Indeed, upon a global time-reversal of Eq. (2.99), we have for the entropy production $\delta\tilde{\sigma}[\tilde{\boldsymbol{m}}_{(\tau)}, t]$ along the backward trajectory $\tilde{\boldsymbol{m}}_{(\tau)}$

$$\ln\frac{\tilde{\mathcal{P}}\left(\tilde{\boldsymbol{m}}_{(\tau)}\right)}{\tilde{\tilde{\mathcal{P}}}\left(\tilde{\tilde{\boldsymbol{m}}}_{(\tau)}\right)} = \delta\tilde{\sigma}[\tilde{\boldsymbol{m}}_{(\tau)}, t]. \tag{2.102}$$

Obviously, $\tilde{\tilde{\boldsymbol{m}}}_{(\tau)} = \boldsymbol{m}_{(\tau)}$, however, $\tilde{\tilde{\mathcal{P}}} = \mathcal{P}$ only if the initial (final) distribution of the forward trajectory matches with the final (initial) distribution of the backward trajectory. In general this will only be the case for stationary states. Since equilibrium states are a special case of stationary states, we conclude that for the forward and backward process as specified above we indeed have $\delta\tilde{\sigma}[\tilde{\boldsymbol{m}}_{(\tau)}, t] = -\delta\sigma$.

Using Eqs. (2.69), (2.96) and (2.97), the fluctuating entropy production (2.72) along the forward process can be rewritten as follows

$$\delta\sigma[\boldsymbol{m}_{(\tau)}, t] = \beta^{(1)}\left[\delta w_\lambda[\boldsymbol{m}_{(\tau)}, t] - \Delta A_1^{eq}\right] + \sum_{\nu=1}^{L}\left\{\beta^{(\nu)}\delta w_f^{(\nu)}[\boldsymbol{m}_{(\tau)}, t] + \left[\beta^{(1)} - \beta^{(\nu)}\right]\delta e^{(\nu)}[\boldsymbol{m}_{(\tau)}, t]\right\}. \tag{2.103}$$

Consequently, Eq. (2.101) implies the finite-time detailed fluctuation theorem [132, 133]

$$\frac{p\left(\beta^{(1)}\delta w_\lambda + \sum\limits_{\nu=1}^{L}\left[\beta^{(\nu)}\delta w_f^{(\nu)} + \left[\beta^{(1)} - \beta^{(\nu)}\right]\delta e^{(\nu)}\right]\right)}{\tilde{p}\left(-\beta^{(1)}\delta w_\lambda - \sum\limits_{\nu=1}^{L}\left[\beta^{(\nu)}\delta w_f^{(\nu)} + \left[\beta^{(1)} - \beta^{(\nu)}\right]\delta e^{(\nu)}\right]\right)} = \mathrm{e}^{\beta^{(1)}\left(\delta w_\lambda - \Delta A_1^{eq}\right) + \sum\limits_{\nu=1}^{L}\left[\beta^{(\nu)}\delta w_f^{(\nu)} + \left(\beta^{(1)} - \beta^{(\nu)}\right)\delta e^{(\nu)}\right]}. \tag{2.104}$$

In fact, the finite-time detailed fluctuation theorem also holds for the joint probability distribution,

$$\frac{p\Big(\beta^{(1)}\,\delta w_\lambda\,,\,\{\delta j_f^\nu\}\,,\,\{\delta j_e^\nu\}\Big)}{\tilde{p}\Big(-\beta^{(1)}\,\delta w_\lambda\,,\,-\{\delta j_f^\nu\}\,,\,-\{\delta j_e^\nu\}\Big)} = \mathrm{e}^{\beta^{(1)}\left(\delta w_\lambda - \Delta A_1^{eq}\right) + \sum_{\nu=1}^{L}\left[\beta^{(\nu)}\delta w_f^{(\nu)} + \left(\beta^{(1)}-\beta^{(\nu)}\right)\delta e^{(\nu)}\right]}, \tag{2.105}$$

where for brevity we use the following notation for the time-integrated currents

$$\{\delta j_f^\nu\} \equiv \Big(\beta^{(1)}\delta w_f^{(1)},\ldots,\beta^{(L)}\delta w_f^{(L)}\Big),\quad \{\delta j_e^\nu\} \equiv \Big([\beta^{(1)}-\beta^{(2)}]\delta e^{(2)},\ldots,[\beta^{(1)}-\beta^{(L)}]\delta e^{(L)}\Big). \tag{2.106}$$

Here, $p(\beta^{(1)}\delta w_\lambda, \{\delta j_{w_f}^\nu\}, \{\delta j_e^\nu\})$ is the probability to observe a nonautonomous driving contribution $\beta^{(1)}\delta w_\lambda$, time-integrated autonomous work currents $\{\delta j_{w_f}^\nu\}$ and time-integrated energy currents $\{\delta j_e^\nu\}$ along the forward process.

Conversely, $\tilde{p}(-\beta^{(1)}\delta w_\lambda, \{-\delta j_{w_f}^\nu\}, \{-\delta j_e^\nu\})$ is the probability to observe a nonautonomous driving contribution $-\beta^{(1)}\delta w_\lambda$, time-integrated autonomous work currents $\{-\delta j_{w_f}^\nu\}$ and time-integrated energy currents $\{-\delta j_e^\nu\}$ along the time-reversed backward process.

We stress that the detailed fluctuation theorem in Eqs. (2.104) and (2.105) are *finite*-time relations. Since all heat reservoirs except the reference one are disconnected, the nonconservative force is switched off and the parameter of the nonautonomous driving has reached its final value at the nonequilibrium state $p_m^{neq}(t)$, all the fluctuating quantities appearing in the argument of the probability distributions in the detailed fluctuation theorems stop evolving during the relaxation process to the final equilibrium distribution of the forward process, i.e. $\beta^{(1)}\delta w_\lambda = 0$ and $\beta^{(\nu)}\delta w_f^{(\nu)} = 0, [\beta^{(1)}-\beta^{(\nu)}]\delta e^{(\nu)} = 0,\ \forall \nu, t' > t$. The same naturally applies to the backward process. As a result, all fluctuating quantities appearing in the detailed fluctuation theorems can be measured along a single finite-time trajectory and allows to access the ensemble equilibrium free energy.

The validity of the detailed fluctuation theorem in Eq. (2.105) can be seen in two ways. First, we define the quantity $\delta X = \beta^{(1)}\delta w_\lambda + \sum_{\nu=1}^{L}\left[\beta^{(\nu)}\delta w_f^{(\nu)} + \left[\beta^{(1)}-\beta^{(\nu)}\right]\delta e^{(\nu)}\right]$. Its corresponding probability is readily determined via marginalization of the joint probability distribution in Eq. (2.105) and can be manipulated as follows,

$$\begin{aligned}
&p(\delta X) = \\
&= \int_{-\infty}^{\infty}\prod_\nu \mathrm{d}\big(\beta^{(1)}\delta w_\lambda\big)\,\mathrm{d}\big(\delta j_f^\nu\big)\,\mathrm{d}\big(\delta j_e^\nu\big)\,\delta\Big(\delta X - \beta^{(1)}\delta w_\lambda - \sum_{\nu=1}^{L}\big(\delta j_f^\nu + \delta j_e^\nu\big)\Big)\,p\Big(\beta^{(1)}\delta w_\lambda, \{\delta j_f^\nu\}, \{\delta j_e^\nu\}\Big) \\
&= \int_{-\infty}^{\infty}\prod_\nu \mathrm{d}\big(\delta j_f^\nu\big)\,\mathrm{d}\big(\delta j_e^\nu\big)\,p\Big(\delta X - \{\delta j_f^\nu\} - \{\delta j_e^\nu\}\,,\,\{\delta j_f^\nu\}\,,\,\{\delta j_e^\nu\}\Big)
\end{aligned}$$

$$= \mathrm{e}^{\delta X} \int_{-\infty}^{\infty} \prod_{\nu} \mathrm{d}\left(\delta j_f^\nu\right) \mathrm{d}\left(\delta j_e^\nu\right) \tilde{p}\Big(\{\delta j_f^\nu\} + \{\delta j_e^\nu\} - \delta X\,,\, -\{\delta j_f^\nu\}\,,\, -\{\delta j_e^\nu\}\Big) = \tilde{p}(-\delta X)\, \mathrm{e}^{\delta X}, \tag{2.107}$$

and thus indeed satisfies the detailed fluctuation theorem (2.104).

Secondly, the detailed fluctuation theorem (2.105) can also be derived via the following symmetry of the associated generating function

$$g\big(1-\gamma_\lambda, \{\gamma_f^\nu\}, \{\gamma_e^\nu\}, t\big) = \tilde{g}\big(1-\gamma_\lambda, \{1-\gamma_f^\nu\}, \{1-\gamma_e^\nu\}, t\big)\, \mathrm{e}^{-\beta^{(1)}\Delta A_1^{eq}(\lambda)}, \tag{2.108}$$

as demonstrated in appendix A.1.

There are some notable special cases of the detailed fluctuation relation (2.105). If the system is autonomous, $\delta w_\lambda = 0$, the fluctuation theorem simplifies to a time-integrated current one [134],

$$\frac{p\Big(\{\delta j_f^\nu\}\,,\, \{\delta j_e^\nu\}\Big)}{p\Big(\{-\delta j_f^\nu\}\,,\, \{-\delta j_e^\nu\}\Big)} = \mathrm{e}^{\sum_{\nu=1}^{L}\left[\beta^{(\nu)}\delta w_f^{(\nu)} + \left(\beta^{(1)}-\beta^{(\nu)}\right)\delta e^{(\nu)}\right]}. \tag{2.109}$$

On the other hand, in the presence of a single reservoir ν and in the absence of any nonconservative force, Eq. (2.105) reduces to the celebrated Crooks detailed fluctuation theorem for the nonautonomous work [135, 136]

$$\frac{p(\beta\delta w_\lambda)}{\tilde{p}(-\beta\delta w_\lambda)} = \mathrm{e}^{\beta\left(\delta w_\lambda - \Delta A_1^{eq}\right)}. \tag{2.110}$$

Finally, if the system is isothermal and driven by both a nonconservative and nonautonomous force, we obtain a generalized work fluctuation theorem

$$\frac{p(\beta\delta w_\lambda, \beta\delta w_f)}{\tilde{p}(-\beta\delta w_\lambda, -\beta\delta w_f)} = \mathrm{e}^{\beta\left(\delta w_\lambda + \delta w_f - \Delta A_1^{eq}\right)}. \tag{2.111}$$

Chapter 3

Enhancing The Thermodynamic Performance via Collective Effects

3.1 Introduction

This chapter represents the thematically first part of this thesis as it addresses the question of how collective effects can affect the thermodynamic performance of a system. First, we consider a minimal stochastic model made of N interacting three-state oscillator units which was shown in Refs. [85–87] to exhibit stable limit cycles in the occupation space indicative of phase synchronization. This model can be seen as a toy model for interacting molecular motors [137], enzymes [138, 139] or switches [140, 141]. In its original formulation, the three-state model is made of three unidirectional stochastic transitions, thus the crucial local detailed balance condition (2.37) is not valid and it is thus not thermodynamically consistent. Our first achievement is to demonstrate that the model can indeed be formulated in a thermodynamically consistent way, while the complex phenomena, such as synchronization, exhibited by the original model are retained.

We find that, at the mean-field level ($N \to \infty$), the system displays as a function of the inverse temperature three phases separated by two nonequilibrium phase transitions: a Hopf bifurcation separating a single stable fixed point phase from a stable limit cycle one, and an infinite-period bifurcation separating the limit-cycle phase from a phase with three stable fixed points. At equilibrium only one phase transition survives which separates a phase with a single stable fixed point from one with multiple stable fixed points via a saddle-node bifurcation.

Next, we provide a detailed understanding of how a linear and irreducible Markov dynamics for finite systems can give rise to the aforementioned complex mean-field dynamics

solutions with increasing system size. This question is particularly intriguing since the Perron-Frobenius theorem ensures that the former irreducible dynamics has a unique stationary solution for finite state spaces [88], while the latter exhibits multiple and time-periodic solutions. We reconcile this apparent contradiction via the spectrum of the Markovian dynamics generator which is shown to encode the information about the two bifurcations that are observed in the mean field. The mean-field dynamics is demonstrated to be characterized by the three eigenvalues with dominant real parts (the null one and a complex conjugated pair). At the Hopf bifurcation, a real-part gap between these eigenvalues and the remaining eigenvalues opens up, enabling the emergence of a metastable mean-field-like oscillatory dynamics over long times. As the second bifurcation is approached, this difference in real parts further increases while the imaginary parts of the dominant eigenvalues significantly drop causing the oscillations to vanish into three metastable fixed points. The fact that the real part of the most dominant complex conjugated eigenvalue pair converges to zero while the gap with respect to the real parts of all other nonzero eigenvalues becomes larger with increasing system size explains the emergence of the mean-field solutions as the perpetuation of the metastable states.

After proving the consistency of stochastic thermodynamics across scales - from the microscopic many-body level over the mesoscopic occupation level to the mean-field one - we analyze the dissipated work across the different dynamical regimes. We find that as a function of increasing inverse temperatures the transition towards synchronization is of first order while the outward transition is of second order. A crucial observation is that in the thermodynamic limit, interactions can significantly decrease the dissipated work per oscillator beyond the synchronization threshold and even more so after the second transition towards multistability. Furthermore, interactions in finite assemblies of oscillators enhance this effect in the former case but reduce it in the latter, in particular when the number of oscillators is too low to sustain a long-lasting metastable solution. Finally, we demonstrate that when operating as an energy converter, synchronization significantly enhances the power output per oscillator. Despite operating far-from-equilibrium, the efficiency at maximum power remains quite close to the linear-response prediction of $1/2$.

Furthermore, we generalize the three-state mean-field model to q-state mean-field clock models. We demonstrate that the two mean-field phases in the three-state model - a single stable and multiple stable fixed points - are prescribed by the thermodynamic consistency and thus universal in all q-state clock models in the high- and low-temperature limit, where the system is entropy- and energy-driven, respectively. In fact, we can even characterize the clock models beyond the universal low- and high-temperature stationary solution. Most importantly, we derive the q-dependent critical Hopf bifurcation temperature as a function of the number of spin states q. Numerically, we show that for even q this phase transition separates the high-temperature and the low-temperature phase, whereas for odd q there are stable limit cycles indicative of synchronization and thus implying the existence of a second phase transition. Since stochastic thermodynamics is built upon Markov dynamics, it follows that there are also two q-dependent universal classes of thermodynamic phenomenology: If q is odd, there are two non-equilibrium phase transition (as in the three-state model), while

for even q there is only one nonequilibrium phase transition. For completeness, we also investigate the dissipated work and the power-efficiency trade-off in the different phases and find that interactions reduce dissipation in all clock models. Conversely, the power-efficiency trade-off for energy transduction processes is only significantly lifted for odd-q models in the synchronization phase, hence further corroborating that collective dynamics can indeed enhance the thermodynamic performance of macroscopic system made up of many interacting microscopic machines.

Finally, we will study the role of collective effects in information processing. More explicitly, we consider a macroscopic bit composed of N noninteracting binary units and compute the thermodynamic costs of both reversible and irreversible information erasure in the bit. Hereby, it will be distinguished if the logical information stored in the macroscopic bit is translated from the underlying microscopic binary units one-to-one (corresponding to a single unit) or via a decoding scheme. The latter is formally a coarse-graining of microscopic physical information and practically increases the safety of the information erasure since the signal-to-noise ratio is proportional to the square-root of the microscopic units contained in a macroscopic bit [142]. The paradigm for the decoding scheme we will consider is the so-called majority-logic decoding, which is the simplest and most used coarse-graining method in information processing [143].

We show that for reversible erasure protocols, information erasure in single units is more efficient than symmetric majority-logic decoding. Conversely, we find that for finite-time erasure protocols the majority-logic decoding can accelerate the process of erasing a fixed amount of information or compress the minimal erasure error of a fixed-time erasure process. While these benefits in terms of speed and precision for most erasure processes come at the expense of a lower erasure efficiency, we show that, remarkably, the majority-logic decoding will however also be more efficient than a single-unit process when the erasing is fast or the erasure error is small. When imposing the optimal erasure protocol that minimizes the dissipated heat, we find that for the two unit models (Fermi- and Arrhenius-rates units), these advantages are preserved. Hence, we conclude that the majority-logic decoding lifts the trade-off between erasure speed, precision and efficiency when compared to a single unit.

3.2 Thermodynamics of Nonequilibrium Phase Transitions in Clock Models

3.2.1 Driven Three-State Model

3.2.1.1 Setup

We consider a system consisting of N three-state units with the energy levels ϵ_i ($i = 1, 2, 3$). Under the constraint of occupying the same state, the units interact globally via an interaction potential u/N. The system is furthermore subjected to a heat bath at inverse temperature β and to a constant non-conservative rotational forcing f, as depicted in figure 3.1.

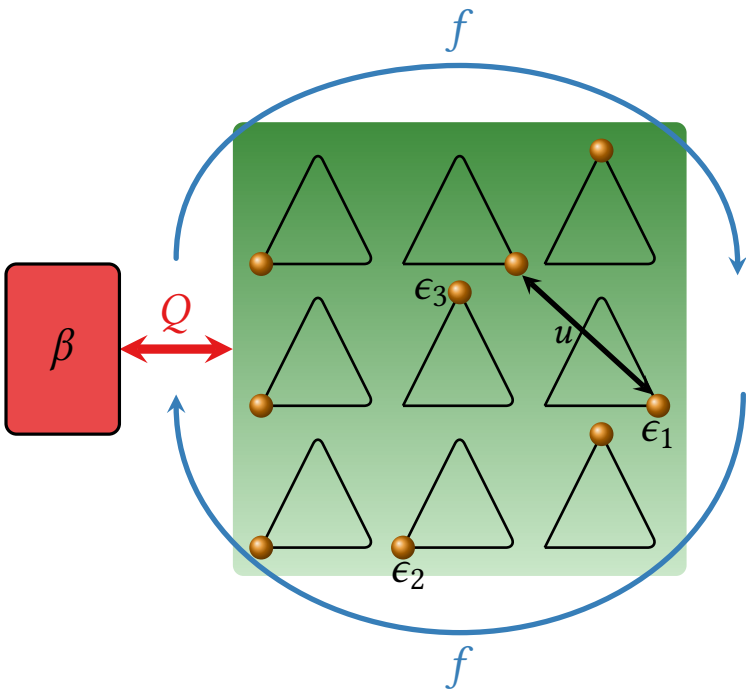

Figure 3.1: *Schematics of a small network of identical and all-to-all interacting three-state units in contact with a heat bath at inverse temperature β and in the presence of a nonconservative force f defined to act in clockwise direction for this given order of the single-unit states ($i \longrightarrow (i+1) \mod 3$).*

We denote a configuration by a multiindex $\alpha = (\alpha_1, \ldots, \alpha_i, \ldots, \alpha_N)$ with $\alpha_i = 1, 2, 3$. These, configurations, referred to as microstates in the following, entirely characterize the system. As an example, labeling the units from left to right and from top to bottom, the microstate displayed in Fig. 3.1 reads $\alpha = (2, 1, 3, 2, 3, 1, 2, 2, 3)$. Let us consider a transition from microstate $\alpha' \longrightarrow \alpha$ amounting to a transition between the single unit energy states $\epsilon_j \longrightarrow \epsilon_i$

and the occupation numbers $N_i \to N_i + 1$ and $N_j \to N_j - 1$. To determine the accompanying change in internal energy along such a transition, the total interaction energy U_α of the network for a given microstate α is required. Noting that determining the total interaction energy boils down to summing over the number of units occupying the state l with $l = 1, 2, 3$, one readily finds

$$U_\alpha(t) = \frac{u}{N} \sum_{k=1}^{3} \sum_{l=1}^{N_{k(\alpha)}(t)-1} l = \frac{u}{2N} \sum_{k=1}^{3} [N_{k(\alpha)}(t) + 1] N_{k(\alpha)}(t) = \frac{u}{2N} \sum_{k=1}^{3} N_{k(\alpha)}^2(t) + U_0, \tag{3.1}$$

where $U_0 = -u/2$ is a constant ($\Delta U_0 = 0$) and the notation $N_{k(\alpha)}(t)$ refers to the number of units occupying the state k for that microstate α at time t. Thus, we obtain for the internal energy

$$e_\alpha(t) = \sum_{k=1}^{3} \left\{ \epsilon_k \, N_{k(\alpha)}(t) + \frac{u}{2N} [N_{k(\alpha)}(t) + 1] N_{k(\alpha)}(t) \right\}. \tag{3.2}$$

and for the change in internal energy during the transition described above

$$e_\alpha(t) - e_{\alpha'}(t) = \epsilon_i - \epsilon_j + \frac{u}{N}(N_i(t) - N_j(t) + 1). \tag{3.3}$$

3.2.1.2 Master Equation

The stochastic jump process arising from the weak coupling of the system to the heat bath is governed by a Markovian master equation (2.31) describing the time evolution of the probability p_α to be in the microstate α as follows

$$\partial_t p_\alpha(t) = \sum_{\alpha'} w_{\alpha\alpha'} \, p_{\alpha'}(t) \,, \tag{3.4}$$

with the microscopic transition rates $w_{\alpha\alpha'}$. Restricting to one transition at a time, it holds that the transition rate matrix is irreducible and stochastic, $\sum_\alpha w_{\alpha\alpha'} = 0$, thus implying the existience of a unique steady state with probability p_α^s as stipulated by the Perron-Frobenius Theorem [88]. We take the microrates to be of Arrhenius form, that is

$$w_{\alpha\alpha'} = \Gamma \, \mathrm{e}^{-\frac{\beta}{2}[e_\alpha - e_{\alpha'} - f \, \Theta_{\alpha\alpha'}]} \,, \tag{3.5}$$

where the kinetic prefactor Γ sets the time-scale of the stochastic jump process. Moreover, the sign function $\Theta_{\alpha\alpha'}$ is defined as $\Theta_{\alpha\alpha'} = 1$ for $\sum_i \alpha_i - \alpha_i' = 1 \mathrm{mod} 3$, else $\Theta_{\alpha\alpha'} = -1$. We note that the microscopic rates satisfy local detailed balance

$$\ln \frac{w_{\alpha\alpha'}}{w_{\alpha'\alpha}} = -\beta \left[e_\alpha - e_{\alpha'} - f \, \Theta_{\alpha\alpha'} \right] , \tag{3.6}$$

where the oddness of the sign function $\Theta_{\alpha\alpha'} = -\Theta_{\alpha'\alpha}$ was used.

In the limit $t \to \infty$, the system will tend to the unique steady state, p^s_α, which is in non-equilibrium due to the non-autonomous driving f. Since there are no other thermodynamic forces present in this model, switching off the force at steady state leads to detailed balance

$$w_{\alpha\alpha'} p^{eq}_{\alpha'} = w_{\alpha'\alpha} p^{eq}_{\alpha} , \tag{3.7}$$

and along with local detailed balance in Eq. (3.6) implies the Gibbs distribution (2.38) as the equilibrium probability distribution

$$p^{eq}_\alpha = \mathrm{e}^{-\beta(e_\alpha - a^{eq})} , \tag{3.8}$$

with the equilibrium free energy

$$a^{eq} = -\frac{1}{\beta} \ln \sum_\alpha \mathrm{e}^{-\beta e_\alpha} . \tag{3.9}$$

3.2.1.3 Coarse-Graining

Formulating the stochastic process as above gives rise to an exceedingly large state space $||\alpha||$ growing exponentially with the number of units in the network as $||\alpha|| = 3^N$. Yet, a closer inspection reveals that the microscopic state space can be exactly marginalized into a mesoscopic one, where the system is unambiguously characterized by mesoscopic states defined as $\boldsymbol{N} \equiv (N_1, N_2, N - N_1 - N_2)$. This is due to the changes in internal energy (3.3) and thus the microscopic transition rates (3.5) which do not depend on the topological details of the network but only on the occupation numbers N_k before and after the transition. We write $\alpha_{\boldsymbol{N}}$ for a microstate α belonging to the mesostate $\boldsymbol{N}$.

The number of equienergetic microstates inside a mesostate, $e_{\alpha_{\boldsymbol{N}}} = E_{\boldsymbol{N}}$ is given by

$$\Omega_{\boldsymbol{N}} = \binom{N}{N_1}\binom{N - N_1}{N_2} = \frac{N!}{\prod_i N_i!} , \tag{3.10}$$

for a network that consists of N units.

We introduce the mesoscopic probability $P_{\boldsymbol{N}}$ to observe the mesostate $\boldsymbol{N}$ and introduce the conditional probability to find the system in a microstate α that belongs to that mesostate,

$$\mathbb{P}_{\alpha_{\boldsymbol{N}}}(t) = \frac{p_{\alpha_{\boldsymbol{N}}}(t)}{P_{\boldsymbol{N}}(t)} . \tag{3.11}$$

From probability conservation

$$\sum_{\alpha_{\boldsymbol{N}}} \mathbb{P}_{\alpha_{\boldsymbol{N}}}(t) = 1, \tag{3.12}$$

follows for the mesoscopic probability

$$P_{\boldsymbol{N}}(t) = \sum_{\alpha_{\boldsymbol{N}}} p_{\alpha_{\boldsymbol{N}}}(t). \tag{3.13}$$

Using the last equation and $e_{\alpha_{\boldsymbol{N}}}(t) = E_{\boldsymbol{N}}(t)$, we can marginalize the microscopic master equation (3.4) towards a mesoscopic one as follows

$$\partial_t P_{\boldsymbol{N}}(t) = \sum_{\alpha_{\boldsymbol{N}}} \sum_{\boldsymbol{N}'} \sum_{\alpha'_{\boldsymbol{N}'}} w_{\alpha_{\boldsymbol{N}},\alpha'_{\boldsymbol{N}'}}\, p_{\alpha'_{\boldsymbol{N}'}}(t) = \sum_{\boldsymbol{N}'} w_{\boldsymbol{N},\boldsymbol{N}'} \sum_{\alpha_{\boldsymbol{N}}} \sum_{\alpha'_{\boldsymbol{N}'}} p_{\alpha'_{\boldsymbol{N}'}}(t)\, \chi_{\alpha_{\boldsymbol{N}},\alpha'_{\boldsymbol{N}'}} = \sum_{\boldsymbol{N}'} W_{\boldsymbol{N}\boldsymbol{N}'}\, P_{\boldsymbol{N}'}(t)\,, \tag{3.14}$$

with the mesoscopic transition rates $W_{\boldsymbol{N}\boldsymbol{N}'} = \Omega_{\boldsymbol{N},\boldsymbol{N}'}\, w_{\boldsymbol{N},\boldsymbol{N}'}$. We note that the coarse-graining conserves the stochasticity and irreducibility of the mesoscopic transition rate matrix

$$\sum_{\boldsymbol{N}} W_{\boldsymbol{N}\boldsymbol{N}'} = 0, \tag{3.15}$$

The characteristic function $\chi_{\alpha_{\boldsymbol{N}},\alpha'_{\boldsymbol{N}'}}$ emerging in the second equality in Eq. (3.14) is a result of pulling the sums over the configurations, $\alpha_{\boldsymbol{N}}$ and $\alpha'_{\boldsymbol{N}'}$, through the transition rate matrix as the information that $w_{\alpha_{\boldsymbol{N}},\alpha'_{\boldsymbol{N}'}} \neq 0$ only for connected microstates $\alpha_{\boldsymbol{N}}$ and $\alpha'_{\boldsymbol{N}'}$ would be lost. Consequently, the function takes the value 1 if $\alpha_{\boldsymbol{N}}$ and $\alpha'_{\boldsymbol{N}'}$ are connected, and is 0 otherwise. To determine the constrained multiplicity factor $\Omega_{\boldsymbol{N},\boldsymbol{N}'}$, we need to calculate the number of microstates $\alpha(\boldsymbol{N})$ under the constraint that the microstates $\alpha_{\boldsymbol{N}}$ and $\alpha'_{\boldsymbol{N}'}$ are connected. Connectivity on the mesostate level means that two of three occupation numbers $(N_1, N_2, N - N_1 - N_2)$ of the two mesostates characterizing the transition differ by exactly one. Transitions between microstates are allowed if, compared entrywise, exactly one number in the tuples representing the two microstates specifying the transition is different. For instance, let us consider a transition $\boldsymbol{N}' \to \boldsymbol{N}$ with $N_1 = N'_1 + 1$ and $N_2 = N'_2 - 1$. The underlying logic of this combinatorial problem can be formulated as follows

$$\Omega_{\boldsymbol{N},\boldsymbol{N}'} = \binom{N - (N - N'_1 - N'_2) - N'_1}{1} \binom{N - (N - N'_1 - N'_2) - N'_1 - 1}{N'_2 - 1} \tag{3.16}$$

$$= \frac{N'_2!}{(N'_2 - 1)!} \cdot \frac{(N'_2 - 1)!}{(N'_2 - 1)!1!} = N'_2\,, \tag{3.17}$$

For the general case, we find that constrained multiplicty factor corresponds to the occupation number that is decremented during the transition, *i.e.*

$$\Omega_{\boldsymbol{N},\boldsymbol{N}'} = N'_1\, \delta_{N'_1,N_1+1} + N'_2\, \delta_{N'_2,N_2+1} + (N - N'_1 - N_2)\, \delta_{N-(N'_1+N'_2),N-(N_1+N_2)+1}\,. \tag{3.18}$$

Hence the coarse-graining of the dynamics is exact and leads indeed to a closed mesoscopic master equation (3.14) governing the time evolution of the mesoscopic probabilities $P_{\boldsymbol{N}}$.

Using Boltzmann's definition for entropy, we introduce the internal entropy of the mesostates due to their internal structure

$$S_{\boldsymbol{N}}^{int} = \ln \Omega_{\boldsymbol{N}}\,, \tag{3.19}$$

and the constrained change in internal entropy for the transition $\boldsymbol{N}' \longrightarrow \boldsymbol{N}$,

$$S_{\boldsymbol{N},\boldsymbol{N}'}^{int,c} = \ln \Omega_{\boldsymbol{N},\boldsymbol{N}'} \tag{3.20}$$

which accounts for all microstates belonging to the mesostate $\boldsymbol{N}'$ that are connected with a microstate inside the mesostate $\boldsymbol{N}'$. We can thus write the mesoscopic transition rates in Eq. (3.14) as follows

$$W_{\boldsymbol{N}\boldsymbol{N}'} = \Gamma\, \mathrm{e}^{-\frac{\beta}{2}\left[E_{\boldsymbol{N}} - E_{\boldsymbol{N}'} - \frac{2}{\beta} S_{\boldsymbol{N},\boldsymbol{N}'}^{int,c} - f\,\Theta_{\boldsymbol{N}\boldsymbol{N}'}\right]}, \tag{3.21}$$

where the mesoscopic sign function $\Theta_{\boldsymbol{N}\boldsymbol{N}'}$ is defined analogously to $\Theta_{\alpha,\alpha'}$ in Eq. (3.5). Explicitly, for a transition $N_i' \longrightarrow N_i = N_i' + 1$ and $N_j' \longrightarrow N_j = N_j' - 1$, it holds that $\Theta_{\boldsymbol{N}\boldsymbol{N}'} = 1$ if $i - j = 1 \mathrm{mod} 3$ and $\Theta_{\boldsymbol{N}\boldsymbol{N}'} = -1$ otherwise. The local detailed balance relation valid at the microscopic level (3.6) is preserved by the coarse-graining and reads at the mesoscopic level

$$\ln \frac{W_{\boldsymbol{N}\boldsymbol{N}'}}{W_{\boldsymbol{N}'\boldsymbol{N}}} = -\beta\left[\Delta A^{eq} - f\,\Theta_{\boldsymbol{N}\boldsymbol{N}'}\right]. \tag{3.22}$$

To arrive at this expression, we first note the relation

$$\ln \frac{\Omega_{\boldsymbol{N}}}{\Omega_{\boldsymbol{N}'}} = \ln \frac{\Omega_{\boldsymbol{N},\boldsymbol{N}'}}{\Omega_{\boldsymbol{N}',\boldsymbol{N}}}, \tag{3.23}$$

which in turn implies that

$$S_{\boldsymbol{N}}^{int} - S_{\boldsymbol{N}}^{int} = S_{\boldsymbol{N},\boldsymbol{N}'}^{int,c} - S_{\boldsymbol{N}',\boldsymbol{N}}^{int,c}\,. \tag{3.24}$$

Secondly, we used the oddness $\Theta_{\boldsymbol{N}\boldsymbol{N}'} = -\Theta_{\boldsymbol{N}'\boldsymbol{N}}$ and introduced the equilibrium free energy $A_{\boldsymbol{N}}^{eq} = E_{\boldsymbol{N}} - S_{\boldsymbol{N}}^{int}/\beta$ of the mesostate $\boldsymbol{N}$. The mesoscopic local detailed balance relation (3.22) stipulates that at $t \longrightarrow \infty$ and for $f = 0$ the mesoscopic equilibrium probability distribution assumes again canonical form

$$P_{\boldsymbol{N}}^{eq} = \mathrm{e}^{-\beta(A_{\boldsymbol{N}} - A^{eq})}, \tag{3.25}$$

with the equilibrium free energy

$$A^{eq} = -\frac{1}{\beta} \ln \sum_{\boldsymbol{N}} \mathrm{e}^{-\beta A_{\boldsymbol{N}}}\,. \tag{3.26}$$

We emphasize that the complexity of the original microscopic model has been significantly reduced by the coarse-graining (3.14). To determine the dimensionality of the mesoscopic state space, we note that for given $N_1 \in 0, \ldots, N$, the second variable N_2 can range from $0, \ldots, N - N_1$ so that

$$||\boldsymbol{N}|| = \left(\sum_{n=0}^{N}\sum_{m=0}^{n} 1\right) = \left(\sum_{n=0}^{N}(n+1)\right) = \frac{(N+1)(N+2)}{2} \overset{N\gg 1}{\sim} \frac{N^2}{2}, \tag{3.27}$$

the mesoscopic state space is asymptotically growing with the square of the system size.

3.2.1.4 Mean-Field Dynamics

In order to further reduce the complexity of the state space of the mesoscopic master equation (3.14) we first operate in the mean-field limit where $N \to \infty$. In this limit, the total change in internal energy due to a transition in Eq. (3.3) simplifies and the corresponding scaled density current $J(n_i, n_j) \equiv \lim_{N\to\infty} W_{\boldsymbol{N}\boldsymbol{N}'}/N$ becomes

$$J(n_i(t), n_j(t)) = \Gamma\, \mathrm{e}^{-\frac{\beta}{2}\left\{\epsilon_i - \epsilon_j + u[n_i(t) - n_j(t)] - f\,\Theta_{ij}\right\}}\, n_j(t)\,, \tag{3.28}$$

where $n_i(t) = N_i(t)/N$ denotes the occupation density of the single-unit state i and the sign function is defined as $\Theta_{ij} = 1$ if $(i-j) = 1 \bmod 3$, else $\Theta_{ij} = -1$. The evolution equation for the mean occupation density reads

$$\langle \dot{n}_i(t)\rangle = \sum_{j\neq i} \langle J(n_i(t), n_j(t))\rangle - \langle J(n_j(t), n_i(t))\rangle. \tag{3.29}$$

In the mean-field approximation we replace any n-point correlation function with a product of n averages thus yielding

$$\partial_t \overline{n}_i(t) \equiv \langle \partial_t n_i\rangle(t) = \sum_{j\neq i} J(\langle n_i\rangle(t), \langle n_j\rangle(t)) - J(\langle n_j\rangle(t), \langle n_i\rangle(t)), \tag{3.30}$$

which represents a closed nonlinear equation. Hence the mean-field system can be described by a single three-state unit, where the (average) occupation density of the single-unit states is assigned to the three states of the mean-field unit. We therefore identify the mean-field occupation density, $\overline{n}_i(t)$, as the probability for any unit to occupy the single-unit state $i = 1, 2, 3$. Its dynamics is ruled by the nonlinear mean-field equation

$$\partial_t \overline{n}_i(t) = \sum_j k_{ij}(t)\, \overline{n}_j(t)\,, \tag{3.31}$$

with the mean-field transition rates

$$k_{ij}(t) = \Gamma \mathrm{e}^{-\frac{\beta}{2}\left[\mathcal{E}_i(t) - \mathcal{E}_j(t) - f\,\Theta(i,j)\right]}, \quad \mathcal{E}_i(t) = \epsilon_i + u\,\overline{n}_i(t), \tag{3.32}$$

obeying local detailed balance

$$\ln \frac{k_{ij}(t)}{k_{ij}(t)} = -\beta\left[\mathcal{E}_i(t) - \mathcal{E}_j(t) - f\,\Theta(i,j)\right]. \tag{3.33}$$

Unit conservation erases one degree of freedom such that there are only two independent variables $\overline{n}_1(t)$ and $\overline{n}_2(t)$. We formally define the mean-field equilibrium free energy via the transcendental equation

$$\overline{n}_i^{eq} = \mathrm{e}^{-\frac{\beta}{2}(\mathcal{E}_i^{eq} - \mathcal{A}^{eq})}, \tag{3.34}$$

with the equilibrium mean-field free energy

$$\mathcal{A}^{eq} = -\frac{1}{\beta} \ln \sum_i \mathrm{e}^{-\frac{\beta}{2}\mathcal{E}_i^{eq}}, \tag{3.35}$$

since it satisfies the detailed balance relation

$$k_{ij}\,\overline{n}_j^{eq} = k_{ji}\,\overline{n}_i^{eq}. \tag{3.36}$$

We proceed by choosing a flat energy landscape of the three-state units, *i.e.* by setting $\epsilon_i = \text{const}\ \forall i$. This allows us to immediately read off the symmetric point $\overline{n}_i^{ss} = 1/3$ as an analytic solution to the nonlinear mean-field Eq. (3.31). Linearizing the Eq. around this fixed point,

$$\partial_t \overline{\boldsymbol{n}}(t) = \boldsymbol{J} \cdot \overline{\boldsymbol{n}}(t), \quad \boldsymbol{J}_{ij} = \left.\frac{\partial k_{ij}(t)}{\partial \overline{n}_j(t)}\right|_{\overline{n}_j(t)=\overline{n}_j^{ss}}, \tag{3.37}$$

we find for the eigenvalues of the linearized Jacobian $\boldsymbol{J}$,

$$\lambda_\pm = -\Gamma(\beta u + 3)\cosh\left(\frac{\beta f}{2}\right) \pm \mathrm{i}\sqrt{3}\,\Gamma \sinh\left(\frac{\beta f}{2}\right). \tag{3.38}$$

For attractive interactions ($u < 0$) between the units the real part of $\lambda_\pm$ changes its sign at $\beta_{c_1} = -3/u$. This crossover suggests that the stable symmetric fixed point destabilizes and degenerates into a limit cycle corresponding to a Hopf bifuraction indicative of synchronization [1]. The proof that the Hopf bifurcation occurs supercritical, *i.e.* gives rise to *stable* limit cycles, is deferred to appendix A.2.

Fig. 3.2 depicts the mean-field phase space for different β and f in units of u. The symmetric fixed point is only stable for $\beta < \beta_{c_1}$. We find in agreement with Eq. (3.38) that for finite f there is a phase characterized by stable limit cycles if $\beta \geq \beta_{c_1}$. For any value of f, we observe an additional phase with three non-symmetric fixed points for $\beta \geq \beta_{c_2}(f)$ that will be investigated further below.

We set $u = -1$ in the following and briefly address a subtlety of the mean-field system. In Fig. 3.2 the analytic solution to Eq. (3.31), $\overline{n}_i = 1/3$, is chosen as initial condition. In fact, at temperatures equal to the first critical temperature β_{c_1} the long-time solution is initial-condition dependent: Figure 3.3a) shows that for $0 < f = 0.15 < f_c \approx 0.21$ there is a finite set of initial conditions different from the symmetric fixed point that will *not* lead to a limit cycle but to a non-symmetric stable fixed point. Combining these two panels that exhibit four different phases each, we find that for the physical initial conditions

[1] We stress that, strictly speaking, synchronization refers to the convergence of an ensemble of phase oscillators with different eigenfrequencies towards a global frequency. Conversely, here, the units have the same eigenfrequency. We nonetheless will interpret the collective dynamics encoded by the limit cycles as synchronization.

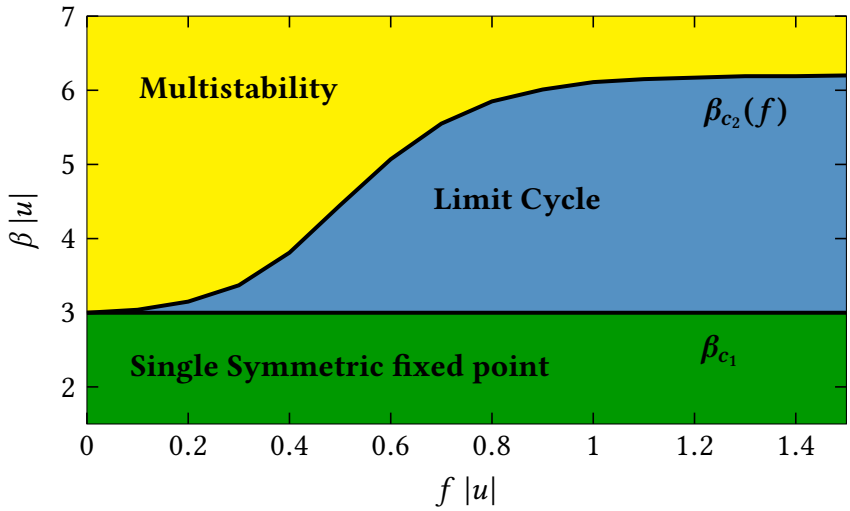

Figure 3.2: *Phase space in the mean-field varying the parameters β and f in units of u. The black lines correspond to the set of critical points β_{c_1} and $\beta_{c_1}(f)$.*

corresponding to the yellow and dark green phases the system will eventually arrive at one of the three non-symmetric fixed points[2]. These differ only by permutations of their components $(\overline{n}_1^{eq}, \overline{n}_2^{eq}, 1-\overline{n}_1^{eq}-\overline{n}_2^{eq})$. Consequently, the light-colored regimes correspond to the limit cycle, which in the vicinity of the Hopf bifurcation point β_{c_1}, has a small dimension, such that $\overline{n}_i(t) \approx 1/3$. If $f \geq f_c$, the dynamics will always exhibit a limit cycle regardless of the chosen initial condition. This can be seen in panel b) where $\overline{n}_i(t) \approx 1/3$ independent of the chosen initial condition for $f = 0.22 > f_c$.

Before studying the different nonequilibrium phases of this model, we discuss it for $f = 0$, *i.e.* at equilibrium. Figure 3.4a) shows, starting from the initial condition $\overline{\boldsymbol{n}}(0) = (1/3, 1/3)^\top$, the long-time solution $\overline{n}_1^{eq}(t)$ for different values of β. At the critical temperature β_{c_1} the system exhibits three non-symmetric stable fixed points that emerge via a saddle-node bifurcation. Starting from the symmetric fixed point, these attractive fixed points are observed to move towards the corners of the triangle in the $\overline{n}_1^{eq} - \overline{n}_2^{eq}$ plane. This is physically plausible since at low temperatures the system tends to occupy its lowest energy state where all units are occupying the same state.

The dependence of the multiple equilibrium states on the initial condition in the low-temperature phase is investigated in Fig. 3.4b). In the lower triangle, $\overline{n}_1^{eq}$ is plotted as a function of all physical initial conditions $(\overline{n}_1(0),\overline{n}_2(0))$. As a complement, the other mean-field probability $\overline{n}_2^{eq}$ is shown in the upper triangle, where the axis labels are omitted for better readability. Each triangle exhibits two phases which are separated by a contour line. With these two panels, we find that for every physical initial condition the system will eventually

[2]Note that a state does not correspond to a folding of the two triangles but a rotation of one of the two planes about 180° and subsequent overlapping of the two layers.

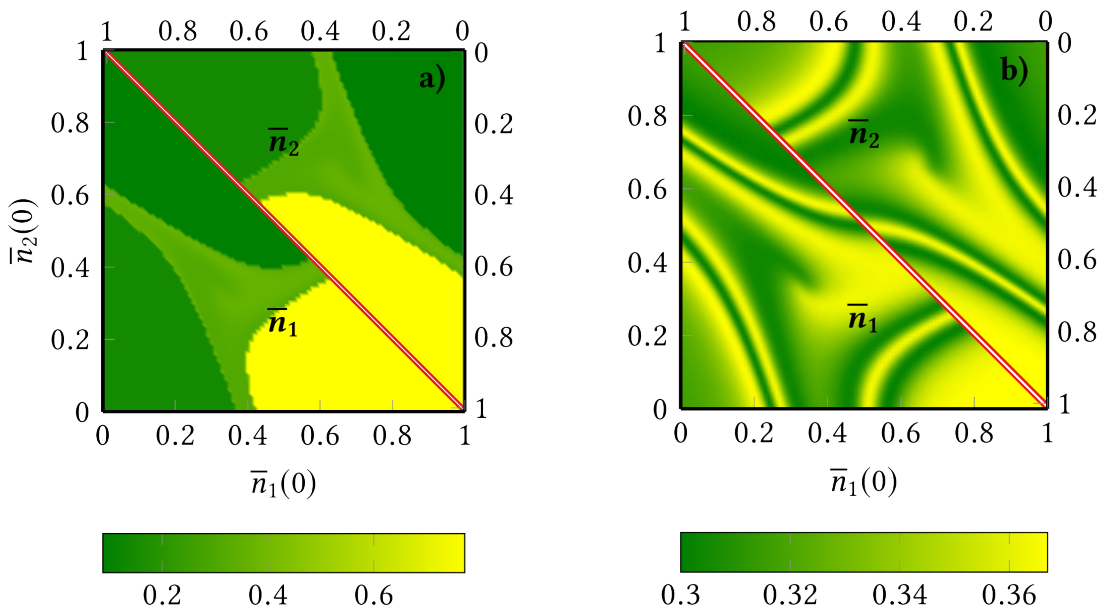

Figure 3.3: *Density plot of the long-time solution of the occupation density $\overline{n}_1(\overline{n}_2)$ for $\beta = \beta_{c_1}$ as a function of all physical initial conditions $\overline{n}_1(0)$ and $\overline{n}_2(0)$ in the lower left triangle (upper right triangle) for $f = 0.15$ in a) and $f = 0.22$ in b). The time is sufficiently long, $t = 10^3$, such that the system exhibits its asymptotic solution.*

arrive at one of the three non-symmetric fixed points. These differ only by permutations of their components $(\overline{n}_1^{eq}, \overline{n}_2^{eq}, 1 - \overline{n}_1^{eq} - \overline{n}_2^{eq})$, where two of them are identical according to the two phases in each of the panels in b). Moreover, our thermodynamic framework allows us also to work within the nomenclature of statistical mechanics. Interestingly, the saddle-node bifurcation corresponds to a first-order equilibrium phase transition since the derivative of the mean-field free energy with respect to β at the critical point β_c is divergent, as can be seen in diagram 3.4c). Emerging from the destabilizing symmetric fixed point, these attractive asymmetric fixed points are observed to move towards the corners of the triangle in the $\overline{n}_1^{eq} - \overline{n}_2^{eq}$ plane. This is physically plausible since at low temperatures the system tends to occupy its lowest energy state where all units are occupying the same state, while in the high-temperature limit any equilibrium system is known to be driven by entropy forcing the system into the high-entropy symmetric fixed point.

For the Hopf bifurcation to appear the system needs to be driven out-of equilibrium. Figure 3.5 depicts in a) the mean-field probability $\overline{n}_1(t)$ as a function of β at long times for $f = 1.0$.

With increasing f, the intermediate synchronization phase (S) is extending over a larger range of values for β (cf. Fig. 3.2). Since the phenomenology of the mean-field dynamics does not change under variation of β, we set $f = 1$ in the following. As depicted in a), the oscillations exhibit an increasing frequency with β up to a point where they slow down.

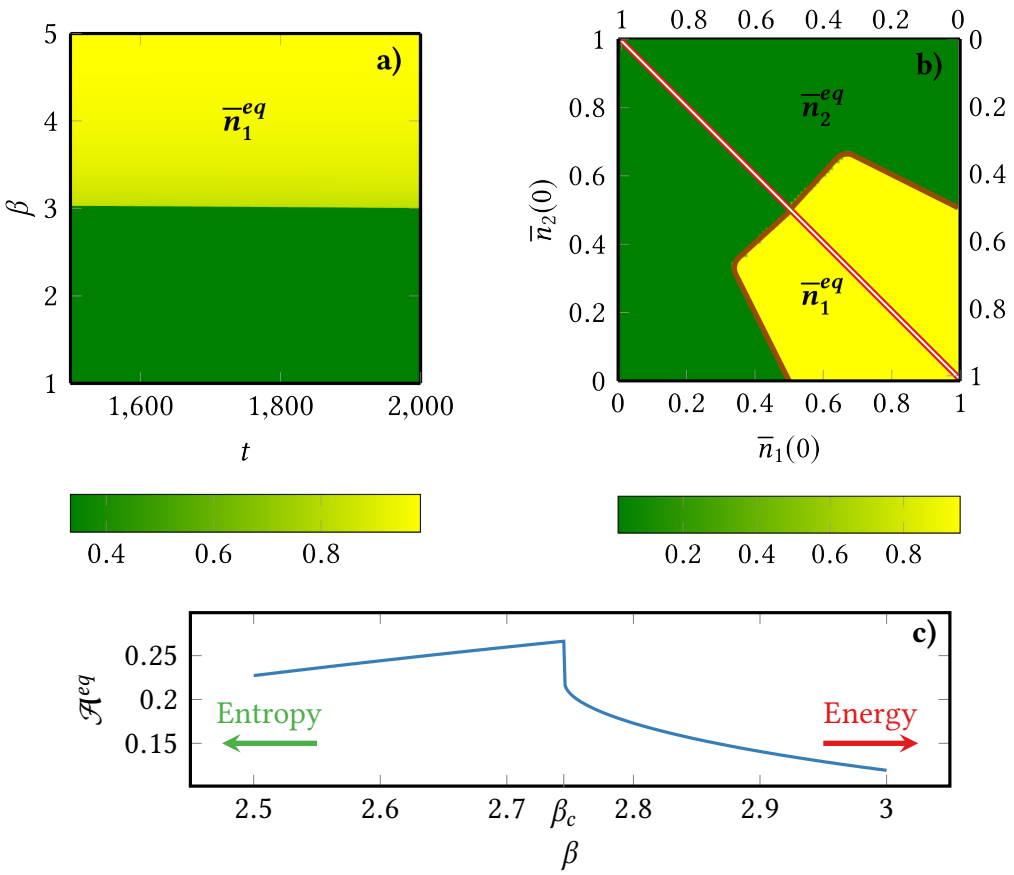

Figure 3.4: *Density plot of the equilibrium occupation density $\overline{n}_1^{eq}$ for different β and times t for an initial condition $\overline{\boldsymbol{n}}(0)$ equal to the symmetric fixed point in a), and as a function of all physical initial conditions $\overline{n}_1(0)$ and $\overline{n}_2(0)$ at time $t = 10^3$ and for $\beta = 4.0$ in b) (lower left triangle). For completeness, the upper right triangle in panel b) shows the other component $\overline{n}_2^{eq}$. The times are chosen to be sufficiently long such that the system has relaxed to equilibrium. The equilibrium free energy close to the critical point β_c is depicted in c).*

At the second critical point, $\beta_{c_2} \approx 6.11$, the oscillation period diverges corresponding to an infinite-period bifurcation [144]. The initial-condition dependence of the stationary states in the non-symmetric asynchronous phase (NA) is depicted in b), with $\beta = 7.0$. Again, depending on the chosen initial condition, the system will eventually arrive in one of the three asymmetric fixed points, which are again related to each other by permutations of their components. Here, in contrast to the equilibrium case, all components are different. This reflects the presence of the force distorting the symmetry of the states. The distortion occurs since it is more likely to jump from the largely populate state into the lower occupied state following the bias rather than the opposite way. This asymmetry naturally increases

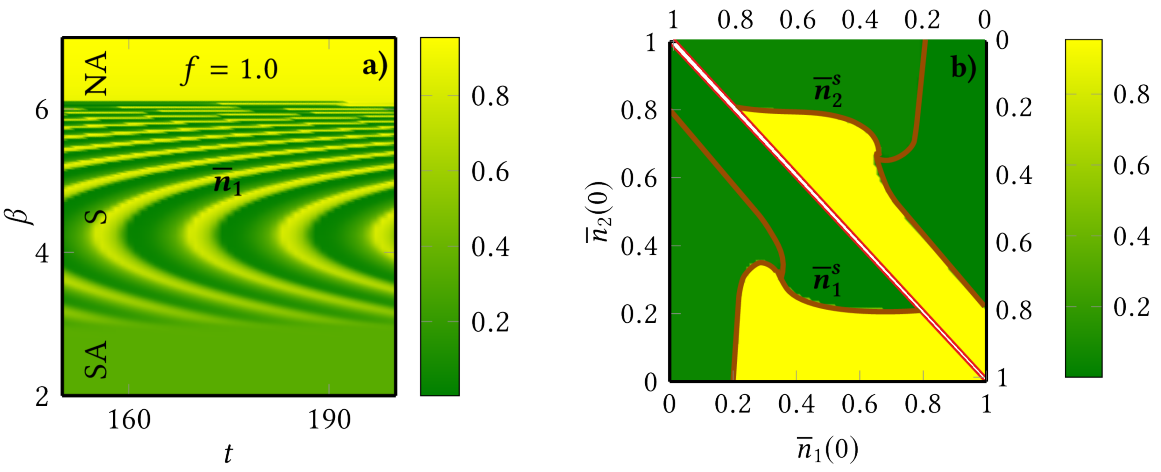

Figure 3.5: *Illustration of the occupation probability $\overline{n}_1$ as a function of β and t for $f = 1.0$ in a), while b) shows the occupation densities $\overline{n}_1$ in the lower left triangle and $\overline{n}_2$ in the upper right triangle as a function of all initial conditions at $\overline{n}_1(0)$ and $\overline{n}_2(0)$ at time $t = 10^3$ and for $\beta = 7.0$. The initial condition underlying the density plot in panels a) is $(\overline{n}_1(0) = 1, \overline{n}_2(0) = 0)$. The times are chosen such that the system has reached either the unique fixed point in the symmetric asynchronous phase (SA), the stable limit cycle in the synchronous phase (S), or one of the three asymmetric fixed points in the non-symmetric asynchronous phase (NA). The triple points defined by the intersecting contour lines in d) correspond to the symmetric unstable fixed point present in the S and NA phase.*

as the system is driven further out-of-equilibrium.

This constitutes the first important result. We have developed a minimal model which, according to Eqs. (3.38) and (A.39), exhibits synchronization and is thermodynamically consistent due to Eqs. (3.6), (3.22) and (3.33). We also note that synchronization only occurs in a finite range of temperatures: Fig. 3.4 illustrates that at low temperatures the equilibrated system is energy-driven and tends to its energetic ground state, while for very high temperatures the system is entropy-driven and takes a uniform stationary probability distribution. By extrapolation from equilibrium to the non-equilibrium scenario where the synchronization phase emerges, we realize that Fig. 3.5 invites for an analogous physical interpretation of the low- and high-temperature limit in the non-equilibrium case. Moreover, the limit $\beta \to 0$ represents the reversible limit since forward and backward transition for each pair of states become equally probable for any f and thus detailed balance holds. We remark furthermore that the term "minimal" refers to the dimensionality of the mean-field dynamics given by Eq. (3.31), which is a natural requirement to observe synchronization since a single-variable nonlinear differential equation can naturally not have complex eigenvalues.

3.2.1.5 Stochastic Dynamics

Spectral Analysis

Dominant Spectrum A crucial aspect of our model is that it allows us to study its (thermo-)dynamic features for large but finite system sizes and in particular to monitor the convergence of the stochastic dynamics to the mean-field dynamics. In order to proceed, we begin by stating the formal solution to the mesoscopic master equation (3.14) that reads

$$\boldsymbol{P}(t) = \mathrm{e}^{\boldsymbol{W}t} \cdot \boldsymbol{P}(0) = \sum_{i,i^*} \mathrm{e}^{\lambda_i t} \underbrace{\left(\Phi_i^L \cdot \boldsymbol{P}(0)\right)}_{\equiv c_i} \Phi_i^R + c_{i^*}\, \mathrm{e}^{\lambda_{i^*} t}\, \Phi_{i^*}^R, \tag{3.39}$$

where $\boldsymbol{P}(0)$ is the initial probability distribution, λ_i are the eigenvalues and Φ_i^L, Φ_i^R are the left- and right eigenvectors of the non-symmetric real transition rate matrix $\boldsymbol{W}$ constituting an orthonormal dual basis $\Phi_i^L \cdot \Phi_j^R = \delta_{ij}$. The index i^* characterizes, if existent, the modes with eigenvalues and eigenvectors being the complex-conjugated of those labeled with i.

The proof of the existence of an orthonormal dual basis is as follows. We consider the eigenvalue equations

$$\boldsymbol{W} \cdot \Phi_i^R = \lambda_i\, \Phi_i^R \tag{3.40}$$

$$\Phi_j^L \cdot \boldsymbol{W} = \lambda_j\, \Phi_j^L, \tag{3.41}$$

and by multiplying Eq. (3.40) with Φ_j^L from left and using Eq. (3.41), we arrive at

$$\Phi_j^L \cdot \boldsymbol{W} \cdot \Phi_i^R - \lambda_i\, \Phi_j^L \cdot \Phi_i^R = 0 \quad \Rightarrow\ (\lambda_j - \lambda_i)\, \Phi_j^L \cdot \Phi_i^R = 0. \tag{3.42}$$

Since $\lambda_i \neq \lambda_j$, it follows that Φ_j^L and Φ_i^R must be orthogonal, $\Phi_i^L \cdot \Phi_j^R = 0, i \neq j$. Normalization implies that the left and right eigenvectors form an orthonormal dual basis, $\Phi_j^L \cdot \Phi_i^R = \delta_{ij}$.

The Perron-Frobenius theorem [88] stipulates that for this irreducible, autonomous and stochastic matrix there is a non-degenerate eigenvalue, the Perron-Frobenius eigenvalue, $\lambda_0 = 0$, which is strictly greater than any other eigenvalue in both real part and absolute value, $|\lambda_i| < |\lambda_0|\ \forall i \neq 0$. Note that the labeling of the eigenvalues is given by the order of their real parts, $0 > \mathrm{Re}(\lambda_1) > \ldots > \mathrm{Re}(\lambda_{N-1})$. Consequently, Eq. (3.39) has a unique, infinite-time solution, $\boldsymbol{P}^s = c_0 \Phi_0^R$, characterized by the Perron-Frobenius eigenvalue and the associated right eigenvector Φ_0^R. Hence the stationary state of the mesoscopic system $\boldsymbol{P}^s$ cannot exhibit stable oscillations (S phase) or multistability (NA phase). On the other hand, one expects that the transition from the mesoscopic system to the mean field is smooth as the system size N grows. This apparent paradox is caused by the non-commutation of the infinite-time limit $t \to \infty$ and the mean-field limit $N \to \infty$, *i.e.*

$$\lim_{t\to\infty} \lim_{N\to\infty} \boldsymbol{P}(t) \neq \lim_{N\to\infty} \underbrace{\lim_{t\to\infty} \boldsymbol{P}(t)}_{\boldsymbol{P}^s}, \quad \text{if } \beta \geq \beta_{c_1}. \tag{3.43}$$

The right-hand side corresponds to the symmetric stationary state of the SA phase for all temperatures, while the left-hand side is temperature-dependent: For $\beta_{c_1} \leq \beta < \beta_{c_2}$ the system is in a time-periodic state (S phase) and for $\beta \geq \beta_{c_2}$ the dynamics will go to one of the non-symmetric steady states (NA phase) depending on the chosen initial condition. At $\beta < \beta_{c_1}$ the right-hand side also corresponds to the symmetric stationary state, hence the two limits commute only at sufficiently high temperatures.

To resolve this apparent contradiction we look for clues in the spectrum of the Markov generator in the mesoscopic master equation (3.14) and establish a link between finite-size systems and mean field via the notion of metastability. Even though the stationary state is inevitably reached in the infinite-time limit, there could be long-living metastable states that display the phenomenology of the mean field. The time-scales to characterize such a state are encoded in the spectrum as follows

$$\tau_r \sim -\frac{1}{\mathrm{Re}(\lambda_1)} \tag{3.44}$$

$$\tau_m \sim -\frac{1}{\mathrm{Re}(\lambda_2)} \tag{3.45}$$

$$\tau_l \equiv \tau_r - \tau_m \sim \frac{1}{|\,\mathrm{Re}(\lambda_1)|}\left(1 - \frac{\mathrm{Re}(\lambda_1)}{\mathrm{Re}(\lambda_2)}\right), \tag{3.46}$$

where τ_r is the relaxation time to reach the unique steady state, *i.e.* it specifies the time-scale at which all finite-time modes have been removed from the dynamics. τ_m is the metastable time at which all modes have decayed except for those forming the metastable state, that is the one associated with the eigenvalue λ_1 and the stationary one characterized by the Perron-Frobenius eigenvalue λ_0. Here, we assume that only one complex-conjugated non-null mode is contributing to metastability, while there could be an arbitrary number of modes forming the metastable state. This assumption will be numerically verified in the following. Physically, τ_l corresponds to the lifetime of that metastable state. To reconcile the stochastic dynamics with its asymptotic solution in the macroscopic limit, the mean-field dynamics, τ_l is required to become increasingly larger with the system size N, while τ_m remains finite since the different mean-field phases emerge at finite time. Using Eqs. (3.44)–(3.46), these prerequisites translate into conditions on the real parts of the dominant eigenvalues of the Markov generator: The real-part gap between the two first non-null mode eigenvalues, $\mathrm{Re}(\lambda_1) - \mathrm{Re}(\lambda_2)$, has to increase by $\mathrm{Re}(\lambda_1)$ converging to zero (corresponding to a diverging relaxation time τ_r), while $\mathrm{Re}(\lambda_2)$ has to approach a finite value (assuring the emergence of the metastable phenomena at finite times). Moreover to mimic mean-field dynamics the metastable state has to be oscillatory ($\mathrm{Im}(\lambda_1) \neq 0$) in the S phase and quasistationary ($\mathrm{Im}(\lambda_1) = 0$) in the NA phase.

Before addressing the question of how the stochastic dynamics converges to the mean field, we study the real parts a) and the imaginary parts b) of the two dominant non-zero eigenvalues of the spectrum in all three different phases ($2 \leq \beta \leq 8$) for a system size of $N = 300$ in Fig. 3.6. We remark that for all β, these two eigenvalues in fact occur as complex-conjugated pairs and only those with a positive imaginary part are depicted in panel b).

Furthermore, to stress that the different phases of the finite-size system for $\beta > \beta_{c_1}$ are only present for finite times, we rename them differently than in the mean field: asynchronous phase (A), synchronous metastable phase (SM) and asynchronous metastable phase (AM).

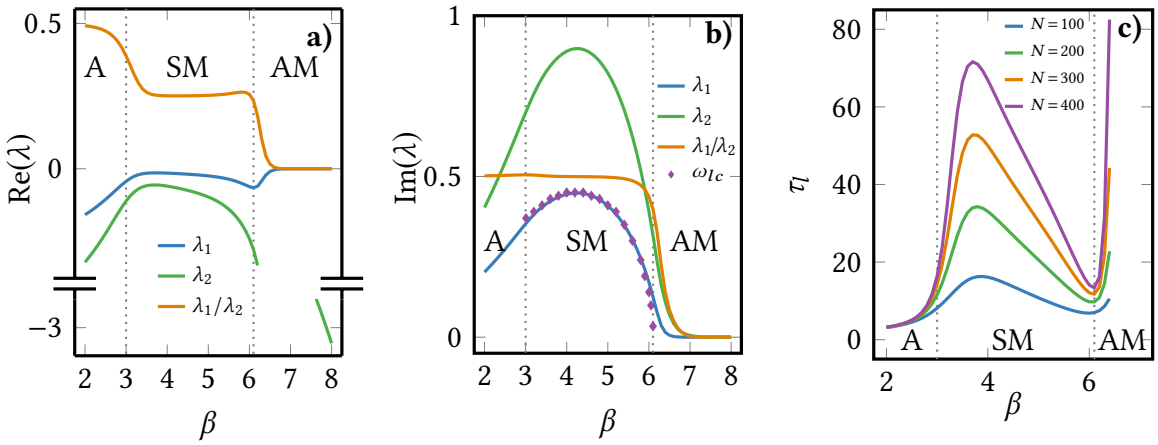

Figure 3.6: *The real part a) and the imaginary part b), as well as the ratio, of the two most dominant eigenvalues with distinct real part, λ_1,λ_2, with positive imaginary part are depicted as a function of β. In addition, the limit cycle frequency ω_{lc} that is numerically extracted from the asymptotic ($t \to \infty$) mean-field dynamics is compared to the imaginary part of the most dominant eigenvalue. All eigenvalues correspond to a generator $\boldsymbol{W}$ for a system of size $N = 300$. Panel c) shows the lifetime of the metastable state τ_l as function of β for different system sizes. The labels of the different phases, that is the asynchronous phase (A), the synchronous metastable phase (SM) and the asynchronous metastable phase (AM) are in correspondence with the labels of the different phases in the mean-field limit introduced in the preceding Sec. 3.2.1.4.*

As can be seen in panel a), the real parts of the two eigenvalues both approach zero up to $\beta \approx 4$ followed by a monotonous decrease of $\mathrm{Re}(\lambda_2)$ while $\mathrm{Re}(\lambda_1)$ changes only slightly and for $\beta > \beta_{c_2}$ rapidly goes to zero. According to Eq. (3.46), this observation along with the fact that $\mathrm{Re}(\lambda_1)/\mathrm{Re}(\lambda_2)$ drops at both critical points (dashed lines) suggests that the lifetime τ_l of the metastable state is increasing from the SM to the AM regime. The small values of $|\mathrm{Re}(\lambda_1)|$ in the SM and AM phase and the sharp changes in the ratio of the real parts at both critical points provides a first hint that the metastable state is constituted by only the stationary mode and the first complex-conjugated non-null modes.

This claim is further strengthened by studying the corresponding imaginary parts of these eigenvalues as shown in Fig. 3.6b). We find an excellent agreement in the SM phase between the limit cycle frequency ω_{lc} in the mean field that is numerically extracted from the dynamics and $\mathrm{Im}(\lambda_1)$. The limit cycle frequency ω_{lc} only coincides with the imaginary part of the Jacobian from the linear stability analysis in Eq. (3.38) at the bifurcation point

β_{c_1}, where the linearization of the nonlinear master equation (3.31) is exact. Moreover, the ratio between the imaginary parts of λ_1 and λ_2 remains nearly constant at 0.5 within the A and SM phase implying that the frequency of oscillation of the mode corresponding to λ_2 is half as that of λ_1. In the AM phase $\mathrm{Im}(\lambda_1)$ quickly goes to zero consistent with our mean-field observations that show no oscillations. Consistent with the discussion of the real parts, Fig. 3.6c) illustrates that the lifetime of the metastable state is nearly zero in the A phase and starts to increase significantly at the first critical point up to a local maximum in the SM phase. The lifetime τ_l is monotonously decreasing for larger β before it sharply rises in the AM phase. All clues thus indicate that in the two phases where the mean field exhibits non-unique solutions at infinite times, the finite system displays metastability. As expected, for all temperatures in the metastable phases the lifetime is monotonously increasing with N.

Next, to shed some light on the convergence of the finite-system dynamics to the mean-field dynamics, we investigate the changes in the spectrum as we approach the mean-field limit. To this end, we look at the first few dominant non-zero eigenvalues as a function of the system size N at $\beta = 4$ representing the SM phase. We observe in Fig.3.7a) that the real parts of these eigenvalues are approaching the Perron-Frobenius eigenvalue. Though the inset reveals an increasing time-scale separation between the mode associated with λ_1 and the faster decaying modes for larger systems. The monotonously increasing behavior of τ_l and τ_r with N implies an increasing lifetime of the metastable state, while this time window is shifted to increasingly larger times, hence the finite-system dynamics are converging to the mean field. To be fully consistent with the mean field, the metastable state must be appearing in the dynamics at a finite time. Taking into account all the aforementioned hints (encoded in Fig. 3.6 and to be made in the following) that indeed only the modes associated with $\lambda_{1,1^*}$ are contributing to the metastability and therefore in correspondence with the mean-field solution, it is reasonable to expect that $\mathrm{Re}(\lambda_2)$ converges to a finite value for larger N. Unfortunately, extracting the dominant eigenvalues of the generator for even larger N is not feasible.

As another striking evidence for the hypothesis that the metastable state comprises only the stationary and the first non-null complex-conjugated mode, the imaginary part of the dominant eigenvalue λ_1 smoothly converges to the limit cycle frequency ω_{lc} in the mean field while the imaginary parts of other modes display a distinct separation as seen in Fig.3.7b). This is confirmed in Figs. 3.7c) – f) depicting the mean occupation densities, $\langle\rho(t)\rangle = \sum_{\boldsymbol{N}} \boldsymbol{N}(t)/N\, P_{\boldsymbol{N}}(t)$, using the full spectral decomposition of the Markov generator in Eq. (3.39) and the truncated one

$$\boldsymbol{P}(t) \overset{\tau_m \ll t \ll \tau_r}{\approx} c_0 \Phi_0^R + \mathrm{e}^{\lambda_1 t} c_1 \Phi_1^R + \mathrm{e}^{\lambda_1^* t} c_1^* \Phi_1^{R^*}, \tag{3.47}$$

for $N = 10^2, 10^3$ at $\beta = 4$.

To understand the metastability in the AM phase, Fig. 3.8 depicts the real and imaginary parts of the eigenvalues associated with the most dominant modes in panels a) and b), respectively, as a function of N for $\beta = 7$.

In contrast to Fig. 3.7a), here, $\mathrm{Re}(\lambda_2)$ clearly converges to a finite value with $\mathrm{Re}(\lambda_1)$ quickly

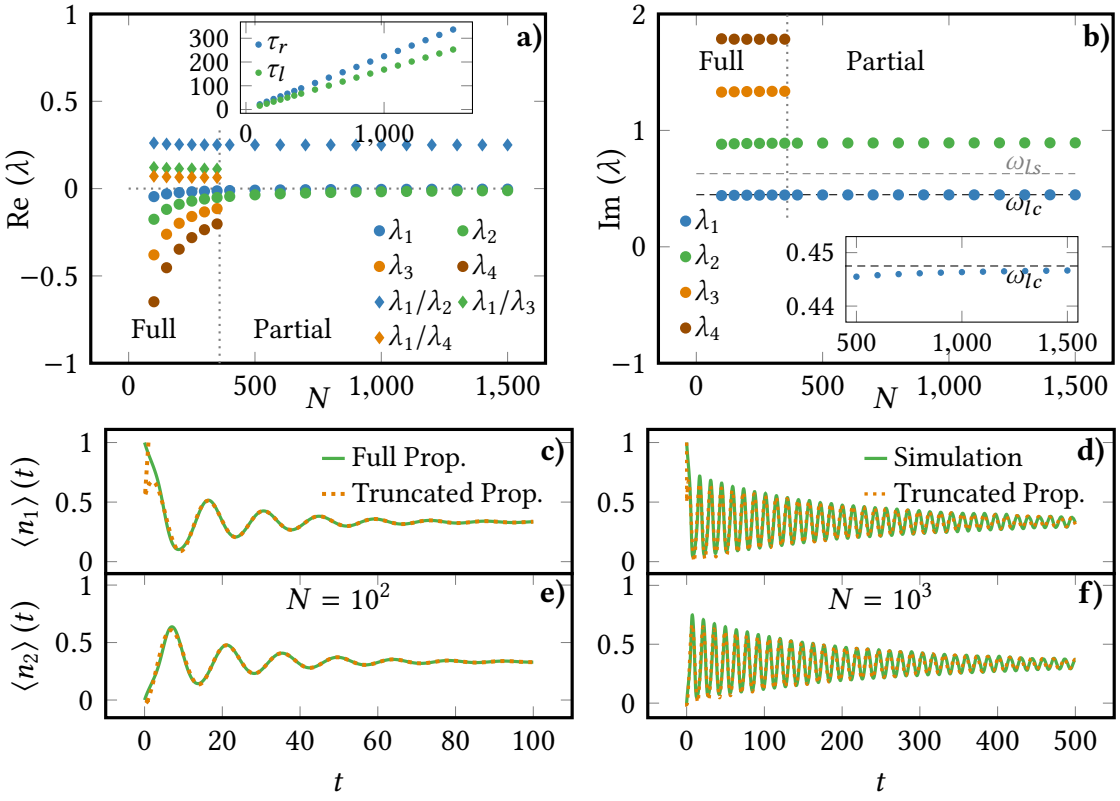

Figure 3.7: *Real a) and corresponding imaginary b) parts of the four most dominant eigenvalues with distinct and finite real part for different N and for $\beta = 4$ as a representative of the SM phase. The data points corresponding to system sizes larger than $N = 350$ are not resulting from a full diagonalization of the matrix but were obtained exploiting the sparseness of the matrices (maximal 6 of the approximately $N^2/2$ entries of every row/column are nonzero), using the iterative Arnoldi algorithm [145], to obtain the dominant part of the spectrum. The inset depicts the relaxation time scale τ_r and the lifetime of the metastable state τ_l as a function of N. In b) the dashed, horizontal lines labeled as ω_{lc} and ω_{ls} correspond to the limit cycle frequency in the mean field and to the imaginary part of the linear stability matrix eigenvalue from Eq. (3.38), respectively. The mean occupation density $\langle n_i \rangle$ as a function of time for both the full (3.39) and truncated (3.47) propagation [$i = 1$ in c), d) and $i = 2$ in e) and f)] for the different network sizes $N = 10^2, 10^3$. The dynamics for $N = 10^3$ was generated using the direct Gillespie method (cf. appendix A.3).*

going to zero already for small N. This is confirmed by the inset showing that τ_l and τ_r take very large values already for smaller systems implying that the metastability in the AM phase is much stronger than in the SM phase. As expected, in compliance with the nonoscillatory

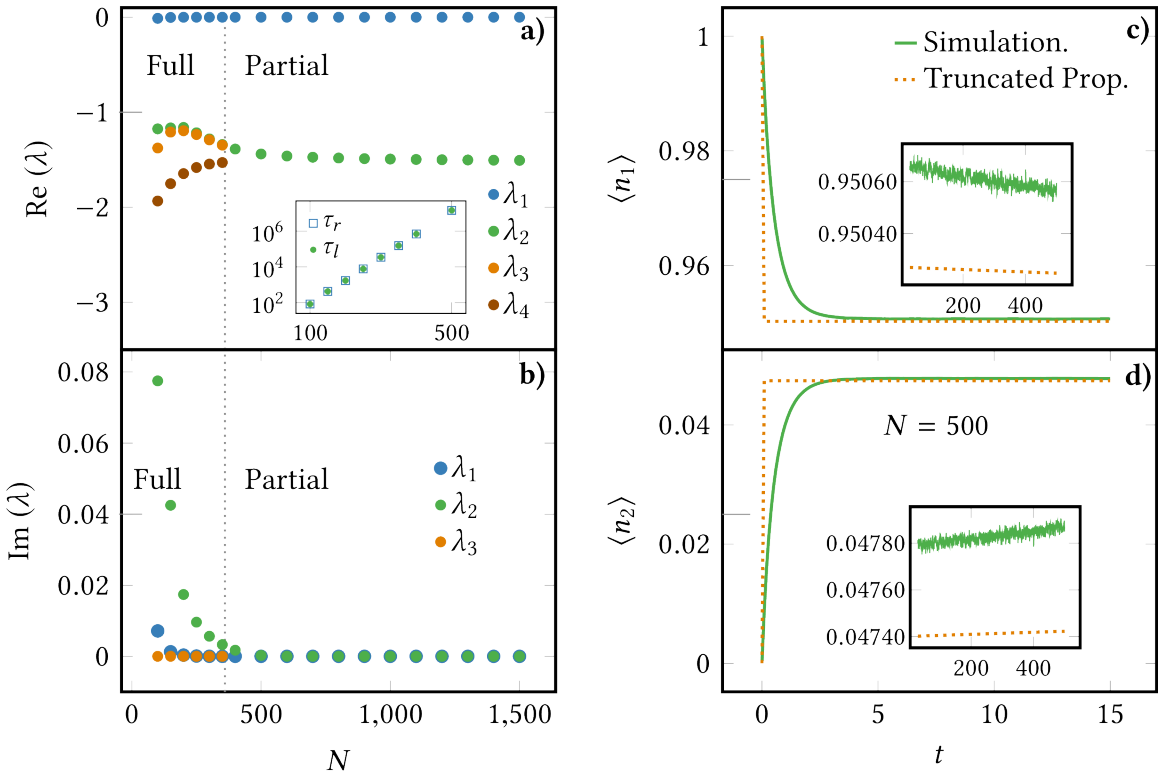

Figure 3.8: *Real a) and corresponding imaginary b) parts of the four most dominant eigenvalues with distinct and finite real part for different N and for $\beta = 7$ as a representative of the AM phase. The inset in a) depicts the relaxation time scale τ_r and the lifetime of the metastable state τ_l as a function of N. Furthermore, the mean occupation density $\langle n_i \rangle$ as a function of time for both the full (3.39) and truncated (3.47) propagation [$i = 1$ in c) and $i = 2$ in d)] for $N = 10^2$ is depicted.*

mean-field solution, the small magnitudes of the imaginary part vanish rapidly with growing system size as displayed in panel 3.8b). Figs. 3.8c) - d) reaffirm that the metastable state in the AM phase is reached at short time-scales and is quasistationary. Moreover, we note the large time-scales (cf. the scale of the axis of the insets) over which the metastable state can be observed in the dynamics in compliance with the observations made in panel 3.8a). Thus, we conclude from the observations made in this section, that for sufficiently large systems in the SM and AM phase at times $\tau_m \ll t \ll \tau_r$, the relaxation dynamics is determined by the metastable state associated with $\lambda_{1,1^*}$ and the Perron-Frobenius eigenvalue. This time span is increasing with N [cf. Figs 3.7a) and 3.8a)] such that the metastable states can be observed over increasingly larger times. Owing to the Perron-Frobenius theorem, any finite system will eventually leave these metastable states at times $t \gg \tau_r$ and relax into the unique stationary state at infinite time. To sum up, we obtain the important result that the different phases and bifurcations of the mean-field dynamics are encoded in the spectrum of

the Markov generator.

Large Spectrum Figure 3.9 shows a 3D-plot of all eigenvalues with $\mathrm{Re}(\lambda) \geq -5$ for a system with $N = 300$ at selected values of the bath parameter β.

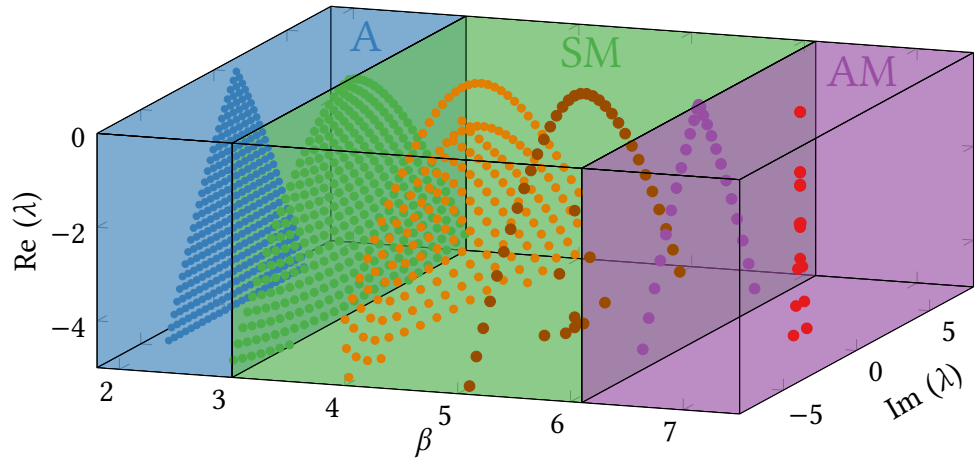

Figure 3.9: *3D-plot of the real all eigenvalues of the Markov generator* $\mathbf{W}$ *from Eq.* (3.14) *for* $N = 300$ *with* $\mathrm{Re}(\lambda) \geq -5$ *as a function of* β *and the corresponding imaginary parts* $\mathrm{Im}(\lambda)$ *in a). The labels of the different regimes, that is A (asynchronous), SM (synchronous metastable) and AM (asynchronous metastable) are in correspondence with the labels of the different regimes in the mean-field limit introduced in the Sec. 3.2.1.4.*

To illustrate the different regimes of the model in the mean-field limit, the 3D-plot is divided in three boxes, where the boundaries are defined by the critical values $\beta_{c_{1,2}}$. First of all, we note that for increasing values of β, the overall shape of the dominant part of the spectrum changes from a triangle to an arc. With increasing β, the latter takes again a triangular shape shrinking to a single line corresponding to vanishing imaginary parts at large β. More strikingly, we observe that at the first bifurcation point, $\beta_{c_1} = 3.0$, a real-part gap separating the bulk and the enveloping band eigenvalues emerges in the vicinity of the Perron-Frobenius eigenvalue λ_0 and widens up with increasing β. This increasing real-part gap suggests that for $\beta \geq \beta_{c_1}$ there is a significant separation between the time scales associated with the relaxation of the slow and the fast decaying modes associated with the eigenvalues in the bulk and the band, respectively. We emphasize that this time-scale separation between the slow (band) and the fast (bulk) decaying modes must not be confused with the time-scale separation between the first non-null mode and the Perron-Frobenius eigenvalue that are both located on the band. It is furthermore interesting to observe that at the second bifurcation point, $\beta_{c_2} \approx 6.1$, the bulk is no longer visible on this scale and the band has transformed into an arc that shrinks as β increases.

We proceed with studying an even larger part of the spectrum in all three phases for different system sizes N in Fig. 3.10.

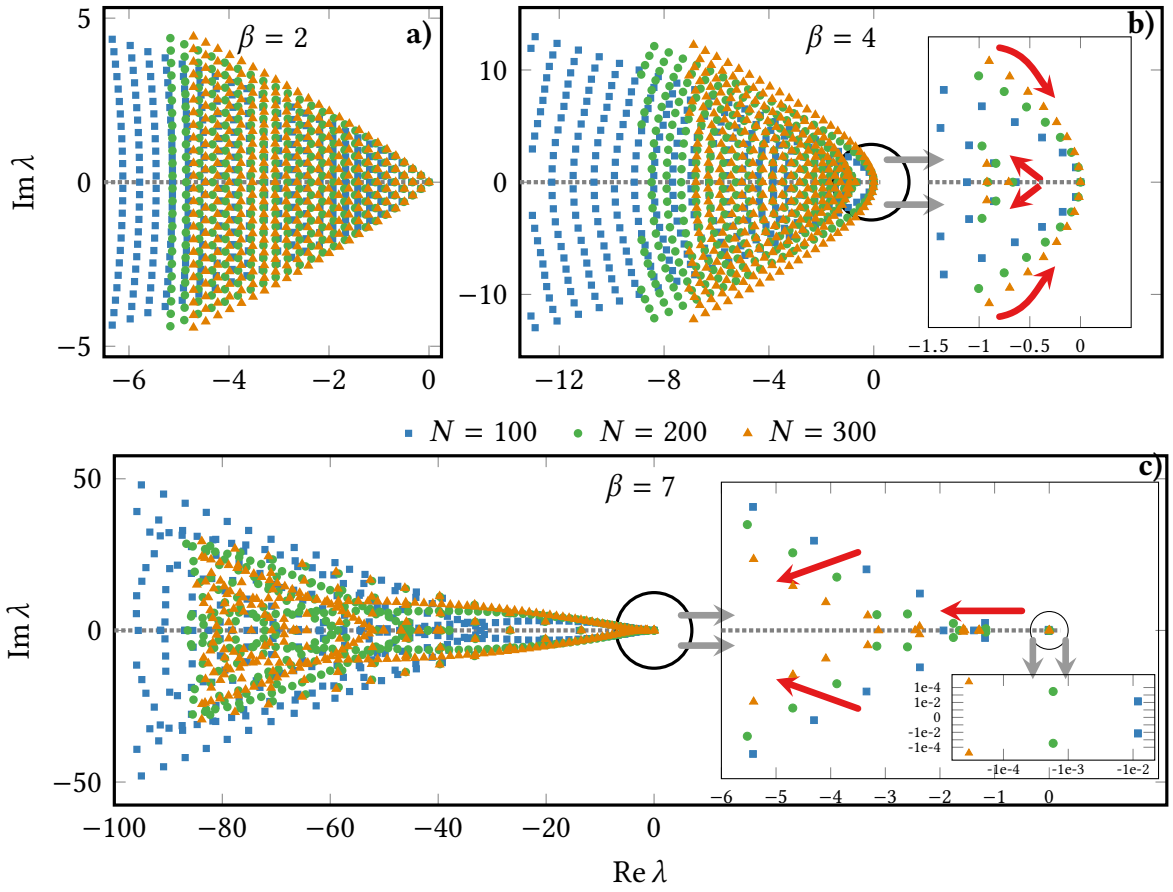

Figure 3.10: *Compilation of 300 eigenvalues of the Markov generator* $\boldsymbol{W}$ *with the largest real parts for different system sizes* $N = 100, 200, 300$ *and for* $\beta = 2$ *in a),* $\beta = 4$ *in b) and* $\beta = 7$ *in c).*

First, the spectrum for β in panel a) does not show any qualitative differences with respect to the corresponding one in Fig. 3.9. Next, for the case $\beta = 4$ depicted in panel b), we note clustering of the eigenvalues as they are moving towards larger real parts with growing N. However, on a smaller scale, the eigenvalues of the surrounding band are approaching the Perron-Frobenius eigenvalue, while the eigenvalues of the bulk are moving away from the Perron-Frobenius eigenvalue. This leads to a separation between the bulk eigenvalues and the enveloping band of eigenvalues. The inset magnifies the vicinity of the gap for better visibility. Here, the most dominant eigenvalues of the bulk are tending to the left with growing system size, while, as already discussed, the overall trend of the bulk eigenvalues seems to be towards the right. This implies a region of large eigenvalue density and strong clustering in the bulk and that the modes associated with the band eigenvalues decay more

slowly and are longer present in the relaxation dynamics for larger systems. Fig. 3.10c) depicts the spectrum for $\beta = 7$, which is remarkably reminiscent of a rocket. Again, the eigenvalues shift towards larger real parts for larger N. Contrary to the observations made in b), the magnified area shown in the inset reveals that, here, all eigenvalues are moving away from the first eigenvalue pair, $\lambda_{1,2}$, and the Perron-Frobenius eigenvalue which coincide on this scale. Moreover the magnification of this region illustrates that for rather moderate changes in N the real part of the eigenvalue pair, $\lambda_{1,2}$, shrinks by orders of magnitudes. Hence the metastability in the AM phase is much stronger than in the SM phase, an observation that, in less quantitative terms, was already made further above.

Dynamic Monte-Carlo Simulations

Solving the master equation (3.14) for systems on the order of $N \sim 10^3$ via full diagonalization of the propagator is computationally not feasible[3]. Hence for extremely large systems we resort to a stochastic simulation algorithm for computing the time evolution of the (Markov) jump processes. This dynamic Monte Carlo method, elaborated in appendix A.3, is often referred to as Gillespie algorithm [146, 147]. This algorithm generates trajectories of a stochastic process that are exact solutions to the stochastic process. By generating sufficiently many trajectories one can infer the statistics of the observables of the stochastic process, in particular the average values generically denoted by $\langle\cdot\rangle$.

Figure 3.11 depicts the $\langle n_2\rangle - \langle n_1\rangle$ plots generated with the Gillespie algorithm sampling over 10^6 trajectories for selected values of β and for different system sizes $N = 10^2, 10^4$. Except for $\beta = 6.1$ shown in e), the larger system, $N = 10^4$, agrees well with the mean-field limit at the displayed times. The smaller system, $N = 10^2$, significantly deviates in both the SM phase ($\beta = 4, 5, 6.1$) and AM phase ($\beta = 7$). In the A phase ($\beta = 2, 3$), there are no visible differences between the different finite system sizes and the mean-field limit solution, as all are relaxing into the unique symmetric fixed point [red closed circle in panel a)]. Of particular interest is the dynamics for $\beta = 7$. While the smaller system directly goes to the stationary state, the larger system quickly approaches and wiggles around the fixed point of the mean-field limit. This can be seen from the inset that displays a magnification around one of the mean-field fixed point [orange closed circle in f)]. Depending on the initial condition the metastable state will approach one of the three mean-field fixed points. This shows that the stochastic dynamics of sufficiently large systems indeed reproduces the mean-field dynamics and thus confirms all predictions made above based on the spectral analysis. As an exception, we observe in e) that close to the infinite-period bifurcation, $\beta \approx \beta_{c_2}$, the large system does not exhibit the characteristics of the solution in the mean-field limit. However, an even larger system, $N = 10^6$, shows signatures of the limit cycle albeit

[3] We mention that the non-symmetric real matrix implies in general a complex eigensystem which shall be determined with float precision. The amount of random-access memory (RAM) required to diagonalize a matrix of dimension 80601×80601 corresponding to a system consisting of 400 units is about 312 GB. We restrict the diagonalization to that size and employ different numerical methods for larger systems.

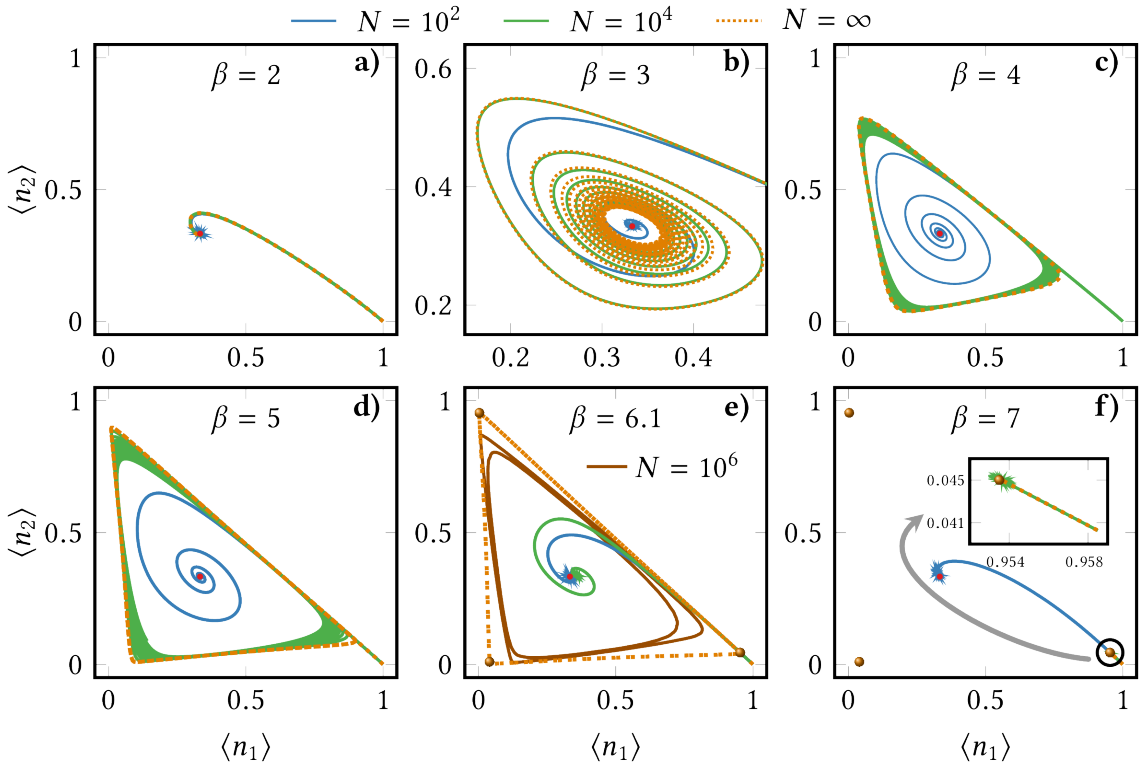

Figure 3.11: *Parametric plot of the mean occupation densities $\langle n_i \rangle$ and comparison between the latter for different finite system sizes $N = 10^2, 10^4$ and the mean-field limit at distinct values of β. The red closed circles correspond to the unique stationary state, $\langle n_i \rangle = 1/3$, for finite N. In all panels we initialize the system in the ground state with $n_1 = 1$ and sample 10^6 trajectories using the direct method of the Gillespie algorithm. The green and blue spikes are due to the steady state noise inherent to this stochastic algorithm. The simulations were carried out using the HPC facilities of the University of Luxembourg [148].*

still deviating. These deviations are due to the strong fluctuations in the vicinity of the phase transition calling for larger N such that the finite system can accurately represent the deterministic limit. We remark that this feature is also manifested in the increasing deviations between the limit cycle frequency, ω_{lc}, and the imaginary part of the crucial eigenvalue, λ_1, as the second critical point, $\beta_{c_2} \approx 6.1$ is approached [cf. Fig. 3.6c)].

However, there is a set of initial conditions for which the stochastic dynamics will not go to one of these metastable states. This set of initial conditions is readily constructed via all possible linear combinations of right eigenvectors of the mesoscopic generator from Eq. (3.14), $\boldsymbol{P}(0) = \sum_{i \neq 1} a_i \, \Phi_i^R$, excluding the mode associated with the crucial eigenvalue pair $\lambda_{1,1^*}$. It follows from the orthonormal dual-basis property of the eigensystem that the weights $c_{1,1^*} = 0$ in Eq. (3.39). Hence the metastability would be removed from the dynamics.

This prompts the question whether the metastability is a generic (up to a negligibly small set of special initial conditions) property of the stochastic process or just an artifact of choosing suitable initial conditions.

This question is addressed in Fig. 3.12, where the initial conditions are sampled and the joint probability distribution $P\{n_1(t = 20), n_2(t = 20)\}$ for different system sizes $N = 10^2, 10^4$ and different $\beta = 4, 7$ is shown in a density plot. In panel 3.12a) the distribution exhibits its maxima indicated by the red spots close to the corners of the limit cycle in the mean-field limit. Overall, the distribution clearly exhibits signatures of the limit cycle but the probability mass is still dispersed aorund the limit cycle contour. Moreover, over the entire state space there are regions with finite probability. If the system size is notably increased to $N = 10^4$, as depicted in Fig. 3.12b), the probability mass is sharply concentrated on the limit cycle contour. Turning to panels c) and d) corresponding to the AM regime with $\beta = 7$, we observe that the joint probability distribution for the smaller system already reproduces to a good approximation the three asymmetric fixed points in the mean-field limit. The distribution for the larger system further concentrates the probability mass on the three fixed points as can be seen by comparing the insets on the left and on the right magnifying the vicinity of the fixed points. The convergence of the probability distribution at smaller N to the mean-field limit for larger β is consistent with the observations already made in the spectral analysis in Fig. 3.6. We thus confirm, once again, that the metastability and therefore the convergence to the mean-field limit increases with N and β. Next, and more importantly, the emergence of the metastable state(s) is, up to a negligible set of special initial conditions, indeed a generic property of the stochastic process. It is insightful to monitor the time evolution of $P\{n_1(t), n_2(t)\}$ starting from a uniform grid at $t = 0$ up to a time as the distribution becomes stationary or time-periodic. To this end, Ref. [149] includes movies displaying the dynamics of the distributions shown in Fig. 3.12.

We have so far established a connection between linear stochastic dynamics and the deterministic nonlinear mean-field dynamics via the study of the spectrum of the Markov generator. Indeed, the different dynamical phases and bifurcations in the mean field are encoded in the spectrum and appear as metastable states for long times in the stochastic dynamics. These predictions are confirmed by our simulations. We now proceed by analyzing the bifurcations as nonequilibrium phase transitions in the thermodynamic observables. In doing so, we link deterministic bifurcation theory to stochastic thermodynamics.

3.2.1.6 Stochastic Thermodynamics

Thermodynamic Laws

We first introduce the basic thermodynamic state functions in this model: the microscopic internal energy and the system entropy

$$\langle e \rangle (t) = \sum_{\alpha} e_{\alpha} \, p_{\alpha}(t) \tag{3.48}$$

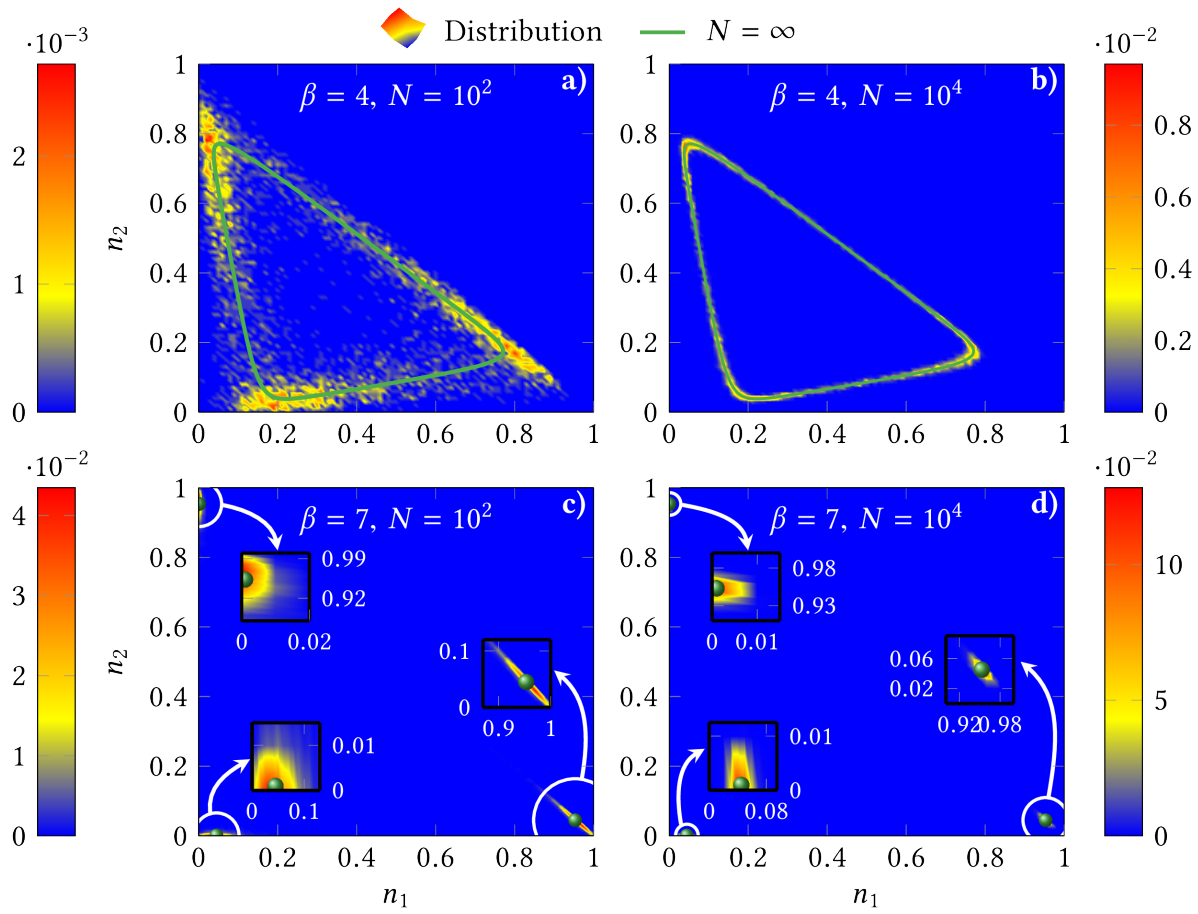

Figure 3.12: *Joint probability distribution $P\{n_1(t), n_2(t)\}$ at $\beta = 4$ in a), b) and at $\beta = 7$ in c), d) for system sizes $N = 10^2, 10^3$ at time $t = 20$. The plots were created using a grid of dimension 101×101 that specifies the set of initial conditions. For comparison, in all plots the long-time mean-field solution is overlaid.*

$$\langle s \rangle (t) = -\sum_{\alpha} p_\alpha(t) \ln p_\alpha(t) \,. \tag{3.49}$$

For our setup with an autonomous driving, f, these functions can only change due to the time-dependence of the probability distribution. The rate of change of internal energy

$$\mathrm{d}_t \langle e \rangle (t) = \sum_{\alpha,\alpha'} e_\alpha \, w_{\alpha\alpha'} \, p_{\alpha'}(t) = \langle \dot{q} \rangle (t) + \langle \dot{w} \rangle (t), \tag{3.50}$$

naturally defines the microscopic first law of thermodynamics with the heat and work current

$$\langle \dot{q} \rangle (t) = \sum_{\alpha,\alpha'} \underbrace{[e_\alpha - e_{\alpha'} - f\,\Theta(\alpha,\alpha')]}_{q_{\alpha,\alpha'}} w_{\alpha\alpha'} \, p_{\alpha'}(t) \tag{3.51}$$

$$\langle \dot{w} \rangle(t) = f \sum_{\alpha,\alpha'} \Theta(\alpha, \alpha')\, w_{\alpha\alpha'}\, p_{\alpha'}(t)\,. \tag{3.52}$$

The microscopic local detailed balance relation (3.6) can be expressed in terms of the heat exchange between the thermal reservoir and the system along the forward transition as follows

$$q_{\alpha,\alpha'} = -\frac{1}{\beta} \ln \frac{w_{\alpha\alpha'}}{w_{\alpha'\alpha}}. \tag{3.53}$$

For the microscopic entropy balance we get

$$\mathrm{d}_t \langle s \rangle(t) = \langle \dot{s}_e \rangle(t) + \langle \dot{\sigma} \rangle(t) \tag{3.54}$$

with the microscopic entropy flow

$$\langle \dot{s}_e \rangle(t) = -\sum_{\alpha,\alpha'} w_{\alpha\alpha'}\, p_{\alpha'}(t) \ln \frac{w_{\alpha\alpha'}}{w_{\alpha'\alpha}} = \beta \langle \dot{q} \rangle(t), \tag{3.55}$$

and the microscopic entropy production rate

$$\langle \dot{\sigma} \rangle(t) = \sum_{\alpha,\alpha'} w_{\alpha\alpha'}\, p_{\alpha'}(t)\, \ln \frac{w_{\alpha\alpha'}\, p_{\alpha'}(t)}{w_{\alpha'\alpha}\, p_{\alpha}(t)} \geq 0, \tag{3.56}$$

whose non-negativity constitutes the microscopic second law of thermodynamics.

The exact marginalization of the microscopic dynamics (3.14) does not *a priori* guarantee that the microscopic thermodynamic observables defined above are invariant under this coarse-graining [112]. Recalling that we have $e_{\alpha_{\boldsymbol{N}}} = E_{\boldsymbol{N}}$, the internal energy (3.48) can be corse-grained as follows

$$\mathrm{d}_t \langle e \rangle(t) = \sum_{\boldsymbol{N},\boldsymbol{N}'} E_{\boldsymbol{N}} \sum_{\alpha_{\boldsymbol{N}}} \sum_{\alpha'_{\boldsymbol{N}'}} w_{\alpha_{\boldsymbol{N}}\alpha'_{\boldsymbol{N}'}}\, p_{\alpha'_{\boldsymbol{N}'}}(t) = \sum_{\boldsymbol{N},\boldsymbol{N}'} E_{\boldsymbol{N}}\, W_{\boldsymbol{N}\boldsymbol{N}'}\, P_{\boldsymbol{N}'}(t) \equiv \mathrm{d}_t \langle E \rangle. \tag{3.57}$$

Thus, the average internal energy is invariant under the coarse-graining and can be equivalently represented in the mesospace. Furthermore, the microscopic heat (3.51) and work current (3.52) can also be exactly coarse-grained to obtain

$$\langle \dot{Q} \rangle = \sum_{\boldsymbol{N},\boldsymbol{N}'} \underbrace{\left(E_{\boldsymbol{N}} - E_{\boldsymbol{N}'} - f\, \Theta(\boldsymbol{N}, \boldsymbol{N}')\right)}_{Q(\boldsymbol{N},\boldsymbol{N}')} W_{\boldsymbol{N}\boldsymbol{N}'} P_{\boldsymbol{N}'}(t) \tag{3.58}$$

$$\langle \dot{W} \rangle = f \sum_{\boldsymbol{N},\boldsymbol{N}'} \Theta(\boldsymbol{N}, \boldsymbol{N}')\, W_{\boldsymbol{N}\boldsymbol{N}'}\, P_{\boldsymbol{N}'}(t). \tag{3.59}$$

Consequently, the first law of thermodynamics has a closed mesoscopic representation

$$\mathrm{d}_t \langle E \rangle(t) = \langle \dot{Q} \rangle(t) + \langle \dot{W} \rangle(t), \tag{3.60}$$

which is identical to the microscopic first law (3.50). We note that after the coarse-graining the heat increment

$$Q(\boldsymbol{N},\boldsymbol{N}') = -\frac{1}{\beta}\ln\frac{W_{\boldsymbol{N}\boldsymbol{N}'}}{W_{\boldsymbol{N}'\boldsymbol{N}}} + S^{int}(\boldsymbol{N}) - S^{int}(\boldsymbol{N}') = -\frac{1}{\beta}\ln\frac{w_{\boldsymbol{N}\boldsymbol{N}'}}{w_{\boldsymbol{N}'\boldsymbol{N}}}, \tag{3.61}$$

is no longer directly given by the local detailed balance relation like in the microspace, cf. Eq. (3.53), but is supplemented by the change in internal entropy (3.19), see also Ref. [112].

We define the average system entropy in the mesospace as follows

$$\langle S\rangle(t) = \sum_{\boldsymbol{N}} P_{\boldsymbol{N}}(t)\left[\Omega(\boldsymbol{N}) - \ln P_{\boldsymbol{N}}(t)\right], \tag{3.62}$$

consisting of the non-equilibrium entropy defined by Eq. (3.49) and the ensemble average of the internal entropy accounting for the multiplicity of distinct microscopic configurations for a given mesostate. Analogously to Eq. (3.54), the mesoscopic entropy balance reads

$$\mathrm{d}_t\langle S\rangle(t) = \langle\dot{S}_e\rangle(t) + \langle\dot{\Sigma}\rangle(t), \tag{3.63}$$

with the mesoscopic entropy flow

$$\langle\dot{S}_e\rangle(t) = -\sum_{\boldsymbol{N},\boldsymbol{N}'} W_{\boldsymbol{N}\boldsymbol{N}'}\,P_{\boldsymbol{N}'}(t)\ln\frac{w_{\boldsymbol{N}\boldsymbol{N}'}}{w_{\boldsymbol{N}'\boldsymbol{N}}} = \beta\langle\dot{Q}\rangle(t), \tag{3.64}$$

and the mesoscopic entropy production rate

$$\langle\dot{\Sigma}\rangle(t) = \sum_{\boldsymbol{N},\boldsymbol{N}'} W_{\boldsymbol{N}\boldsymbol{N}'}\,P_{\boldsymbol{N}'}(t)\,\ln\frac{W_{\boldsymbol{N}\boldsymbol{N}'}P_{\boldsymbol{N}'}(t)}{W_{\boldsymbol{N}'\boldsymbol{N}}P_{\boldsymbol{N}}(t)} \geq 0. \tag{3.65}$$

A closer inspection yet reveals, that the definitions in Eqs. (3.62),(3.65) are in general not coinciding with those made at the microscopic level, *i.e.* $\langle S\rangle \neq \langle s\rangle, \langle\Sigma\rangle \neq \langle\sigma\rangle$. The nonlinearity of the system entropy and the entropy production [Eqs. (3.49), (3.56)] in the microstate probability p_α is incompatible with the coarse-graining (3.14). Instead, a marginalization of the microscopic second law (3.56) gives, in general, rise to additional entropic contributions which are dependent on microscopic information. As a result, the mesocsopic second law can, in general, not be closed.

However, there are two cases for which the entropies and thus the second law can be equivalently represented in the mesoscopic space. First, for the natural choice of a microscopic initial condition,

$$p_{\alpha'}(0) = P_{\boldsymbol{N}'}(0)/\Omega_{\boldsymbol{N}}, \tag{3.66}$$

where all microstates inside the respective mesostates are stationary and thus uniformly distributed. This also holds at later times since the energetics (3.2) and therefore the microscopic transition rates (3.5) do not discriminate between them. In this case, Eqs. (3.62) and (3.65) hold so that

$$\langle S\rangle(t) = \sum_{\boldsymbol{N}}\left[S^{int}_{\boldsymbol{N}} - \ln P_{\boldsymbol{N}}(t)\right]\,P_{\boldsymbol{N}}(t) = -\sum_{\alpha} p_\alpha(t)\,\ln p_\alpha(t) = \langle s\rangle(t), \tag{3.67}$$

and

$$\langle \dot{\Sigma} \rangle(t) = \sum_{\boldsymbol{N},\boldsymbol{N}'} \ln \frac{W_{\boldsymbol{N}\boldsymbol{N}'}\, P_{\boldsymbol{N}'}(t)}{W_{\boldsymbol{N}'\boldsymbol{N}}\, P_{\boldsymbol{N}}(t)}\, W_{\boldsymbol{N}\boldsymbol{N}'}\, P_{\boldsymbol{N}'}(t) = \sum_{\alpha,\alpha'} \ln \frac{w_{\alpha\alpha'}\, p_{\alpha'}(t)}{w_{\alpha'\alpha}\, p_{\alpha}(t)}\, w_{\alpha\alpha'}\, p_{\alpha'}(t) = \langle \dot{\sigma} \rangle(t) \geq 0. \tag{3.68}$$

Secondly, in the stationary state the microscopic master equation (3.4) reduces to $0 = \sum_j w_{ij} p_j$. Since the microscopic transition rates (3.5) do not depend on the individual microstate $\alpha_{\boldsymbol{N}}$ belonging to a given mesostate $\boldsymbol{N}$, it follows that the microscopic probability does not either in the stationary state so that

$$\mathbb{P}^s_{\alpha_{\boldsymbol{N}}} = \frac{1}{\Omega_{\boldsymbol{N}}}, \qquad p^s_{\alpha_{\boldsymbol{N}}} = \frac{P^s_{\boldsymbol{N}}}{\Omega_{\boldsymbol{N}}}. \tag{3.69}$$

For this particular case, it is true that

$$\langle S^s \rangle = \sum_{\boldsymbol{N}} \left[S^{int}_{\boldsymbol{N}} - \ln P^s_{\boldsymbol{N}} \right] P^s_{\boldsymbol{N}} = - \sum_{\alpha} p^s_{\alpha} \ln p^s_{\alpha} = \langle s^s \rangle, \tag{3.70}$$

and the mesoscopic second law reduces to the statement that the stationary entropy production rate $\langle \dot{\Sigma}^s \rangle$ is equal to minus the stationary entropy flow $\langle \dot{S}^s_e \rangle$,

$$\langle \dot{\Sigma}^s \rangle = \sum_{\boldsymbol{N},\boldsymbol{N}'} W_{\boldsymbol{N}\boldsymbol{N}'}\, P^s_{\boldsymbol{N}'} \ln \frac{w_{\boldsymbol{N}\boldsymbol{N}'}}{w_{\boldsymbol{N}'\boldsymbol{N}}} = -\langle \dot{S}^s_e \rangle = \sum_{\alpha,\alpha'} w_{\alpha\alpha'}\, p^s_{\alpha'} \ln \frac{w_{\alpha\alpha'}\, p^s_{\alpha'}}{w_{\alpha'\alpha}\, p^s_{\alpha}} = \langle \dot{\sigma}^s \rangle \geq 0. \tag{3.71}$$

We now turn to the mean-field case and consistently define the first law in this limit

$$\mathrm{d}_t\, \mathcal{E} = \sum_{i,j} \mathcal{E}_i\, k_{ij}\, \overline{n}_j = \dot{\mathcal{Q}} + \dot{\mathcal{W}}, \tag{3.72}$$

with the heat and work currents

$$\dot{\mathcal{Q}} = \sum_{i,j} \left(\mathcal{E}_i - \mathcal{E}_j - f\, \Theta(i,j) \right) k_{ij}\, \overline{n}_j \tag{3.73}$$

$$\dot{\mathcal{W}} = f \sum_{i,j} \Theta(i,j)\, k_{ij}\, \overline{n}_j\,, \tag{3.74}$$

where $i, j = 1, 2, 3$ specifies the state of the single mean-field unit. In analogy to Eq. (3.49), we write the system entropy in the mean-field limit as

$$\mathcal{S} = - \sum_{i} \overline{n}_i \ln \overline{n}_i, \tag{3.75}$$

which we split into the mean-field entropy flow

$$\dot{\mathcal{S}}_e = - \sum_{i,j} k_{ij}\, \overline{n}_j \ln \frac{k_{ij}}{k_{ji}} = \beta \dot{\mathcal{Q}}, \tag{3.76}$$

and the non-negative mean-field entropy production rate

$$\dot{\mathcal{S}}_i = \sum_{i,j} k_{ij}\,\overline{n}_j\,\ln\frac{k_{ij}\overline{n}_j}{k_{ji}\overline{n}_i} \geq 0. \tag{3.77}$$

As the mean-field represents the asymptotic limit of the mesospace, it holds that all the mesoscopic averages of the intensive observables $\langle O\rangle/N$, that are compatible with the coarse-graining (3.14), converge to the corresponding observables $\mathcal{X}$ in the mean-field limit, $\lim_{N\to\infty}\langle\dot{O}\rangle/N = \dot{\boldsymbol{O}}$, with $O = E, Q, W, S_e$.

Moreover, for the mean-field definitions in Eqs. (3.75) and (3.77) to represent the physical entropies, that is $\lim_{N\to\infty}\langle\dot{O}\rangle/N = \dot{\boldsymbol{O}}$, with $O = S, \Sigma$, we have to restrict to the two aforementioned cases [Eqs. (3.67), (3.68) and Eqs. (3.70), (3.71)] for which the mesoscopic second law coincides with the microscopic one. In the stationary case, the second law in the mean-field limit again boils down to the equivalence of the stationary entropy production and minus the stationary entropy flow,

$$\dot{\mathcal{S}}_i^s = \sum_{i,j} k_{ij}\,\overline{n}_j^s\,\ln\frac{k_{ij}}{k_{ji}} = -\dot{\mathcal{S}}_e^s \geq 0. \tag{3.78}$$

We have thus developed three different levels (microspace, mesospace and mean field) to consistently characterize the energetics of our model. For the first law, the lower levels of description are equivalent, while for the second law they only coincide for a class of initial conditions (3.66) and in the stationary limit (3.69). The same applies asymptotically in the macroscopic limit to the thermodynamic observables defined at the mean-field level.

Dissipated Work

With the thermodynamic framework developed in the preceding section at hand, we can now proceed by addressing one of the crucial research questions of this dissertation, that is the thermodynamics of non-equilibrium phase transitions. We are naturally interested in the (metastable) synchronization regime bounded by the two phase transitions. Since the nonstationary entropy production represented in the microscopace is, in general, not identical to the one in the mesospace, we characterize the nonequilibrium phase transitions via the dissipated work given by Eqs. (3.59) and (3.74). At metastable or infinite time, the work is observed to be always dissipative on average, that is the system takes up the energy from the nonconservative force, $\langle W\rangle > 0$, and dissipates it into the bath in the form of heat, $\langle Q\rangle < 0$, for all temperatures and system sizes.

In the following, we compare the dissipated work of an unit embedded in an interacting network and of an independent one. For the latter, an analytic expression can be derived. We consider a single unit with states $i = 1, 2, 3$ whose evolution is governed by the master equation

$$\partial_t P_i(t) = \sum_{i,j} W_{ij}\,P_j(t)\,, \tag{3.79}$$

where P_i is the probability to find the unit in the unit state i with the transition rates

$$W_{ij} = \mathrm{e}^{-\frac{\beta}{2}\left[\epsilon_i - \epsilon_j - f\,\Theta(i,j)\right]}, \tag{3.80}$$

with the sign function $\Theta(i,j)$ which is defined as in Eq. (3.33). The steady-state work current for a single unit reads

$$\langle \dot{W}_1^s \rangle = f \sum_{i,j} \Theta(i,j)\, W_{ij}\, P_j^s \,. \tag{3.81}$$

Using the spanning tree formula [150], one obtains for the stationary probabilities

$$P_1^s = \frac{a_1}{a_1 + a_2 + a_3}, \quad P_2^s = \frac{a_2}{a_1 + a_2 + a_3}, \tag{3.82}$$

where

$$\begin{aligned} a_1 &= W_{13}W_{12} + W_{12}W_{23} + W_{13}W_{32} \\ a_2 &= W_{23}W_{31} + W_{21}W_{13} + W_{23}W_{21} \\ a_3 &= W_{31}W_{12} + W_{32}W_{21} + W_{32}W_{31}, \end{aligned} \tag{3.83}$$

For a flat energy landscape, $\epsilon_i = \epsilon$, we indeed find that the symmetric stationary solution $P_i = 1/3$ is independent of β and f like the symmetric high-temperature solution of the mean-field equation (3.31). Next, the stationary work current is given by

$$\langle \dot{W}_1^s \rangle = 3f\, \frac{W_{13}W_{21}W_{32} - W_{31}W_{12}W_{23}}{W_{12}(W_{13} + W_{23} + W_{31}) + W_{13}(W_{21} + W_{32}) + (W_{21} + W_{31})(W_{13} + W_{23} + W_{31})}, \tag{3.84}$$

that can be simplified to

$$\langle \dot{W}_1^s \rangle = 2\,\Gamma f \sinh\left(f\beta/2\right). \tag{3.85}$$

Figure 3.13a) depicts the difference between the stationary work current of a single unit $\langle \dot{W}_1^s \rangle$ and the work current per unit in a N-unit network at times at the order of the metastable timescale (3.45), $\overline{W}_N \equiv \langle W \rangle / (Nt)|_{t\sim\tau_m}$, as a function of β and for different N. The work current $\overline{W}_N$ is numerically determined by solving Eqs. (3.14) and (3.31), respectively. As seen in Fig. 3.13a), the large ($N = 10^4$) system agrees excellently with the mean-field limit for all temperatures, while the smaller systems, albeit showing a qualitatively similar behavior, unlike the dynamics, deviate significantly.

Since the stationary single-unit work current $\langle \dot{W}_1^s \rangle$ is governed by a smooth and convex function, we observe that the dissipated mean-field work exhibits striking changes at the critical points $\beta_{c_{1,2}}$. The vicinities of these critical points are magnified in the two insets. The phase transitions in the dissipated mean-field work at β_{c_1} and β_{c_2} exhibit a kink and a saddle, respectively, and are therefore reminiscent of a first- and second-order equilibrium

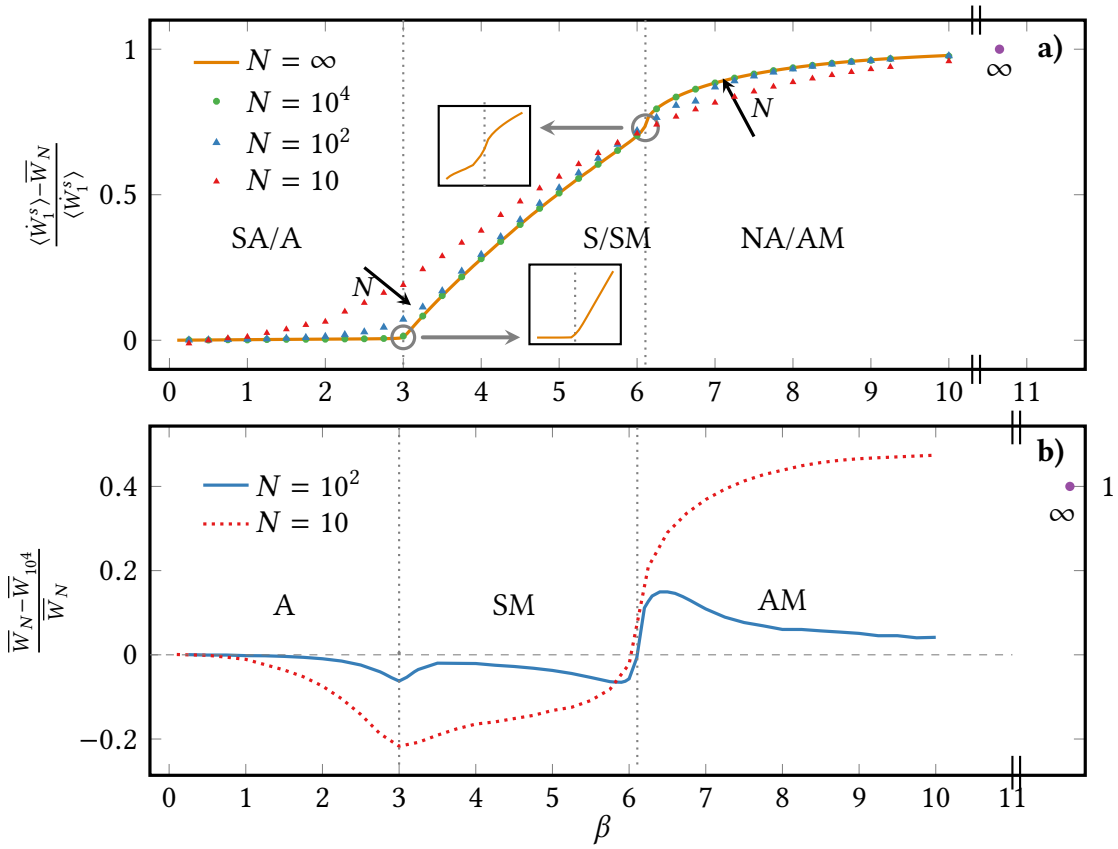

Figure 3.13: *Panel a): Difference of the dissipated work for a single-unit, $\langle \dot{W}_1^s \rangle$, and for a unit in a network of N interacting units, $\overline{W}_N$, for inverse temperatures $\beta = 0 \ldots 10$. The time t is chosen to be at the order of τ_m to ensure that $\overline{W}_N$ corresponds to its metastable value. Panel b): Difference of the dissipated work per unit for networks of different size with $N < 10^4$, for β ranging from 0 to 10 and thus covering all three phases. As in panel a), the time is at the order of τ_m. The purple closed circles in panels a) and b) represent the analytic expression given by Eqs. (3.87) and (3.88), respectively. In each plot all finite systems were simulated sampling 10^6 trajectories.*

phase transition, cf. Fig. 2.1 and the corresponding introduction in Sec. 2.1. Remarkably, owing to the metastability in the stochastic dynamics, sufficiently large but finite systems also exhibit signatures of these nonequilibrium phase transitions at the bifurcation points for times $t \sim \tau_m$ which blur out with decreasing system size.

In the high-temperature limit $\beta \to 0$ and for large N, the difference $\langle \dot{W}_1^s \rangle - \overline{W}_N$ goes to zero since the interaction energy (3.3) vanishes, more explicitly $N_i = N_j$ as $\beta \to 0$ and $u/N \to 0$ as $N \to \infty$. While for the mean field this holds true in the entire SA phase ($\beta < \beta_{c_1}$), for finite systems the range of β values in the A phase for which the interaction

energy is negligible decreases with N. Crucially, for temperatures below the critical Hopf one ($\beta > \beta_{c_1}$), we find that interactions always reduce the costs to maintain the system in its nonequilibrium state, that is $\langle \dot{W}_1^s \rangle - \overline{W}_N > 0$. We furthermore observe in Fig. 3.13a) that this difference in work dissipation is a monotonously increasing function of β and becomes infinitely large in the low-temperature limit. This asymptotic limit can be seen as follows. First, we have for the dissipated work per unit in a finite network

$$\lim_{\beta\to\infty} \overline{W}_N = \lim_{\beta\to\infty} \Gamma f \left(\mathrm{e}^{\beta f} - 1\right) \mathrm{e}^{-\frac{\beta(fN-Nu+u)}{2N}}, \tag{3.86}$$

which is subdominant to $\langle \dot{W}_1^s \rangle$ that has been analytically determined in Eq. (3.84),

$$\lim_{\beta\to\infty} \frac{\Delta\overline{W}}{\langle \dot{W}_1^s \rangle} = \lim_{\beta\to\infty} \left[1 - \mathrm{e}^{\frac{\beta u(N-1)}{2N}}\right] = 1. \tag{3.87}$$

Hence we have shown that at low and intermediate temperatures an interacting network of any size is energetically favorable with respect to a noninteracting one. Interestingly, in the the two phases of higher temperature, the operational costs per unit can be further decreased by employing *smaller* networks. As one approaches the second critical point the different curves intersect and in the NA/AM phase the operation of *larger* networks gives rise to less work dissipation per unit.

This is also illustrated in Fig. 3.13b) that depicts the difference in the dissipated work between a system of size $N = 10^4$ exhibiting metastability and a smaller system $N = 10^2, 10^3$ which does not display the latter. In agreement with panel a), the smaller system requires less input per unit to be maintained in the two higher temperature phases (A and SM phase), that is $\overline{W}_N - \overline{W}_{10^4} < 0$ while the opposite holds true in the AM phase, $\overline{W}_N - \overline{W}_{10^4} > 0$. Again, we observe significant changes at the critical points: At the first critical point the difference $\overline{W}_N - \overline{W}_{10^4} < 0$ takes a local minimum and at the second critical point it changes sharply around an inflection point. It is plausible that these changes are more pronounced for decreasing N since in this case the "distance" to metastable behavior and thus signatures of nonequilibrium phase transition exhibited by the reference system ($N' = 10^4$) is increasing.

For the same reasons as stated in the context of Fig. 3.13a), $\overline{W}_N - \overline{W}_{10^4}$ goes to zero in the high-temperature regime, while in the low-temperature limit one obtains

$$\lim_{\beta\to\infty} \frac{\overline{W}_N - \overline{W}_{10^4}}{\overline{W}_N} = \lim_{\beta\to\infty} 1 - \mathrm{e}^{-\frac{\beta u}{2}\left(\frac{1}{10^4} - \frac{1}{N}\right)} = 1, \tag{3.88}$$

if $N < 10^4$. This limit is illustrated by the purple closed circle in plot b). For the larger system the work difference is decreasing in the range of available data. Generating data for larger β to monitor the convergence to the low-temperature limit is not possible since the simulation becomes numerically unstable owing to the large values the exponentials take in the mesoscopic transition rates (3.21).

To illustrate the data underlying the plots in Fig. 3.13b), we show in Figs. 3.14a)–c) the time-scaled work asymptotics per unit for different system sizes as chosen for the blue curve

in panel b) as well as the mean-field limit for selected values of $\beta = 2, 4, 7$. We note the excellent agreement between the mean-field limit and the large system in compliance with the observations made in panel 3.14a). On the other hand, the small system clearly deviates from the large systems in all three different regimes, even though we observed that in the SA/A phase the dynamics of large and small systems can not be distinguished, see Fig. 3.11. Due to the approximate time-periodicity of the dynamics in the S/SM phase, the dissipated work is also oscillating.

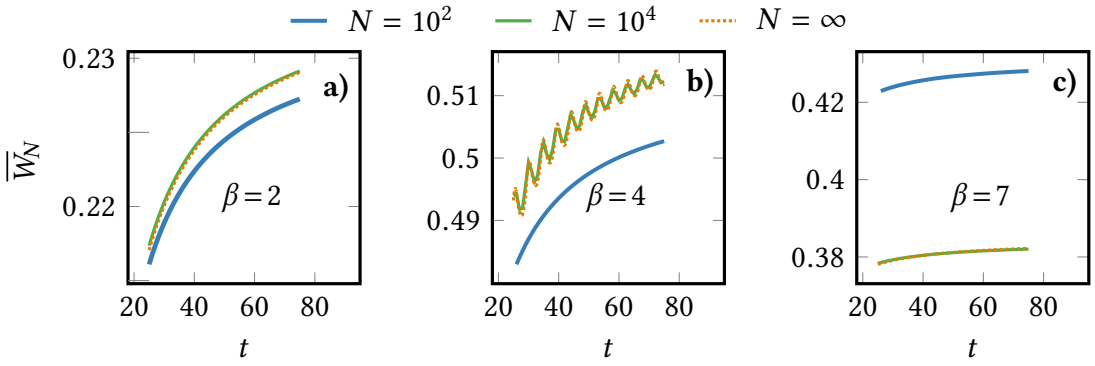

Figure 3.14: *Panels a)–c): Plot of $\overline{W}_N$ for selected values at $\beta = 2, 4, 7$ and system sizes $N = 10^2$ and $N = 10^4$. This is the same data as the one underlying the blue solid curve in Fig. 3.13b) but, for better visualization, the time t is restricted. For comparison, the mean-field limit is overlaid in the panels a)–c). In each plot all finite systems were simulated sampling 10^6 trajectories.*

Finally, Fig. 3.15 depicts the difference between the stationary single-unit and the asymptotic mean-field work current, $\Delta\overline{W}_{1\infty}$, as a function of β for different f. Again, $\Delta\overline{W}_{1\infty} = 0$ in the SA phase since the single and the mean-field unit are indistinguishable above the critical Hopf temperature β_{c_1} as already observed in Fig. 3.13a). For $\beta \geq \beta_{c_1}$ the second critical point is gradually shifting to smaller β [cf. Fig. 3.2] while the difference $\Delta\overline{W}_{1\infty}$ is monotonously increasing with decreasing f. Therefore, if compared to the mean field, the additional costs to maintain the nonequilibrium stationary state of the noninteracting system at a given temperature are the smaller the further it is driven out-of-equilibrium. This implies in particular that the dissipation of the synchronized system at fixed temperature is approaching the one of the non-synchronized system as they are further driven out-of-equilibrium.

To summarize, we have obtained two major results in this section. First, though the nonequilibrium phase transitions are naturally only present in the mean-field limit where the nonlinear dynamics exhibits the supercritical Hopf and the infinite-period bifurcation, we find that the metastability observed in the finite-system dynamics translates into signatures of genuine nonequilibrium phase transition in finite systems. This consistently connects linear stochastic dynamics, nonlinear deterministic dynamics, and thermodynamics and furthermore demonstrates that thermodynamics of nonequilibrium phase transitions and

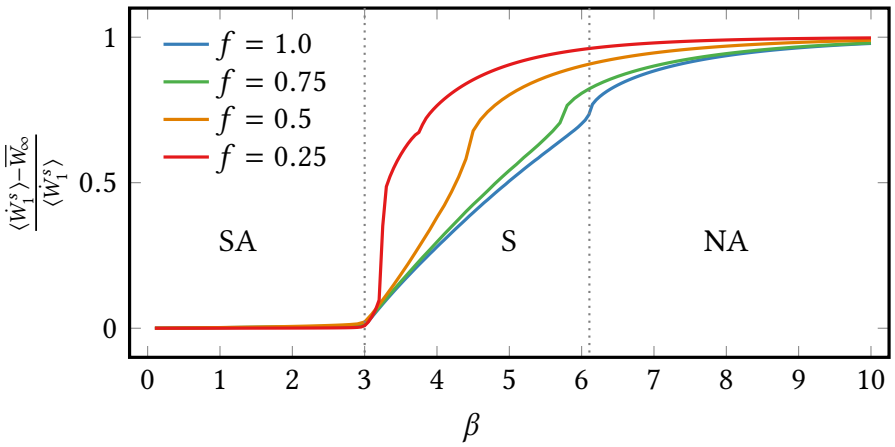

Figure 3.15: *Comparison between stationary single-unit work current $\langle \dot{W}_1^s \rangle$ and asymptotic mean-field work current $\overline{W}_\infty$ for different values of f.*

bifurcation theory are closely related. Secondly, any finite and attractive interaction in a network reduces the dissipated work per unit. Interestingly, if operating in the synchronous phase, it is even more economic to employ interacting but *smaller* networks. What is still open to investigate is how the nonequilibrium phase transitions affect the power-efficiency trade-off, if the system operates as an energy-converting machine.

3.2.1.7 Efficiency at Maximum Power

In order to construct such an energy converter with our system both a positive force $f_1 > 0$ and a negative force $f_2 < 0$ are applied on the same unit. Examples for this type of work-to-work conversion are could be double quantum dot channel capacitively coupled to a quantum point contact [34] or the biological motors kinesin and myosin. In the latter case, the motor is driven forward with f_1 by extracting energy via ATP hydrolysis while the load carried by the motor is modeled as f_2 [66, 151]. In general, these two forces obey two different distributions accounting for the crucial fluctuations these motors exhibit. Since the following discussion is restricted to the mean-field limit, we consider the homogeneous case where the same positive and negative force are applied on all units. We remark that homogenous motors modeled as diffusing particles on a lattice subjected to an exclusion rule were studied in [65] while a noisy Kuramoto model resembling molecular motors with nontrivial force distributions was investigated in [66].

We thus decompose the net force $f = f_1 + f_2$ into the driving force $f_1 > 0$ and the load force $f_2 < 0$. The asymptotic work contributions are given by integrating Eq. (3.74) over times $t \gg \tau_r$ and are denoted by $\overline{\mathcal{W}}_1$ and $\overline{\mathcal{W}}_2$, respectively. Substituting Eq. (3.33) into Eq. (3.78), yields the following decomposition of the asymptotic entropy production in the

mean-field limit

$$\overline{\mathcal{S}_i} = \beta\,\overline{\mathcal{W}}_1 + \beta\,\overline{\mathcal{W}}_2. \tag{3.89}$$

Using the definition (2.55) for the asymptotic efficiency of this work-to-work conversion,

$$\overline{\eta} = -\frac{\overline{\mathcal{W}}_2}{\overline{\mathcal{W}}_1} = -\frac{f_2}{f_1} = 1 - \frac{f}{f_1}. \tag{3.90}$$

At equilibrium ($f = 0$), the reversible limit, $\eta_c = 1$, is attained while out of equilibrium ($f \neq 0$) the efficiency is bounded, $0 < \eta < \eta_c$. Of particular interest is the efficiency at maximum power, η^*, which results from the optimization of the asymptotic output power

$$\overline{\mathcal{P}} \equiv \frac{\partial \overline{\mathcal{W}}_2}{\partial t}, \tag{3.91}$$

with respect to the output force,

$$\eta^* = -\left.\frac{f_2}{f_1}\right|_{f_2 = f_2^*}. \tag{3.92}$$

The maximization parameter f^* is determined by the condition $\partial\overline{\mathcal{P}}/\partial f_2|_{f_2=f_2^*} = 0$, while fixing $f_1 = 1$ and thus varying the total dissipation characterized by f.

In the SA phase, $\beta < \beta_{c_1}$, the asymptotic power output coincides with the stationary work current of a single unit (3.84). For the other two phases (S and NA), we have to resort to simulations to obtain the asymptotic power output. Moreover, owing to the time-periodic state in the S phase, the power is periodically changing in time. Hence we consider the time-average of the power over one limit cycle period. Figure 3.16a) shows the asymptotic output power $\overline{\mathcal{P}}$ as a function of β and f_2 in a density plot.
The white dashed lines indicate the critical points $\beta_{c_{1,2}}$ as a function of the output force. Thus the area enclosed by those lines corresponds to the S phase. Remarkably, we find that the maximum output power is generated in this phase. In particular, the global maximum of the output power indicated by the purple closed circle lies inside the S phase. At large β that represents the NA phase, the generated power rapidly drops. In panel b) the output power maximized with respect to the output force for different values of the inverse temperature is depicted. The numerical data from panel a) is overlaid with the (semi-)analytic results in the SA phase (green solid line) and the low-temperature limit and shows an excellent agreement. These limiting cases can be obtained as follows. In the SA phase, the condition for maximization of the power

$$\frac{\partial \mathcal{P}}{\partial f_2} = \left[2 + \beta f_2 + \mathrm{e}^{\beta(1-f_2)}(\beta f_2 - 2)\right] = 0, \tag{3.93}$$

results in a transcendental equation that must be treated numerically. In the low-temperature limit, the extremum condition

$$\frac{\partial \mathcal{P}}{\partial f_2} = \mathrm{e}^{\frac{\beta}{2}(f_2-1)}\left[\mathrm{e}^{\beta(1-f_2)}(\beta f_2 - 2) + (\beta f_2 + 2)\right] = 0, \tag{3.94}$$

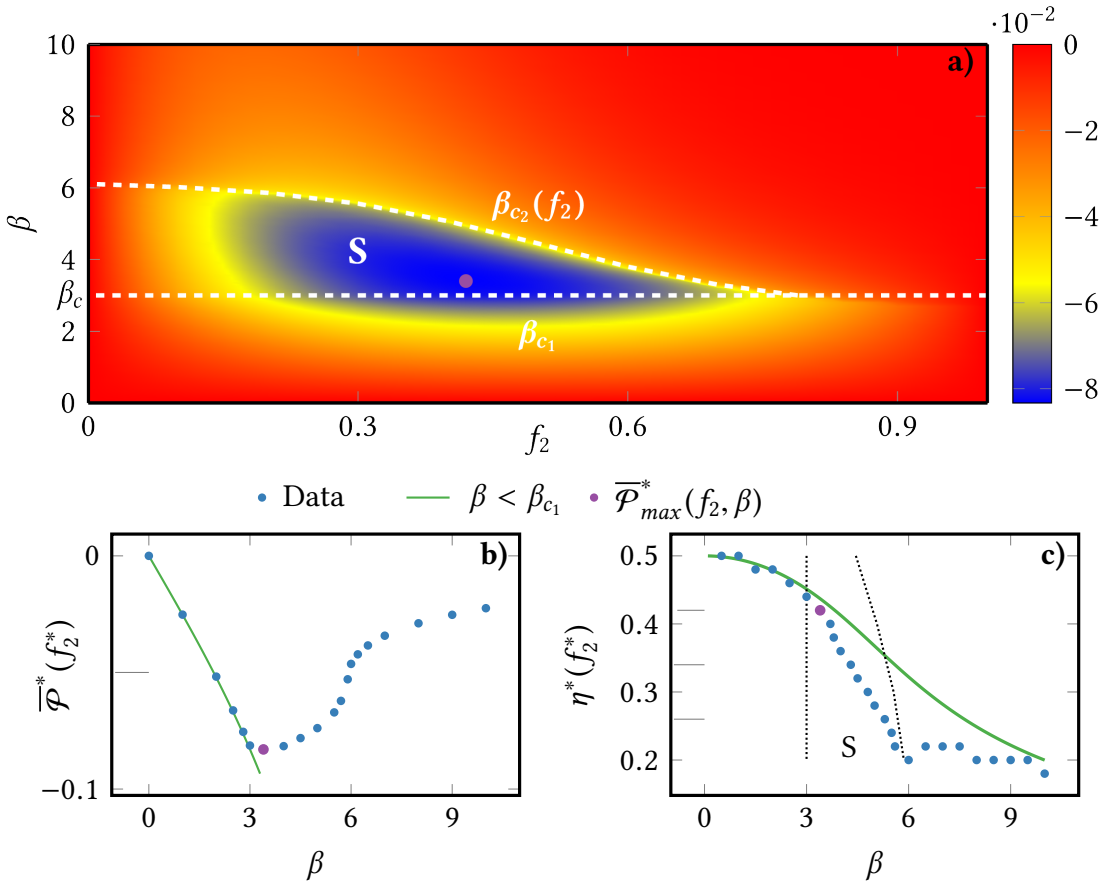

Figure 3.16: *Depiction of the asymtpotic output power $\overline{\mathcal{P}}$ in a) as a function of the output force f_2 and the inverse temperature β. The white dashed lines correspond to the numerically determined critical points as a function of the output force. Hence the enclosed area defines the synchronization phase S. The global maximum of the output power is indicated by the purple closed circle. In panel b) the maximum output power $\overline{\mathcal{P}}^*$ is optimized with respect to f_2 and in panel c) the associated efficiency at maximum power $\eta^*(f_2^*)$ is displayed. In panel c) the dashed lines specify the critical points and the synchronization phase S. The efficiency at the global maximum power is indicated by the closed purple circle. The analytic solution for $\beta < \beta_c$ is overlaid with the numerical data in the lower panels.*

can not be satisfied for any f_2 compatible with the constraint $\beta = \infty$.

The efficiencies associated with the processes corresponding to the data points in panel 3.16b) are depicted in panel c). Again, the semianalytic solution for the temperatures corresponding to the SA phase (green solid line) is compared with the numerical results and shows an excellent agreement at these tmeperatures. As β approaches zero, the efficiency at maximum power becomes $\eta^* = \eta_c/2 = 1/2$ corresponding to the universal linear-response

value for tightly-coupled (only one net-current) systems, see also the first-order expansion of the Curzon-Ahlborn efficiency (2.56). This can be seen by expanding the analytic solution for the the stationary power output in the SA phase (3.84) up to first order in β and maximizing the truncated expansion with respect to f_2. This, in turn, yields the linear-response relation $J^s = L\,f$ with the Onsager coefficient $L = 2\Gamma\beta$. Therefore, small products βf correspond to linear response in our model and lead to efficiency at maximum power values very close to 1/2. With increasing β, the system starts to respond nonlinearly and the efficiency decreases monotonously and nonlinearly.

It is worth emphasizing that the efficiency for the global maximum power output achieved in the far-from-equilibrium S phase and indicated by the purple closed circle is still close to the universal linear-response efficiency at maximum power value. This finding points out the importance of non-equilibrium phase transitions for the performance of an assembly of nano-machines and suggests synchronization as an operating mode facilitating very efficient energy-conversion processes with appreciable power output.

3.2.2 Class of Driven Potts Models

3.2.2.1 Setup

The three-state model as proposed in Fig. 3.1 is a special realization ($q = 3$) of a class of Potts models [152] that generalize the Ising model ($q = 2$) [153, 154] by considering interacting spins on a lattice that can take q different values distributed uniformly about a circle. It is intriguing to study how and if the (thermo)dynamical phenomenology changes with q. Analogously to the procedure detailed in Secs. 3.2.1.3 and 3.2.1.4, we can formulate the problem at the microscopic and mesoscopic level and arrive at a mean-field description that is exact in the macroscopic limit $N \to \infty$. Obviously, the convergence of the irreducible linear Markov flow in finite-systems to possibly complex mean-field solutions can again be understood via notions of metastability that is encoded in the spectrum of the Markov generator, as demonstrated in Sec. 3.2.1.5. To avoid redundancy, we restrict the discussion to the mean-field description and focus on studying the (thermo)dynamical phenomenology for different q.

For this purpose, we consider infinitely many q-state units on a ring with energies ϵ_i ($i = 1, 2, \ldots q$). The units correspond to clocks since transitions are only allowed between adjacent unit states. Units that occupy the same state interact globally via an attractive interaction potential u. The system is autonomously driven by a global and non-conservative force f. In addition, the system is connected with a heat bath at inverse temperature β. The schematics of the setup for a finite-size eight-state model ($q = 8$) is depicted in Fig. 3.17.

The mean-field system is fully characterized by the occupation densities of the states $i = 1, \ldots, q$ which we again identify as the probability $\overline{n}_i(t)$ for any unit in the continuum to occupy the state i at time t. Following the path from a microscopic over a mesoscopic to

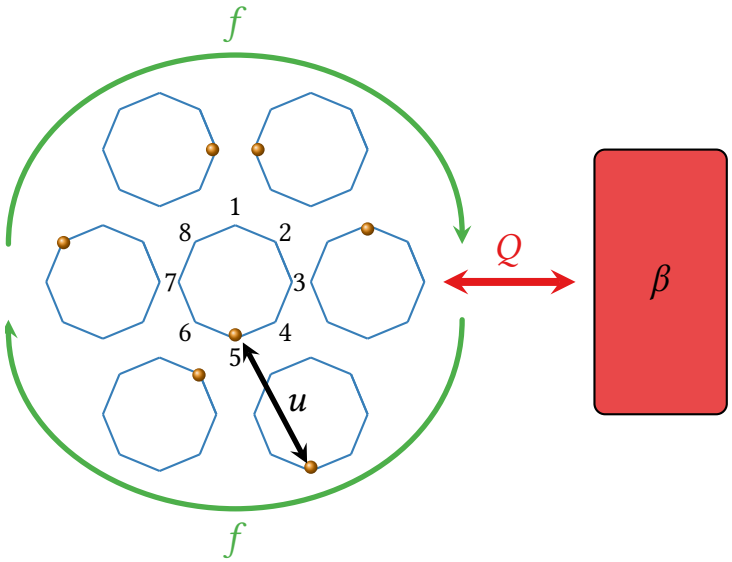

Figure 3.17: *Schematics of identical and globally interacting q-state units that are connected with a heat bath at inverse temperature β and are furthermore subjected to a nonconservative rotational force f.*

a macroscopic representation of the system as in three-state case [cf. Eqs. (3.4), (3.14) and (3.31)], we arrive at the following mean-field equation

$$\partial_t \overline{n}_i(t) = \sum_j k_{ij}\, \overline{n}_j(t)\,, \tag{3.95}$$

with the mean-field transition rates from Eq. (3.32) that satisfy the local detailed balance property,

$$\ln \frac{k_{ij}}{k_{ij}} = -\beta \left[\mathcal{E}_i(t) - \mathcal{E}_j(t) - f\,\Theta(i,j)\right], \quad \mathcal{E}_i(t) = \epsilon_i + u\overline{n}_i(t), \tag{3.96}$$

ensuring that the system is thermodynamically consistent. We note that probability conservation, $\sum \overline{n}_i = 1$, erases one degree of freedom such that the nonlinear system has $q-1$ dimensions.

3.2.2.2 Mean-Field Dynamics

Universal High- and Low-Temperature Phases

The highly nonlinear character of Eq. (3.95) can give rise to complex dynamics and does, in general, not admit an analytic solution. Yet, for a flat energy landscape of the units, $\epsilon_i = \epsilon$, the uniform probability distribution

$$\overline{n}_i^s = \frac{1}{q}, \quad i = 1, 2, \ldots, q-1, \tag{3.97}$$

is a solution to Eq. (3.95). However, to infer the stability of that fixed point the computation of the spectrum of the linearized Jacobian (3.37) is required. Though, for $q > 4$ there is no systematic way to determine the eigenvalues of the Jacobian, since there is no general formula to solve polynomial equations of order higher than four.

However, the local detailed balance condition (3.96) ensures that the system is thermodynamically consistent. As a consequence, the possible dynamical phenomenology is constrained by the thermodynamic principles that even allow specify the long-time solution in the low- and high-temperature regime as already demonstrated in the three-state model in Sec. 3.2.1.4. First, we note that the high-temperature limit $\beta \to 0$ represents a reversible limit for finite f since detailed balance $k_{ij}\,\overline{n}_j^{eq} = k_{ji}\,\overline{n}_i^{eq}$, $\forall i, j$ is attained. Equilibrium statistical mechanics predicts that in the high-temperature limit the system behaves entropically and the probability distribution is uniform. Physically, at high temperatures the autonomous driving becomes negligible and the forward-backward transition are indistinguishable. This means that the solution (3.97) represents a stable and unique fixed point in the high-temperature limit.

On the other hand, the low-temperature limit, $\beta \to \infty$ represents a totally irreversible limit for finite driving, where only those transitions occur that are aligned with the bias f. Eq. (3.95) now reduces to

$$\partial_t \overline{n}_i(t) = k_{i,i-1}\,\overline{n}_{i-1}(t) - k_{i+1,i}\,\overline{n}_i(t), \tag{3.98}$$

where $i \pm 1 = (i \pm 1) \mod q$. A closer inspection reveals that since the transition rates are irreversible and the force has no spatial resolution, the latter is acting like a renormalization of the kinetic prefactor $\Gamma \to \Gamma \cdot \exp[\beta f/2]$ corresponding to an increasing number of transitions per unit time aligned with f along the ring. Then, the rewritten transition rates $k_{ij} = \Gamma \exp[-\beta u/2(\overline{n}_i - \overline{n}_j)]$ suggest that occupation is the only relevant factor that determines where the dynamics goes to in the long-time limit, *i.e.* the system with irreversible rates behaves energetically like an equilibrium one would. Here, one has to distinguish between repulsive ($u > 0$) and attractive ($u < 0$) interactions. For the former the system has a unique energy ground state that coincides with the entropic state from Eq. (3.97). Conversely, for attractive interactions there are q energy ground states where all units occupy the same state

$$\overline{n}_i^s = 1,\ \overline{n}_{j\neq i}^s = 0, \quad i = 1, 2, \ldots, q. \tag{3.99}$$

To which one the dynamics is striving depends on which single-unit state i is the most populated one at initial times and hence depends on the initial condition. The physical interpretation of the energy ground state is obvious: For repulsive interactions the system tries to avoid accumulation of units occupying the same state, while for attractive interactions the system favors clustering. To sum up, our thermodynamic framework allows us to exclude complex solutions (*e.g.* limit cycles) in the limits of low and high temperature and to even specify the stationary solutions in these regimes without explicitly solving the highly nonlinear equations of motion (3.95) that become exceedingly complicated for larger q. We emphasize that these results hold for any finite number of states q and finite autonomous driving f.

q-Dependent Intermediate-Temperature Phase

With the results for high and low temperatures at hand, we proceed by studying the qualitative behavior of the system at intermediate temperatures. For repulsive interactions there is the possibility that the system exhibits no significant qualitative changes in the dynamics as the temperature varies. If attractive interactions are considered the different dynamical behavior in the low- and high-temperature limit encoded in Eqs. (3.97) and (3.99) implies at least one bifurcation (phase transition) at intermediate temperatures. The stability of the symmetric fixed point, $\overline{\boldsymbol{n}}^{ss}$, is encoded in the spectrum of the linearized Jacobian (3.37).

In the high-temperature limit, $\beta \to 0$, the linear stability matrix for $q > 2$ reads

$$\begin{aligned} \boldsymbol{J}_{ij} &= -2\Gamma\delta_{ij} + \Gamma(\delta_{i+1,j} + \delta_{i-1,j})\big|_{i,j\neq 0,q} + \Gamma(\delta_{i,0}\delta_{q,j} + \delta_{i,q}\delta_{0,j}) \\ \boldsymbol{J} &= \begin{pmatrix} -2\Gamma & \Gamma & 0 & \dots & 0 & 0 & \Gamma \\ \Gamma & -2\Gamma & \Gamma & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \Gamma & -2\Gamma & \Gamma \\ \Gamma & 0 & 0 & \dots & 0 & \Gamma & -2\Gamma \end{pmatrix}_{q\times q} . \end{aligned} \tag{3.100}$$

This is a circulant matrix - a special kind of Toeplitz matrix [155] - where all remaining columns are given by cyclical permutations of the first column vector. This particular case of a symmetric circulant matrix admits the eigenvalues [156]

$$\lambda_k = -2\Gamma + \Gamma\, \mathrm{e}^{\frac{2\pi \mathrm{i} k}{q}} + \Gamma\, \mathrm{e}^{\frac{2\pi \mathrm{i}(q-1)k}{q}} = -2\Gamma + 2\Gamma \cos k\frac{2\pi}{q}, \quad k = 0, 1, \dots, q-1. \tag{3.101}$$

For $q = 2$ the linear stability matrix $\boldsymbol{J}_{ij} = -2\Gamma\delta_{ij} + 2\Gamma(\delta_{12} + \delta_{21})$ admits the eigenvalues $\lambda_{1,2} = 0, -4\Gamma$.

We now consider finite inverse temperatures, $\beta = 0 + \delta\beta$, and expand the linear stability matrix in Eq. (3.100) up to linear order in $\delta\beta$ and arrive at

$$\boldsymbol{J}(0 + \delta\beta) = \boldsymbol{J}(\beta)\big|_0 + \delta\beta\, \widehat{\boldsymbol{J}} + \boldsymbol{O}(\delta\beta^2), \tag{3.102}$$

with the first-order correction matrix

$$\begin{aligned} \widehat{\boldsymbol{J}}_{ij} &= a\delta_{ij} + b(\delta_{i+1,j} + \delta_{i,q}\delta_{0,j}) + c(\delta_{i-1,j} + \delta_{i,0}\,\delta_{q,j}), \\ \widehat{\boldsymbol{J}} &= \begin{pmatrix} a & b & 0 & \dots & 0 & 0 & c \\ c & a & b & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & c & a & b \\ b & 0 & 0 & \dots & 0 & c & a \end{pmatrix}_{q\times q}, \quad a = -\tfrac{2\Gamma u}{2}, \quad b = \Gamma\tfrac{2u - fq}{2q}, \quad c = \Gamma\tfrac{2u + fq}{2q}, \end{aligned} \tag{3.103}$$

that is again a circulant matrix and consequently has the eigenvalues [156]

$$\widehat{\lambda}_k = a + b\, \mathrm{e}^{\frac{2\pi \mathrm{i} k}{q}} + c\, \mathrm{e}^{\frac{2\pi \mathrm{i}(q-1)k}{q}} = -\tfrac{4\Gamma u}{q} \sin^2\left(k\tfrac{\pi}{q}\right) - \mathrm{i}\,\Gamma f \sin\left(k\tfrac{2\pi}{q}\right), \; k = 0, 1, \dots, q-1. \tag{3.104}$$

Since the real part of the eigenvalues in Eq. (3.103) is always nonpositive, $\mathrm{Re}[\widehat{\lambda}_k] \leq 0 \ \forall k$, the real part of all perturbed eigenvalues

$$\mathrm{Re}[\lambda_k(0+\delta\beta)] = \mathrm{Re}[\lambda_k(\beta)|_0] + \delta\beta \ \mathrm{Re}[\widehat{\lambda}_k] + O\left(\delta\beta^2\right), \tag{3.105}$$

remain negative for finite inverse temperature, hence reconfirming that in the high-temperature regime the symmetric fixed point is stable.

For $q \leq 4$ the linear stability analysis is analytically tractable for all temperatures since the characteristic equation can always be solved for up to four dimensions. We make the crucial observation that the critical point β_c that characterizes the bifurcation of the symmetric fixed point $\overline{\boldsymbol{n}}^{ss}$ obeys the relation

$$q + \beta_c u = 0, \quad q \leq 4. \tag{3.106}$$

For repulsive interactions this relation is never satisfied and the system remains in the symmetric fixed point at all temperatures. However, for attractive interactions there is a much richer phenomenology as demonstrated in the following. Evaluated at this temperature, the linear stability matrix for $q > 2$ generically takes the form

$$\boldsymbol{J}_{ij} = c\left[(\delta_{i+1,j} - \delta_{i-1,j})\big|_{i,j\neq 0,q} + \delta_{i,q}\,\delta_{0,j} - \delta_{i,0}\,\delta_{q,j}\right] \tag{3.107}$$

$$\boldsymbol{J} = \begin{pmatrix} 0 & c & 0 & \dots & 0 & 0 & -c \\ -c & 0 & c & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -c & 0 & c \\ c & 0 & 0 & \dots & 0 & -c & 0 \end{pmatrix}_{q\times q}, \quad c = \Gamma \sinh\left(q\,\frac{f}{2u}\right).$$

Here, the circulant matrix is skew-symmetric and its eigenvalues read

$$\lambda_k = c\,\mathrm{e}^{\frac{2\pi \mathrm{i} k}{q}} - c\,\mathrm{e}^{\frac{2\pi \mathrm{i}(q-1)k}{q}} = 2\mathrm{i}\,c \sin\left(k\frac{2\pi}{q}\right), \quad k = 0, 1, \dots, q-1, \tag{3.108}$$

and are thus either identical zero or purely imaginary. For $q = 2$ the linear stability matrix, $\boldsymbol{J}_{ij} = 0$, has only zero eigenvalues. We now consider temperatures in the vicinity of the critical temperature, $\beta_c + \delta\beta$, and expand the linear stability matrix in Eq. (3.107) up to linear order in $\delta\beta$ and arrive at

$$\boldsymbol{J}(\beta_c + \delta\beta) = \boldsymbol{J}(\beta)\big|_{\beta_c} + \delta\beta\ \widehat{\boldsymbol{J}} + O\left(\delta\beta^2\right), \tag{3.109}$$

with the first-order matrix

$$\widehat{\boldsymbol{J}}_{ij} = a\delta_{ij} + b(\delta_{i+1,j} + \delta_{i,q}\delta_{0,j}) + c(\delta_{i-1,j} + \delta_{i,0}\,\delta_{q,j}) \tag{3.110}$$

$$\widehat{\boldsymbol{J}} = \begin{pmatrix} a & b & 0 & \dots & 0 & 0 & c \\ c & a & b & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & c & a & b \\ b & 0 & 0 & \dots & 0 & c & a \end{pmatrix}_{q\times q}, \ a = \tfrac{2\Gamma u}{q} d, \ b = \Gamma \tfrac{2u - qf}{2q} d, \ c = \Gamma \tfrac{qf + 2u}{2q} d, \ d = \cosh\left(q\tfrac{f}{2u}\right),$$

that is again a circulant matrix and consequently has the eigenvalues

$$\widehat{\lambda}_k = a + b\,\mathrm{e}^{\frac{2\pi \mathrm{i}k}{q}} + c\,\mathrm{e}^{\frac{2\pi \mathrm{i}(q-1)k}{q}} = -\frac{2\Gamma\cosh\left(q\frac{f}{2u}\right)\sin^2\left(\frac{k\pi}{q}\right)}{q}\left[2u + \mathrm{i}fq\cot\left(\frac{k\pi}{q}\right)\right],\ k = 0, 1, \ldots, q-1. \tag{3.111}$$

We note that the real part of the eigenvalues in Eq. (3.110) is always nonnegative, $\mathrm{Re}[\widehat{\lambda}_k] \geq 0\ \forall k$, such that the real part of all perturbed eigenvalues

$$\mathrm{Re}[\lambda_k(\beta_c + \delta\beta)] = \mathrm{Re}[\lambda_k(\beta)\big|_{\beta_c}] + \delta\beta\ \mathrm{Re}[\widehat{\lambda}_k] + O\left(\delta\beta^2\right), \tag{3.112}$$

changes its sign at the critical point from negative ($\delta\beta < 0$) to positive values ($\delta\beta > 0$) or remains zero for all q.

Hence Eq. (3.106) indeed characterizes the critical point that destabilizes the symmetric fixed point. If the system is in equilibrium, $f = 0$, the real parts of the eigenvalues of the linear stability matrix evaluated at the critical point change their sign while the imaginary part is identically zero corresponding to a saddle-node bifurcation that degenerates the single symmetric fixed point into q fixed points. In the out-of-equilibrium scenario, $f \neq 0$, the real parts of the eigenvalues of the linearized Jacobian evaluated at the critical temperature change their sign while the imaginary parts remain finite implying a Hopf-bifurcation that degenerates the symmetric fixed point into a limit cycle.

Moreover, Eq. (3.106) states that the uniform probability distribution can be observed over a larger range of temperatures as q increases. The uniform distribution, however, removes the interactions from the dynamics according to Eq. (3.96). Hence for repulsive interactions the mean-field system is noninteracting as well as for attractive interactions at temperatures above the first critical temperature defined by Eq. (3.106) that approaches zero as q becomes large. Due to the richer phenomenology for attractive interactions we focus on this type of interaction and we set $u = -1$ if not explicitly stated otherwise. Physically, Eq. (3.106) can be understood by noting that the energy required to move between adjacent states is shrinking to zero as the density of states increases. Thus, spontaneous fluctuations can induce these transitions such that there is no instantaneous accumulation of any state (*e.g.* limit cycles) and the symmetric fixed point [Eq. (3.97)] is stable.

To infer the stability of the oscillations, an analysis of the normal form of the Hopf bifurcation and the determination of the first Lyapunov coefficient would be required as done for the three-state model in Sec. A.2. The normal form analysis relies on Eq. (3.95), hence for $q > 3$ the dimensionality of the system renders analytic progress difficult. A numerical analysis in Fig. 3.18 that depicts in a parametric $\overline{n}_1 - \overline{n}_2$ plot the dynamics of the q-state model ($q = 3, 4, 5, 6$) motivates the following *conjecture*: If q is even like in panels b) and d), the Hopf bifurcation occurs subcritical and any perturbation will lead to the destruction of the limit cycle and to the emergence of q stable fixed points, hence there is only one phase transition at β_c. Conversely, if q is odd, like in panels a) and c), the Hopf-bifurcation occurs supercritical and the symmetric fixed point degenerates at β_{c_1} into a stable limit cycle that at a second critical point β_{c_2} degenerates via an infinite-period bifurcation into q asymmetric

stable fixed points. They are related to each other by permutations of their coordinates in the q-dimensional probability space. For low temperatures these fixed points move towards the respective energy ground states in Eq. (3.99).

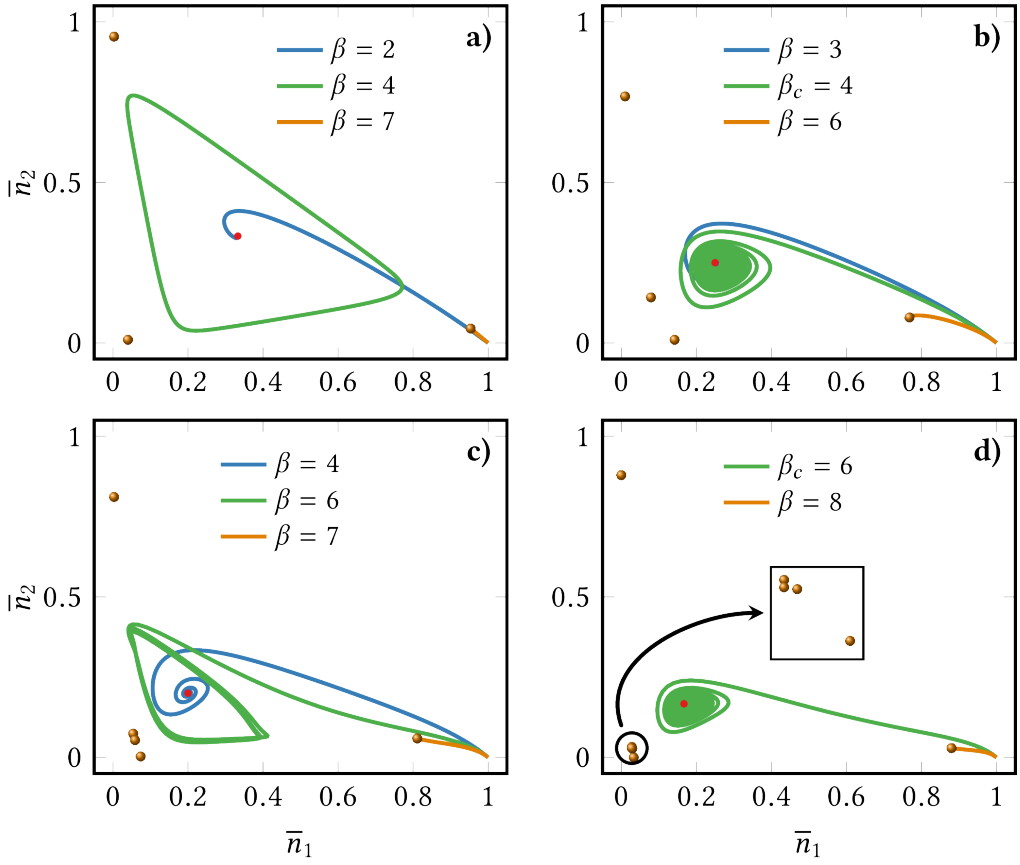

Figure 3.18: *Parametric plot of the probabilities $\overline{n}_{1,2}$ in the q-state model with $q = 3$ [panel a)], $q = 4$ [panel b)], $q = 5$ [panel c)] and $q = 6$ [panel d)] for different temperatures β. The orange-shade spheres indicate the position of the q fixed points in the low-temperature regime and the symmetric fixed point is indicated by the red closed circle. In all plots the intial condition $\overline{n}_1 = 1$ is used and we set $\Gamma = 0.1$ and $f = 1$.*

3.2.2.3 Mean-Field Thermodynamics

Dissipated Work

As already pointed out above, stochastic thermodynamics is a superstructure built upon the Markov process. As a consequence, the different dynamical phenomenology for even and

odd q is to be expected to translate into different thermodynamic features depending on q. Analogously to the coarse-graining of the thermodynamic observables demonstrated in Sec. 3.2.1.6, we can establish a thermodynamic framework of the Potts model across different scales, *i.e.* microscopic, mesoscopic and mean-field. Consequently, we find that only the statistics of the first-law quantities are invariant under the coarse-graining at finite times, while the mesoscopic (or mean field) system entropy and entropy production only represent the real entropies for a special class of initial conditions (3.66) and stationary states (3.69). Hence, as in the three-state model, we will restrict to the dissipated work as a proxy for the total dissipation (entropy production).

At the mean-field level, the first law of thermodynamics reads

$$\mathrm{d}_t \mathcal{E} = \sum_{i,j} \mathcal{E}_i \, k_{ij} \, \overline{n}_j = \dot{\mathcal{Q}} + \dot{\mathcal{W}}, \tag{3.113}$$

with the heat and work currents

$$\dot{\mathcal{Q}} = \sum_{i,j} \left(\mathcal{E}_i - \mathcal{E}_j - f \, \Theta(i,j) \right) k_{ij} \, \overline{n}_j \tag{3.114}$$

$$\dot{\mathcal{W}} = f \sum_{i,j} \Theta(i,j) \, k_{ij} \, \overline{n}_j \, . \tag{3.115}$$

The bifurcations observed in the stochastic dynamics naturally translate into nonequilibrium phase transitions that can be characterized via the work. Here, the work is dissipative since the mean-field system takes rotational energy, $\dot{\mathcal{W}} > 0$, and dissipates it into the bath in form of heat, $\dot{\mathcal{Q}} < 0$. First, we recall that except for attractive interactions below the critical temperature in Eq. (3.106) the system behaves like a noninteracting one. The stationary dissipated work current for a single unit was already determined in Eq. (3.84) and reads,

$$\langle \dot{W}_1^s \rangle = 2\Gamma f \sinh\left(\frac{\beta f}{2}\right) \geq 0, \tag{3.116}$$

and is thus independent of the number of states q. Next, in the multistability phase, $\beta \gg \beta_{c(c_2)}$, the stationary mean-field work current can be approximated as

$$\dot{\mathcal{W}}^s \approx 2\Gamma \, f \, \mathrm{e}^{\frac{\beta u}{2}} \sinh\left(\frac{\beta f}{2}\right) = \mathrm{e}^{\frac{\beta u}{2}} \, \langle \dot{W}_1^s \rangle. \tag{3.117}$$

Hence operating an interacting system in the low-temperatures regime is exponentially in the interaction strength less costly than maintaining a noninteracting system. This can be seen in Fig. 3.19 that depicts the difference between the stationary work current of a single unit $\langle \dot{W}_1^s \rangle$ and the asymptotic mean-field work current $\overline{\mathcal{W}} \equiv \mathcal{W}/t$ with $t \gg \tau_r$ for different β and q. In agreement with Eqs. (3.116) and (3.117), we find that for all q the mean-field system is noninteracting at inverse temperatures below the inverse Hopf-bifurcation temperature $\beta_{c(c_1)}$ [Eq. (3.106)], while the dissipated work is significantly reduced above that inverse critical temperature. In fact, we conclude from the monotonotic behavior of the curves

that it is always energetically beneficial to consider attractive interactions. Since the (first) critical point is shifting to larger values of β as q increases, it is overall favorable to employ small-q units. At the (first) critical point $\beta_{c(c_1)}$ the mean-field dissipated work exhibits for all q a kink that is reminiscent of a first-order equilibrium phase transition, cf. Fig. 2.1 and the corresponding introduction in Sec. 2.1. As a consequence of the two bifurcations for odd q there is a second non-equilibrium phase transition at $\beta_{c_2}(f)$ which displays characteristics of both a saddle and a jump that is more pronounced for larger q.

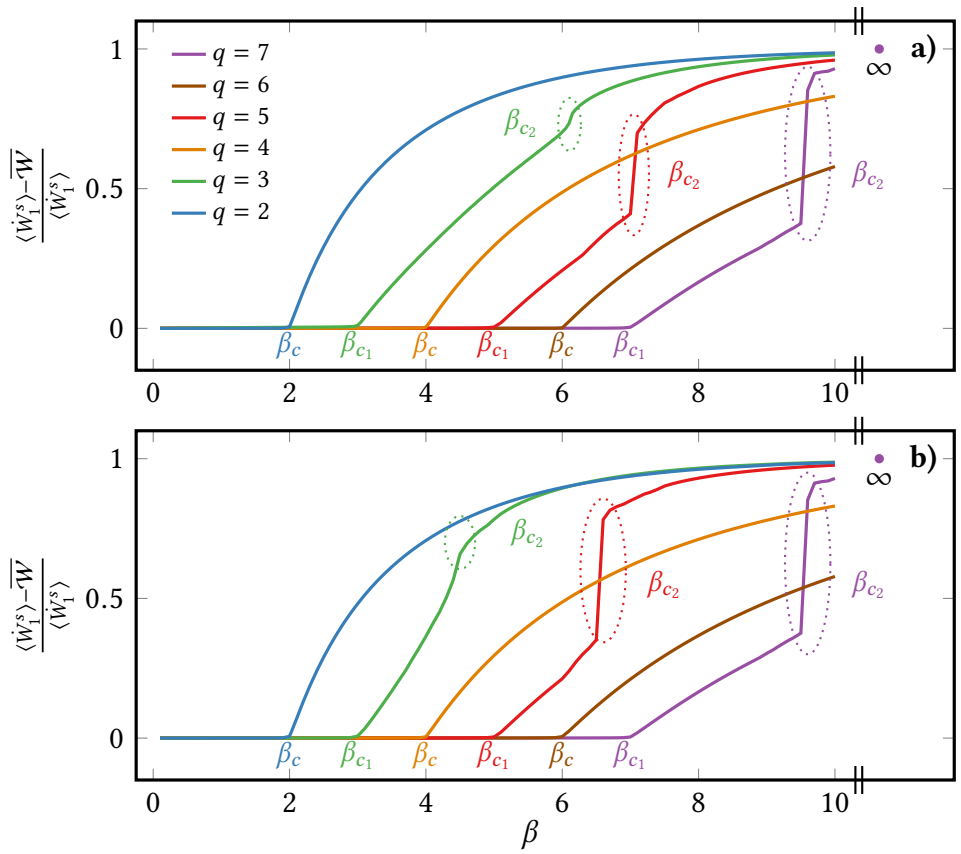

Figure 3.19: *Difference of the dissipated work for a stationary single-unit, $\langle \dot{W}_1^s \rangle$, and the asymptotic, time-averaged mean-field work $\overline{\mathcal{W}}$ as a function of the inverse temperature $\beta = 0, \ldots, 10$ for different number of states $q = 2, \ldots, 7$ and for different forces $f = 1$ in a) and $f = 0.5$ in b). The time $t = 500$ is chosen to be sufficiently long in order to ensure that $\overline{\mathcal{W}}$ has converged to its asymptotic value. The purple closed circle represents the q-independent analytic result for the work currents in the low-temperature limit* (3.117).

It is interesting to observe that for smaller driving $f = 0.5$ in panel b) the second critical phase transition characterized by β_{c_2} is shifting to smaller values of β. Furthermore, the sensitivity of the second critical point on the driving is diminishing as q becomes large.

This means that the range of intermediate temperatures for which stable limit cycles can be observed is only dependent on how far the system is driven out-of-equilibrium for small q-Potts models. Hence, for smaller q there is an additional reduction in the dissipated work for attractive interactions when the system is driven less far out-of-equilibrium.

3.2.2.4 Efficiency at Maximum Power

Finally, we want to investigate how the number of spin states q affects the power-efficiency trade-off in energy transduction processes. To this end, we again consider a work-to-work conversion by decomposing the nonconservative force f into a force aligned with ($f_1 > 0$) and a force acting against ($f_2 < 0$) the bias, *i.e.* $f = f_1 + f_2$ and define via Eq. (3.115) the corresponding asymptotic work contributions $\overline{\mathcal{W}}_1$ and $\overline{\mathcal{W}}_2$, respectively. Next, we maximize the asymptotic power output $\overline{\mathcal{P}}$ (3.91) with respect to f_2^* and compute the efficiency at maximum power η^* (3.92).

Fig. 3.20 shows the asymptotic power output $\overline{\mathcal{P}}$ as a function of β and f_2 for $q = 3$ in a) and $q = 5$ in b). We observe that in both odd-q systems the maximum power output is achieved in the synchronization regime indicated by the area enclosed by the dashed white lines. The same applies to the global maximum indicated by the closed purple circle.

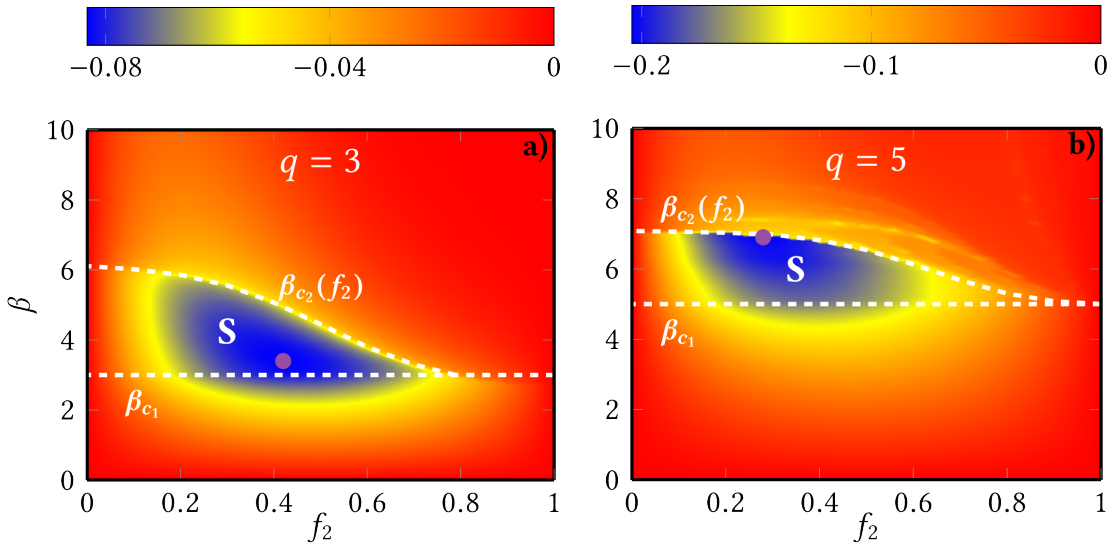

Figure 3.20: *The asymptotic, time-averaged output power $\overline{\mathcal{P}}$ as a function of the output force f_2 and the inverse temperature β for $q = 3$ in a) and $q = 5$ in b). In both plots the white dashed lines indicate the set of critical points. Hence the enclosed area defines the synchronization phase S. In addition, the global maximum of the output power is indicated by the purple closed circle.*

Next, we recall that according to Eq. (3.106) in both cases in the high-temperature limit the

systems are noninteracting and their stationary power output is thus determined by Eq. (3.116),

$$\mathcal{P}_1^s = 2\Gamma f_2 \sinh\left(\frac{\beta f}{2}\right). \tag{3.118}$$

Further, in both cases the power rapidly drops in the low-temperature regime ($\beta \gg \beta_{c_2}$) as prescribed by Eq. (3.117) that implies for the asymptotic power output,

$$\overline{\mathcal{P}} \approx 2\Gamma\, f_2\, \mathrm{e}^{\frac{\beta u}{2}} \sinh\left(\frac{\beta f}{2}\right) = \mathrm{e}^{\frac{\beta u}{2}}\, \mathcal{P}_1^s. \tag{3.119}$$

Hence operating an interacting system in the low-temperatures regime provides a power output that is exponentially higher in the interaction strength compared to a noninteracting system. The range of validity of these approximations can also be seen in Figs. 3.21a) and b) that show the maximum power for different β.

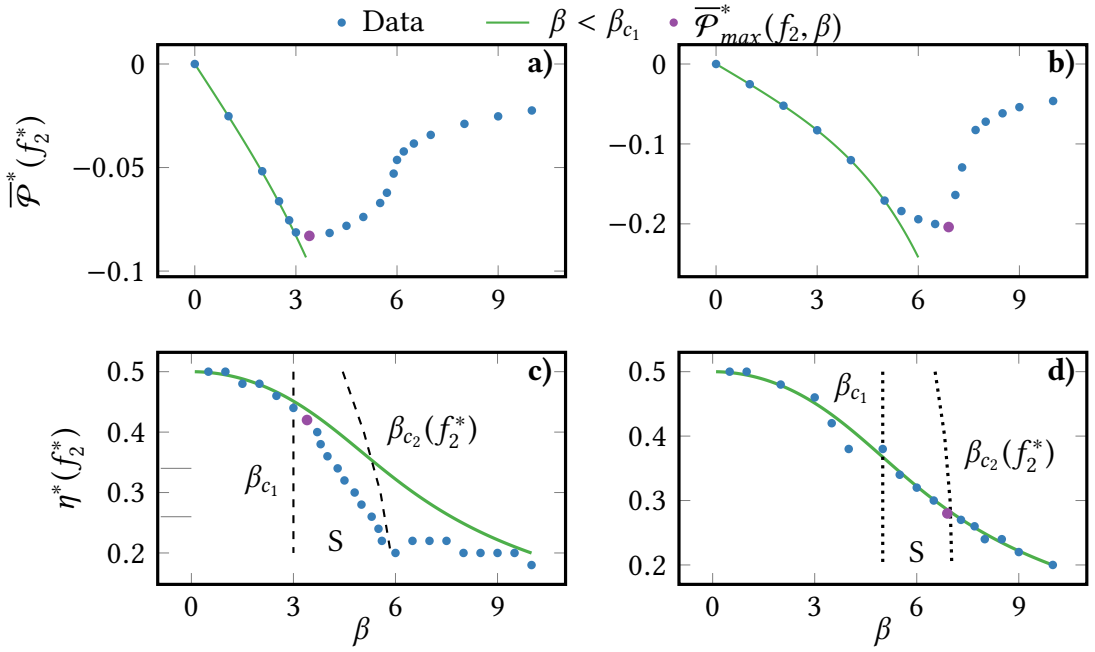

Figure 3.21: *The maximum output power $\overline{\mathcal{P}}^*$ optimized with respect to f_2 for $q = 3$ in a) and $q = 5$ in b). Moreover, panels c) and d) depict the associated efficiency at maximum power $\eta^*(f_2^*)$ for $q = 3$ and $q = 5$, respectively. Here, the dashed lines specify the critical points and the synchronization phase S and the efficiency at the global maximum power is indicated by a purple closed circles. The analytic solution from Eq. (3.118) for $\beta < \beta_{c_1}$ is overlaid with the numerical data in all panels.*

Here, the analytic low-β solution from Eq. (3.118) is overlaid and shows a good agreement up to the proximity of the global maximum power indicated by the closed purple circle. This

reflects that the global maximum power is attained far-out-of equilibrium where the system responds nonlinearly. Expanding Eq. (3.118) up to first order in β, yields the linear-response relation $J^s = Lf$ with the Onsager coefficient $L = 2\Gamma\beta$. Hence small products βf represent the linear-response regime. Thus, the result obtained in Sec. 3.2.1.7 that maximum power is achieved in the far out-of-equilibrium synchronization regime is not a mere coincidence for $q = 3$ but seems to be a universal one.

We conclude from comparing Figs. 3.21 a) and b) that, on absolute terms, the maximum power is increasing with q. Yet, turning to the associated efficiency at maximum power shown in panels c) and d), we see that as q increases, the associated efficiencies of these far out-of-equilibrium processes are deviating more and more from the maximum value $\eta_c/2 = 1/2$ that is universal for a system with a single current responding linearly (2.56), *i.e.* for $\beta f \to 0$. This shows that while the power-efficiency trade-off is optimal for synchronous processes, it can not be further lifted by simply changing the topology of the synchronized units.

Conversely, Fig. 3.22 depicts the power output in a density plot as function of β and f_2 for different even values of q. First we note again that for the even number of spin states there is no synchronization phase. Instead, the systems exhibit a subcritical Hopf bifurcation at β_c that separates the single fixed-point phase, *i.e.* a noninteracting system as in Eq. (3.118), and the multistability phase where all units coalesce into one of the q states. As a consequence, the power output immediately drops for $\beta > \beta_c$ according to Eq. (3.119) and there is no operating mode where - compared to a noninteracting system - significant additional power can be generated. It is thus detrimental for the energy transduction to employ a Potts model with an even number of states due to its absence of collective dynamics.

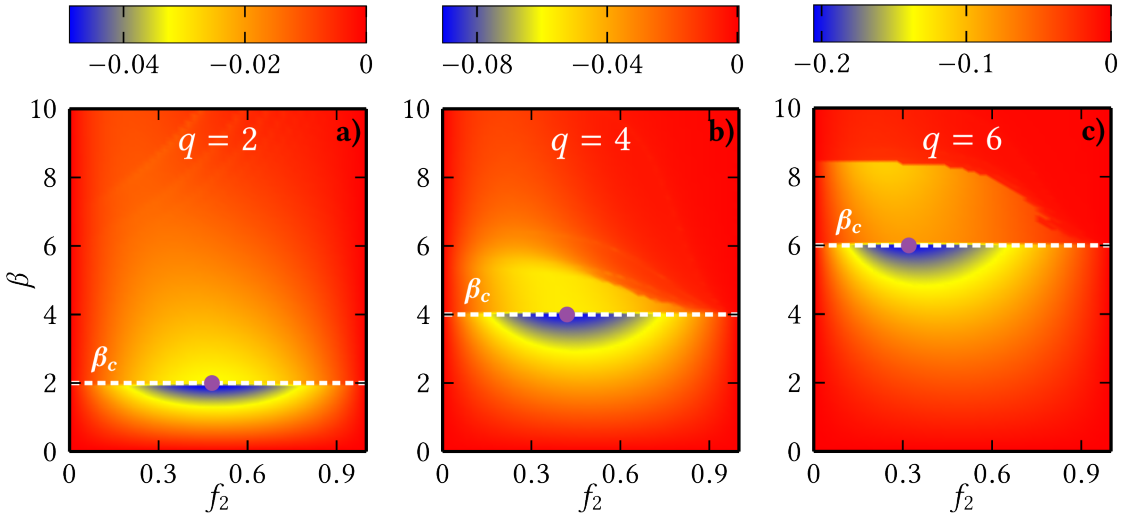

Figure 3.22: *The asymptotic output power $\overline{\mathcal{P}}$ as a function of the output force f_2 and the inverse temperature β for $q = 2$ in a), $q = 4$ in b) and $q = 6$ in c).*

To gain deeper insight of the dependence of the power-efficiency trade-off on the number

of states q, Fig. 3.23 depicts both the global maximum power - indicated by purple closed circles in Figs. 3.20 and 3.22 - and the associated efficiency at maximum power as a function of q.

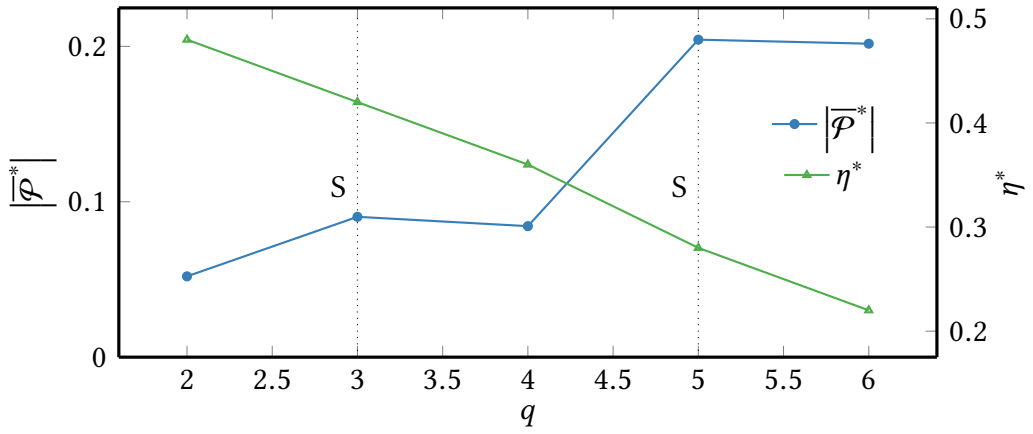

Figure 3.23: *The global maximum power output and the associated efficiency at maximum power for different q. The black dotted lines indicate the q-values that exhibit a synchronization (S) phase.*

Overall the maximum power output is increasing with q, though there are striking jumps from even to odd q-systems, *i.e.* from asynchronous to synchronized systems. These jumps are followed by plateaus where the maxima remain roughly the same. Conversely, the associated efficiency at maximum powers are decreasing monotonically with q. In particular, for $q \leq 3$, the efficiency at maximum power is close to the optimal linear-response value $\eta^* = 1/2$. Therefore, we confirm the conclusion made in Sec. 3.2.1.7 that the optimal power-efficiency trade-off is achieved for synchronization, *i.e.* for odd-q systems. It is furthermore energetically beneficial to employ small-q units as they minimize the dissipation and therefore enhance the efficiency of the energy conversion processes.

3.3 Thermodynamics of Majority-Logic Decoding in Information Erasure

3.3.1 Majority-Logic Decoding

As an elementary storage unit, we consider a microscopic binary unit with states 0 and 1 and denote by p the probability to observe it in state 1. Based on these microscopic binary units we can construct mesoscopic (logical) bits in two different ways as illustrated in Fig. 3.24.

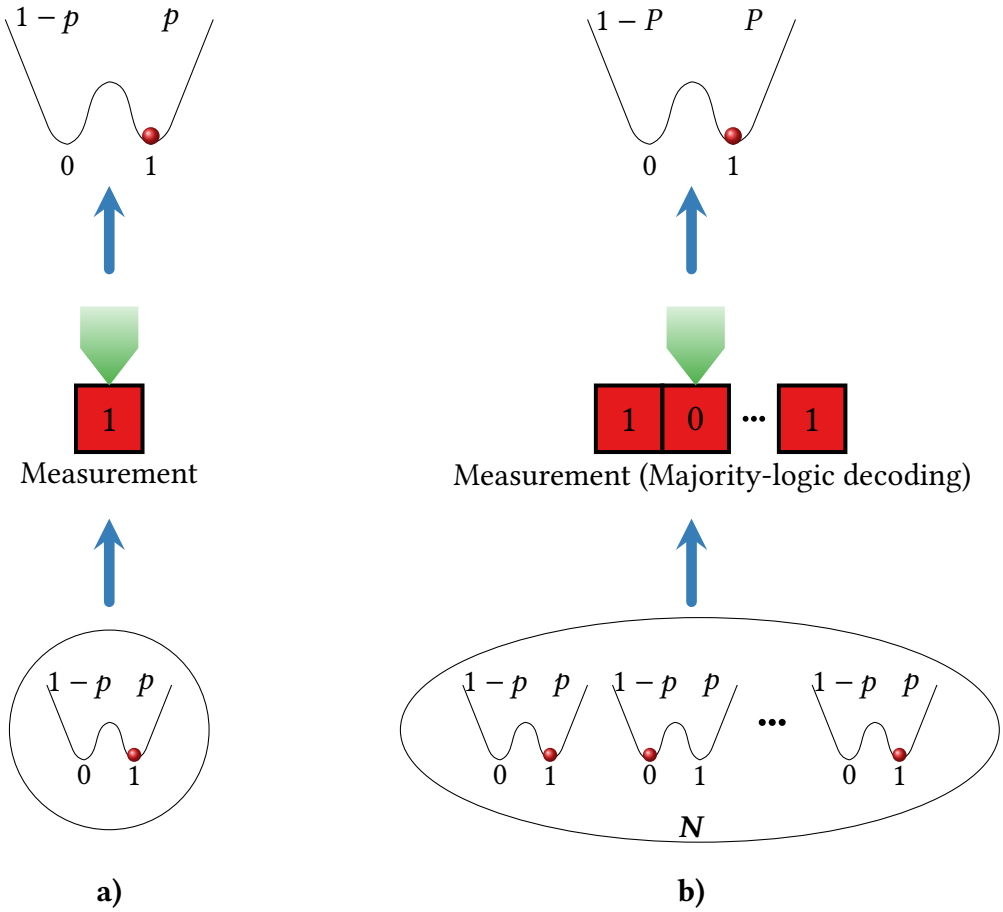

Figure 3.24: *Schematics of a single-unit bit (SUB) in a) and a N-majority-logic decoding bit (MLB) in b).*

First, we consider a single-unit bit (SUB) in Fig. 3.24a) that consists of only one microscopic unit and is in contact with a heat bath at inverse temperature β. The probability of finding the

SUB in state 1 is represented by P, and is, of course, equal to probability p for the microscopic unit to be in state 1. Alternatively, a majority-logic decoding bit (MLB) can be thought of as an array of N identical and non-interacting microscopic units that are subjected to the same experimental protocol and connected to a heat bath at inverse temperature β as sketched in Fig. 3.24b). The probability of the MLB to be in state 1 is denoted by P and is determined via majority-logic decoding. This decoding scheme prescribes that the information encoded in the MLB corresponds to the state that is occupied most at the level of the microscopic units and is therefore a coarse-graining procedure that is mathematically formulated in the following.

The relation between the microscopic probability p associated with each microscopic unit and the macroscopic probability P specifying the occupation probability of the MLB can be expressed as

$$P(p,N,\zeta) = \sum_{k=\zeta}^{N} \binom{N}{k} p^k (1-p)^{N-k}, \tag{3.120}$$

where ζ represents the threshold number of the detector of the same microscopic state during the measurement process.

There is an useful identity that relates the regularized incomplete beta function

$$\mathcal{I}_x(a,b) = \frac{\int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t}{\int_0^1 t^{a-1}(1-t)^{b-1}\mathrm{d}t}, \tag{3.121}$$

to the macroscopic probability

$$P(p,N,\zeta) = \mathcal{I}_p(\zeta, N+1-\zeta). \tag{3.122}$$

This equality can be seen as follows. We start from the definition of the regularized incomplete beta function in (3.121) and solve the integral in the denominator via partial integration yielding

$$\int_0^1 t^{a-1}(1-t)^{b-1}\mathrm{d}t = \frac{(a-1)!(b-1)!}{(a+b-1)!}. \tag{3.123}$$

The integral in the numerator is determined via consecutive application of partial integration

$$\int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t = \frac{1}{a}x^a(1-x)^{b-1} + \frac{b-1}{a(a+1)}x^{a+1}(1-x)^{b-2} + \ldots + \frac{(b-1)(b-2)\cdot\ldots\cdot 2\cdot 1}{a(a+1)\cdot\ldots\cdot(a+b-1)}x^{a+b-1}. \tag{3.124}$$

Collecting results, one finds that

$$\mathcal{I}_x(a,b) = \sum_{k=a}^{a+b-1} \binom{a+b-1}{a} x^a(1-x)^{b-1}\,. \tag{3.125}$$

If we set $a = \zeta, b = N - \zeta + 1, x = p$, we immediately arrive at Eq. (3.122).

The bijectivity of the regularized incomplete beta function $\mathcal{I}$ allows furthermore to determine the microscopic probability p given the macroscopic probability P of a MLB as follows

$$p(P, N, \zeta) = \mathcal{I}_P^{-1}(\zeta, N + 1 - \zeta). \tag{3.126}$$

The qualitative features of the majority-logic decoding are illustrated in Fig. 3.25 that compares the macroscopic probability $P(p, N, \zeta)$ of a SUB and MLB for different array sizes N and threshold values ζ.

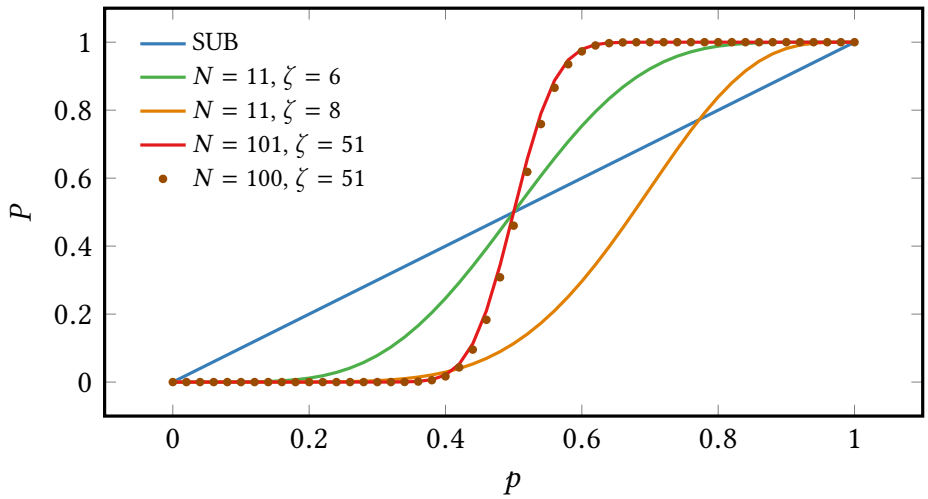

Figure 3.25: *The curves of P with respect to p at different values of N and ζ. The blue line refers to the SUB. While the green ($N = 11, \zeta = 6$) and red curves ($N = 101, \zeta = 51$) refer to symmetric MLBs, the orange ($N = 11, \zeta = 8$) and the brown dotted curves ($N = 100, \zeta = 51$) refer to asymmetric MLBs with even N.*

First, we observe that the macroscopic probability P is monotonically increasing with p. Next, for any value of N and ζ it holds for $x = 0, 1$ that $P(x, N, \zeta) = x$ since the incomplete beta function becomes a unity operator, $\mathcal{I} = \boldsymbol{I}$. Physically, this means that perfect information erasure in a MLB ($P = 0, 1$) is realized by perfect erasure in each microscopic unit the MLB consists of. For odd values of N the symmetric case, $\zeta = (N + 1)/2$, corresponds to *majority*-logic decoding in the strict sense, that is

$$P\left(p, N, \tfrac{N+1}{2}\right) = \mathcal{I}_p\left(\tfrac{N+1}{2}, \tfrac{N+1}{2}\right). \tag{3.127}$$

We note that for symmetric majority-logic decoding, $\zeta = (N + 1)/2$, the curve is symmetric with respect to $(1/2, 1/2)$ $\forall N$. This is readily derived by noting that according to the binomial theorem, one has

$$P\left(\tfrac{1}{2}, N, 0\right) = \frac{1}{2^N} \sum_{k=0}^{N} \binom{N}{k} = \frac{1}{2^N}(1 + 1)^N = 1. \tag{3.128}$$

Thus, it immediately follows from the symmetry of the binomial probability distribution [Eq. (3.120)], that one has $P\big(1/2, N, \zeta = (N+1)/2\big) = 1/2$. It can furthermore be seen in Fig. 3.25 that for large N the curves converge to a step function centered at $p = 1/2$. For large N, Taylor-expanding Eq. (3.120) for the symmetric case around $p = 1/2$ up to linear order in $|p - 1/2|$ yields

$$\left|\frac{1}{2} - P\right| = \sqrt{\frac{2N}{\pi}}\left|\frac{1}{2} - p\right| + O\left(\left|\frac{1}{2} - p\right|^3\right), \tag{3.129}$$

where we denote by $O(x^2)$ all terms that are at the order x^2 or higher. According to Eq. (3.129), the slope at the symmetric point p is increasing with $\sqrt{N}$ in the symmetric decoding case. We therefore arrive at the first important result that large macroscopic erasure in the MLB can be achieved at the cost of small microscopic erasure in each microscopic unit. The good agreement of the curves corresponding to $N = 100$ and $N = 101$ suggests that for $\zeta \approx (N+1)/2$ the symmetry decoding case is approached if N becomes large.

3.3.2 Information Thermodynamics

3.3.2.1 Definitions

The equivalent probabilistic descriptions on the microscopic and macroscopic level prompt the question of how to define the underlying physical processes at these levels. In information erasure processes, important quantities are the heat that is generated during the operation, the change in Shannon entropy that measures the amount of information erased by that process as well as the erasure efficiency [96, 157].

On the level of a single microscopic unit, the heat dissipated during an erasure process from p_i to p_f is denoted by q and defined as a negative quantity $q < 0$. With this convention, the first law of thermodynamics in differential form reads

$$\mathrm{d}_t\, e = \dot{q} + \dot{w}, \tag{3.130}$$

where $\dot{w}$ denotes the work current and $\dot{e}$ the rate of energy change. The microscopic Shannon entropy

$$s(p) = -p \ln p - (1-p)\ln(1-p), \tag{3.131}$$

allows to quantify $\Delta s(p_i, p_f) \equiv s(p_f) - s(p_i)$ as the amount of information that is erased during an erasure process from p_i to p_f. The second law of thermodynamics reads

$$\mathrm{d}_t s = \beta\dot{q} + \dot{\sigma}, \tag{3.132}$$

where $\dot{\sigma} \geq 0$ refers to the irreversible entropy production rate in the microscopic unit.

Throughout this thesis, we consider the case in which the initial state is the maximum information state given by $p_i = 1/2$ and $s(1/2) = \ln 2$. For this case the change in entropy is always negative $\Delta s(p_i, p_f) < 0$ if $p_f \neq 1/2$. Therefore, a suitable definition of the microscopic erasure efficiency for this process reads

$$\eta_s(p_i, p_f) = \frac{\Delta s(p_i, p_f)}{\beta\, q}. \tag{3.133}$$

The heat q generated by the microscopic unit is naturally dependent on specific models and operating protocols. The optimal protocols that minimize the dissipated heat and thus maximize the erasure efficiency are investigated in Sec. 3.3.2.3.

On the level of macroscopic bits, the heat dissipation refers to the cumulated heat generated by all the microscopic units the bit consists of. Thus, the microscopic definitions from above are, of course, also physically significant for the erasure process at the level of the SUB, that is $s(P) = s(p)$ and $\eta_s(1/2, p_f) = \eta_s(1/2, P_f)$. Conversely, in order to perform a macroscopic erasure in the MLB from $P_i = 1/2$ to P_f, an amount of information specified by the majority-logic decoding needs to be erased in each microscopic unit contained in the MLB. Here, P_f is the probability that the final state of the MLB after the erasure is logically decoded as state 1. According to Eq. (3.126), this amounts to change in each microscopic unit from the initial state $p_i = \mathcal{I}_{1/2}^{-1}(\zeta, N+1-\zeta)$ to the final state $p_f = \mathcal{I}_{P_f}^{-1}(\zeta, N+1-\zeta)$. The heat dissipated by the MLB, Q, is thus determined as follows

$$Q(1/2 \to P_f) = N\, q \left(\mathcal{I}_{1/2}^{-1}(\zeta, N+1-\zeta) \to \mathcal{I}_{P_f}^{-1}(\zeta, N+1-\zeta) \right). \tag{3.134}$$

The first law of thermodynamics at the level of the MLB reads

$$\mathrm{d}_t E = \dot{Q} + \dot{W}, \tag{3.135}$$

where the thermodynamic quantities of the MLB are naturally given by the sum of the microscopic ones, *i.e.* one has $E = Ne, Q = Nq, W = Nw$. It furthermore holds that $\Sigma = N\sigma$ since the irreversible entropy production of the MLB has to be equal to the sum of the irreversible entropy production of the microscopic units contained in the MLB. Next, we define the entropy quantifying the information stored in the MLB as for the microscopic unit, *i.e.*

$$S(P) = -P \ln P - (1-P) \ln (1-P). \tag{3.136}$$

From the definition for the entropy associated with a MLB made in Eq. (3.136) follows that $S \neq Ns$. This however implies that the entropy balance at the level of the MLB is broken

$$\mathrm{d}_t S \neq \beta\, \dot{Q} + \dot{\Sigma}. \tag{3.137}$$

Hence the macroscopic Shannon entropy in Eq. (3.136) should be thought of as logical but not strictly physical information. Finally, the macroscopic efficiency associated with the erasure in the MLB is defined as

$$\eta_m(1/2, P_f) = \frac{\Delta S(1/2, P_f)}{\beta\, Q}, \tag{3.138}$$

with the change in macroscopic Shannon entropy $\Delta S(1/2, P_f) \equiv S(P_f) - S(1/2) = S(P_f) - \ln 2$.

3.3.2.2 Reversible Erasure Protocols

In the limit of reversible erasure, the irreversible entropy production vanishes and one has for the heat in a SUB $\beta q = s(p_f) - \ln 2$. Thus, from Eq. (3.133) follows that $\eta_s^{rev} = 1$ for any erasure process in this limit. Turning to the MLB, the dissipated heat during the erasure process reads

$$\beta Q = N\left[s\left(\mathcal{I}_{P_f}^{-1}(\zeta, N+1-\zeta)\right) - s\left(\mathcal{I}_{1/2}^{-1}(\zeta, N+1-\zeta)\right)\right]. \tag{3.139}$$

Fig. 3.26 depicts in the reversible limit the heat dissipated during an erasure process from $P_i = 1/2$ to P_f by a SUB and MLB [panel a)] and the associated efficiencies for symmetric decoding processes [panel b)].

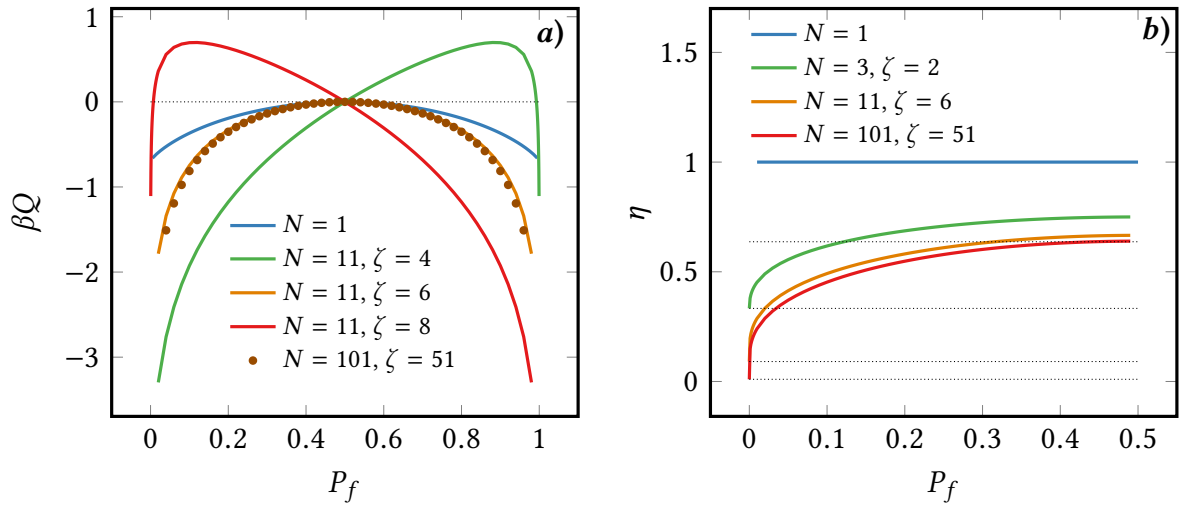

Figure 3.26: *a) Heat dissipated by the macroscopic bits in the reversible limit. The initial macroscopic state is $P_i = 1/2$. b) Efficiencies associated with reversible erasure processes for the SUB and symmetric MLBs. The dotted lines from top to bottom represent the values $\eta = 2/\pi$, $\eta = 1/3$, $\eta = 1/11$ and $\eta = 1/101$, respectively.*

We observe in panel a) that the SUB always dissipates less heat than the symmetric MLBs which generate more heat as N increases. For asymmetric majority-logic decoding, $\zeta \neq (N+1)/2$, the heat generated by a MLB is reduced with respect to the one of a SUB and even takes positive values for probabilities in the range $\left[1/2, \mathcal{I}_{1-\mathcal{I}_{P_f}^{-1}(\zeta,N+1-\zeta)}(\zeta, N+1-\zeta)\right]$ or $\left[\mathcal{I}_{1-\mathcal{I}_{P_f}^{-1}(\zeta,N+1-\zeta)}(\zeta, N+1-\zeta), 1/2\right]$, if the decoding is left-asymmetric or right-asymmetric, respectively. The positivity of the heat means that one could even extract work during certain erasure processes by employing asymmetric majority-logic decoding. This property is similar to the division of logical entropy of the system entropy and the physical entropy of each subspaces as discussed in Refs. [127, 158, 159]. However, here, the analogue of physical

entropy in each subspaces varies according to Eqs. (3.120) and (3.131) as the macroscopic probability P changes.

The erasure efficiency of MLBs under symmetric majority-logic decoding, $\zeta = (N+1)/2$, for different array lengths N is shown in Fig. 3.26b). Consistently to the observations made in panel b), the SUB is always more efficient than the MLB whose erasure efficiency decreases as N increases. The same applies to the other erasure branch given by $P_f > 1/2$ that can be seen as follows: According to Eqs. (3.127) and (3.131), the macroscopic Shannon entropy of a symmetric MLB is symmetric with respect to $P = 1/2$. Therefore, the energetics of the erasure process from the initial state $P_i = 1/2$ to the final state P_f is the same as that of the erasure process from $P_i = 1/2$ to the final state $1 - P_f$. Hence, we will restrict to the symmetric majority-logic decoding characterized by Eq. (3.127) and to the erasure branch, $P \leq 1/2$, in the following. To ease notation, we omit the explicit notation $\zeta = (N+1)/2$ and $P_i = p_i = 1/2$.

As already pointed out earlier, the macroscopic entropy associated with a MLB can be thought of as a coarse-graining of the sum of the physical entropy of each microscopic unit contained in the MLB. We write the difference between the physical entropy of all microscopic units contained in the bit and the logical (Shannon-like) entropy of the macroscopic bit as

$$\mathbb{S}(P,N) = Ns\left(\mathcal{I}_P^{-1}\left(\tfrac{N+1}{2}, \tfrac{N+1}{2}\right)\right) - S(P). \tag{3.140}$$

In Appendix A.4, a proof for the monotonic behavior of $\mathbb{S}(P,N)$ with respect to P in the case of symmetric majority-logic decoding is provided. In the range of $0 \leq P \leq 1/2$, the entropy function $\mathbb{S}(P,N)$ is monotonically increasing, else it is decreasing monotonically. Evidently, $\mathbb{S}(P,N)$ takes its maximum value, $(N-1)\ln 2$ if $p = P = 1/2$ and its minimum value, $\mathbb{S}(P,N) = 0$, if $p = P = 0, 1$. Hence, for symmetric majority-logic decoding the logical information underestimates or is equal to the physical one $\mathbb{S}(P,N) \geq 0\ \forall N$. As can be seen in Fig. 3.26a), such an inequality does not hold for the more general asymmetric case where $\zeta \neq (N+1)/2$. In fact, the inequality $\mathbb{S}(P,N) \geq 0$ is physically equivalent to the well-known universal relationship between fully microscopic and coarse-grained Shannon entropies [112]. It is furthermore important to note that no general statement can be made about the changes of entropy for a given erasure process, since, according to Eq. (3.127) the final distribution for the microscopic unit is, in general, different from the one of the MLB.

From Eq. (3.138) and the monotonicity of $\mathbb{S}(P,N)$ follows that the reversible erasure efficiency of the symmetric MLB is bounded as follows

$$0 < \eta_m^{rev}(1/2, P_f) \leq 1, \tag{3.141}$$

where the equality only holds when no information is erased, $\Delta S(1/2, P_f = 1/2) = 0$. Hence we find that a SUB is always more efficient in reversible information erasure than a symmetric MLB. This can be attributed to the majority-logic decoding that neglects some microscopic degrees of freedom and thus associates less logical entropy to the MLB than the cumulated physical entropy of the microscopic units that constitute the MLB.

We now state two limiting results for the reversible erasure efficiency of a MLB. First, in case of perfect erasure, $P_f = 0$, we derive from Eqs. (3.127) and (3.138) that the erasure efficiency simplifies to

$$\eta_m^{rev}(1/2, 0) = \frac{1}{N}. \tag{3.142}$$

Secondly, if the amount of erased information is small, $P_f \approx P_i = 1/2$, and N is large, the erasure efficiency

$$\eta_m^{rev}(1/2, P_f) = \frac{2}{\pi} + O\left(\left|\frac{1}{2} - P_f\right|^2\right), \tag{3.143}$$

becomes independent of the final macroscopic probability $P_f \approx 1/2$ as can be seen in Fig. 3.26b).

3.3.2.3 Finite-Time Erasure Protocol under Majority-Logic Decoding

Since we have captured the phenomenology of the reversible majority-logic decoding in the previous section, we now proceed by studying more realistic, finite-time information erasure processes. In dynamical processes the heat generation depends on the erasure protocol and the specific model for the microscopic units. In this section we formulate finite-time erasure processes for two commonly employed microscopic models: We consider a two-state system with either Arrhenius rates or Fermi rates and denote in both cases by $p(t)$ the probability of the unit to be in state 1 at time t.

Preliminaries

Master Equation The transition rate from state 0 to state 1 and vice versa is referred to as $w_{10}(t)$ and $w_{01}(t)$, respectively, which depends, in general, on time via the erasure protocol. We assume that the process is Markovian, such that the dynamics of $p(t)$ is ruled by a master equation:

$$\partial_t p(t) = [1 - p(t)]\, w_{10}(t) - p(t)\, w_{01}(t), \tag{3.144}$$

with the transition rates satisfying the local detailed balance relation

$$\frac{w_{10}(t)}{w_{01}(t)} = \mathrm{e}^{-\beta\, \Delta\epsilon(t)}, \tag{3.145}$$

where $\Delta\epsilon(t)$ is the energy gap from state 0 to 1 that is modulated in time according to the specific erasure protocol. With Eq. (3.145) the master equation (3.144) can be cast into the form

$$\partial_t p(t) = \left[\mathrm{e}^{-\beta\, \Delta\epsilon(t)} - (1 + \mathrm{e}^{-\beta\, \Delta\epsilon(t)})p(t)\right] w_{01}(t). \tag{3.146}$$

Finite-Time Erasing No protocol can achieve a perfect erasure corresponding to $s(p_f) = s(P_f) = 0$, since this requires an infinite amount of time [96]. This can be seen as follows: For a given microscopic model, there are several ways to decrease the microscopic probability $p(t)$ at different speeds, which corresponds to a protocol that ensues different amounts of heat and thus a different erasure efficiency. This, in turn, implies that the time required to transform the initial probability $p_i(0) = 1/2$ into the final one $p_f(\tau) \equiv p_f$ can not be smaller than a minimal time τ_c

$$\tau \geq \tau_c = -\log 2p_f, \quad p_f < \frac{1}{2}, \tag{3.147}$$

As $p_f \to 0$, the minimal time diverges, $\tau_c \to \infty$, such that perfect erasure of a finite initial amount of information can only be realized by infinite-time protocols.

Eq. (3.147) can be rearranged to obtain a lower bound for the final probability obtained after an erasure with fixed duration τ as follows

$$p_f \geq p_c = \frac{1}{2}\mathrm{e}^{-\tau_c}, \quad p_f < \frac{1}{2}, \tag{3.148}$$

Consequently, in finite-time information erasure processes, the Shannon entropy of the final state $s(p_f)$ can be seen as the erasure error. According to Eq. (3.148), the minimal erasure error of the process is $s(p_c)$ and the erasable information within τ_c is $\Delta s(p_f, 1/2) = \log 2 - s(p_c)$. According to Eqs. (3.147) and (3.148), the bounds τ_c and p_c are rate- and thus model-independent.

These results hold at the level of a microscopic unit. We now proceed by discussing the bounds on the level of macroscopic bits. First, since each SUB consists of only one microscopic unit, Eqs. (3.147) and (3.148) are also applicable to the macroscopic quantities, that is

$$\tau \geq \tau_c^s = \tau_c = -\log 2P_f, \quad P_f \geq P_c^s = p_c = \frac{1}{2}\mathrm{e}^{-\tau}, \quad P_f < \frac{1}{2}. \tag{3.149}$$

In order to calculate the lower bound on the macroscopic erasure time $\tau_c^m(1/2, P_f, N)$ for a symmetric MLB, we recall that a macroscopic erasure from $P_i = 1/2$ to P_f is achieved by the corresponding erasure from $p_i = 1/2$ to p_f in each microscopic unit contained in the MLB. Therefore, one has for the minimal erasure time in a MLB

$$\tau_c^m(1/2, P_f, N) = \tau_c(1/2, p_f), \tag{3.150}$$

where p_f and P_f are related to each other via the symmetric majority-logic decoding in Eq. (3.127). Hence, in order to compute τ_c^m, the final microscopic probability p_f needs to be determined via inversion of Eq. (3.127) and plugged into Eq. (3.147). Next, using Eqs (3.148) and (3.127), one straightforwardly obtains the minimal final probability $P_c^m(\tau, N)$ after erasure time τ. From Eq. (3.127) follows the inequalities

$$\tau_c^m(P_f, N) \leq \tau_c^s(P_f), \quad P_c^m(\tau, N) \leq P_c^s(\tau), \tag{3.151}$$

which suggest that majority-logic decoding can both accelerate the erasing process and additionally reduce the minimal erasure error. Finally, for finite-time erasure processes we define the erasure power for a SUB and MLB as follows

$$\mathcal{P}_s = \frac{\Delta s(1/2, p_f)}{\tau}, \quad \mathcal{P}_m = \frac{\Delta S(1/2, P_f)}{\tau}, \tag{3.152}$$

respectively.

Arrhenius-Rates Unit Model The Arrhenius model consists of two potential well, which are regarded as states 0 and 1, respectively, that are separated by a barrier. The energy gap between state 0 and 1 is defined as $\Delta\epsilon$. In order to be consistent with the assumption of Arrhenius transition rates, the energy barrier height of the potential well associated with state 1 is assumed constant, ϵ, throughout the erasure process. This setup is different from the the one used in Ref. [160], where the two states are merged together and then separated during the erasure. Here, the transition rates read

$$w_{01}^A(t) = r_0\, \mathrm{e}^{-\beta\epsilon}, \quad \text{and} \quad w_{10}^A(t) = r_0\, \mathrm{e}^{-\beta[\epsilon+\Delta\epsilon(t)]}, \tag{3.153}$$

where r_0 is a constant setting the time scale of the process. Then, the master equation (3.146) can be written as follows

$$\partial_t p(t) = \mathrm{e}^{-\beta\Delta\epsilon(t)} - \left[1 + \mathrm{e}^{-\beta\Delta\epsilon(t)}\right] p(t), \tag{3.154}$$

where the constant $r_0\, \mathrm{e}^{-\beta\epsilon}$ is absorbed into the time scale.

Fermi-Rates Unit Model The Fermi-rates unit model can be experimentally realized via a single quantum dot with a single energy level E that is in contact with a moving metallic lead corresponding to a time-dependent chemical potential $\mu(t)$ and a heat bath at inverse temperature β [157]. If the dot is filled by an electron we consider the unit to be in state 1, else 0. The transition rate for an electron leaving or entering the dot reads

$$w_{01}^F(t) = \frac{r_0}{\mathrm{e}^{-\beta\,\Delta\epsilon(t)} + 1}, \quad w_{10}^F(t) = \frac{r_0}{\mathrm{e}^{\beta\,\Delta\epsilon(t)} + 1}, \tag{3.155}$$

respectively, where $\Delta\epsilon(t) \equiv E - \mu(t)$ represents the energy barrier to enter the dot and r_0 is a constant that sets the time scale of the process. Then, the master equation (3.146) can be written as follows

$$\partial_t p(t) = -p(t) + \frac{1}{\mathrm{e}^{\beta\,\Delta\epsilon(t)} + 1}, \tag{3.156}$$

where the constant r_0 is absorbed into the time scale. Since the second term in Eq. (3.156) is bounded between 0 and 1, the fastest way to decrease the microscopic probability $p(t)$ is realized by $\partial_t p(t) = -p(t)$, which corresponds to a protocol that ensues a divergent heat generation and thus a vanishing erasure efficiency.

Variable Erasure Duration

With the tools to address finite-time information processing at hand, we want to start with the simplest erasure protocol given by an instantaneous switching of the energy gap to the same value $\Delta\epsilon$ at time $t = 0$. For generic transition rates with constant energy gaps, $w_{01}(\Delta\epsilon)$, the master equation of a microscopic unit from Eq. (3.146) is solved by

$$p(t) = \frac{1}{2}\tanh\left(\frac{\beta\Delta\epsilon}{2}\right)\mathrm{e}^{-\left(1+\mathrm{e}^{-\beta\Delta\epsilon}\right)w_{01}(\Delta\epsilon)\,t} + \frac{1}{1+\mathrm{e}^{\beta\Delta\epsilon}}, \tag{3.157}$$

where we used the initial condition $p(0) = 1/2$. In the infinite-time limit the probability converges to the lower bound

$$p(\infty) = \frac{1}{1+\mathrm{e}^{\beta\Delta\epsilon}}. \tag{3.158}$$

We emphasize that for fixed $\Delta\epsilon$ the definitions (3.147) and (3.148) no longer apply. However, Eq. (3.158) represents also a bound on the minimal erasure error and thus plays a similar role as p_c. From Eq. (3.157) we obtain the following expression for the erasure duration

$$\tau = -\frac{1}{\left(1+\mathrm{e}^{-\beta\Delta\epsilon}\right)w_{01}(\Delta\epsilon)}\ln\left(2\frac{(1+\mathrm{e}^{\beta\Delta\epsilon})p_f - 1}{\mathrm{e}^{\beta\Delta\epsilon}-1}\right). \tag{3.159}$$

The heat dissipated by the microscopic unit reads

$$q = \int_0^\tau \Delta\epsilon\,\partial_t p(t)\,\mathrm{d}t = \Delta\epsilon\int_{p_i}^{p_f}\mathrm{d}p = \left(p_f - \frac{1}{2}\right)\Delta\epsilon, \tag{3.160}$$

which, with Eq. (3.134), results in the total heat generated by the SUB and MLB

$$q = \left(p_f - \frac{1}{2}\right)\Delta\epsilon, \quad Q = N\left[\mathcal{I}_{P_f}^{-1}\left(\frac{N+1}{2},\frac{N+1}{2}\right) - \frac{1}{2}\right]\Delta\epsilon, \tag{3.161}$$

respectively. With the definitions in Eqs. (3.133), (3.138) and (3.152) one has for the macroscopic erasure power

$$\mathcal{P}_s^s\left(\frac{1}{2},p_f\right) = \frac{\Delta s(1/2,p_f)}{\tau}, \quad \mathcal{P}_m^m\left(\frac{1}{2},P_f\right) = \frac{\Delta S(1/2,P_f)}{\tau}, \tag{3.162}$$

and the macroscopic efficiencies

$$\eta_s\left(\frac{1}{2},p_f\right) = \frac{\Delta s(1/2,p_f)}{\beta\, q}, \quad \eta_m\left(\frac{1}{2},P_f\right) = \frac{\Delta S(1/2,P_f)}{\beta\, Q}. \tag{3.163}$$

The results of the finite-time erasure process for a fixed energy gap are depicted in Fig. 3.27.

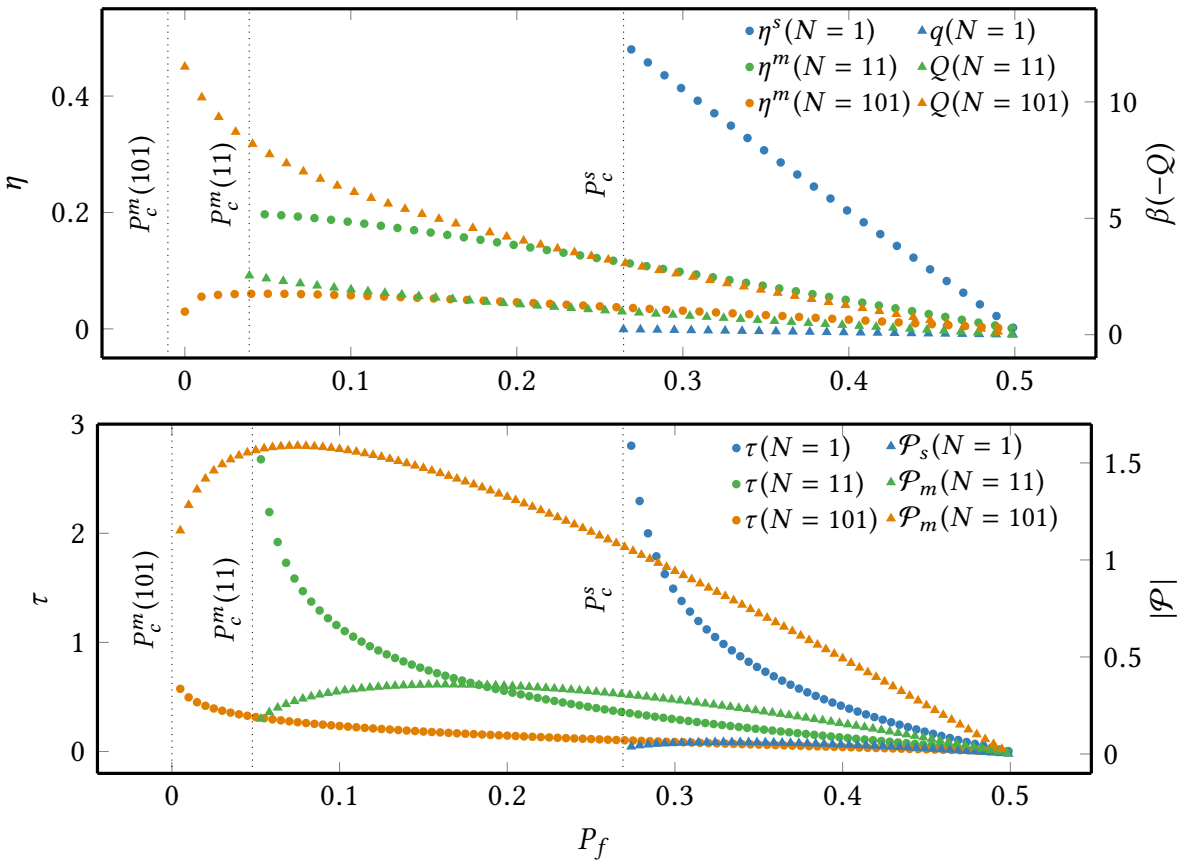

Figure 3.27: *a) Comparison of both the modulus of heat dissipation, $-\beta Q$, and the associated erasure efficiency between a SUB and a symmetric MLB for different N. b) Comparison of both the erasure time and the modulus of the power, $|\mathcal{P}|$, between a SUB and a symmetric MLB for different N. The data was generated by using the generic solution of the master equation* (3.157) *and setting $w_{01}, \beta\epsilon \equiv 1$. The dotted vertical lines correspond to the minimal erasure error P_c^s and $P_c^m(N)$ for the systems under consideration.*

As can be seen in panel a), in order to perform an erasure with the same erasure error the SUB dissipates less heat and thus has a higher efficiency than the symmetric MLBs, for which the heat production increases and the efficiency decreases with growing N. Hence the SUB is more efficient than the symmetric majority-logic decoding, as already observed in the reversible case, cf. Fig. 3.26. However, Fig. 3.27b) shows that the erasure duration is reduced and the erasure power thus enhanced by employing the majority-logic decoding for large ensembles of microscopic units. The minimal erasure error characterized by Eq. (3.158) is indicated by the dotted vertical lines that correspond to the minimal final probability $P_c^m(N)$. Therefore, as already derived in Eq. (3.151), majority-logic decoding reduces the minimal erasure error that goes to zero as N becomes large corresponding to perfect erasure.

3.3.2.4 Fixed Erasure Duration

We now assume that the explicit form of the transition rates of the microscopic unit model is known and compare the performance of the two types of macroscopic bits under the protocol with fixed erasing time and instantaneous switching of the energy gap. For a specific erasure process, the erasure time and thus the erasure power are equal for the two macroscopic bits, hence we restrict the discussion to the macroscopic erasure efficiencies. The microscopic dynamics of a Fermi-rates unit reads

$$p(t) = \frac{1}{2}\tanh\left(\frac{\beta\Delta\epsilon}{2}\right)\mathrm{e}^{-t} + \frac{1}{1+\mathrm{e}^{\beta\Delta\epsilon}}. \tag{3.164}$$

Plugging the erasure time τ into Eq. (3.164), yields the final microscopic probability and the erasure error. The calculation of the generated heat, erasure power and efficiency both on the microscopic and macroscopic level is analogous to the one in the last section. As the specific dynamics of the unit is known and τ is fixed, the definition of minimal final probability in Eq. (3.148) is valid under this protocol.

Fig. 3.28 shows that the symmetric majority-logic decoding has additional advantages for finite-time erasing: As already observed earlier, the minimal erasure error of the symmetric MLB is smaller than that of the SUB and approaches 0 with increasing N as illustrated by the vertical dotted lines corresponding to the minimal final probability $P_c^m(N)$ after the erasure time τ. More importantly, the symmetric MLBs are more efficient in the region of small erasure error region, $P_f \approx P_c^s$, as opposed to the region of large erasure error where the SUB is more efficient. We find that for Arrhenius rates the results are qualitatively similar and thus omitted.

To sum up, Figs. 3.27 and 3.28 illustrate the important result that the precision-speed-efficiency trade-off in finite-time information erasure processes is lifted by symmetric majority-logic decoding.

3.3.2.5 Optimal Erasure Protocol under Majority-Logic Decoding

In view of applications, the least work-intense erasure processes are of particular interest. The general method to determine the optimal erasure protocol that minimizes the generated heat in a Fermi-rates unit model has been established in Ref. [95]. The detailed derivation of the heat-minimizing protocol for Arrhenius rates is deferred to Appendix A.5. Given the optimal protocol, both the microscopic and macroscopic heat dissipation, erasure power and efficiency follow readily from Eqs. (3.160), (3.162) and (3.163).

The erasure efficiency of a SUB and a symmetric MLBs for different N are compared in Fig. 3.29. It is important to note that the advantages in terms of information erasure inherent to symmetric majority-logic decoding discussed in the previous section are preserved and enhanced by the optimal erasure protocol: First, the minimal erasure error is strongly reduced for a symmetric MLB at the expense of an erasure efficiency that decreases with increasing N in the regime of small erasure, $P_f \gg P_c^s$. Conversely, for small erasure error, $P_f \approx P_c^s$, this

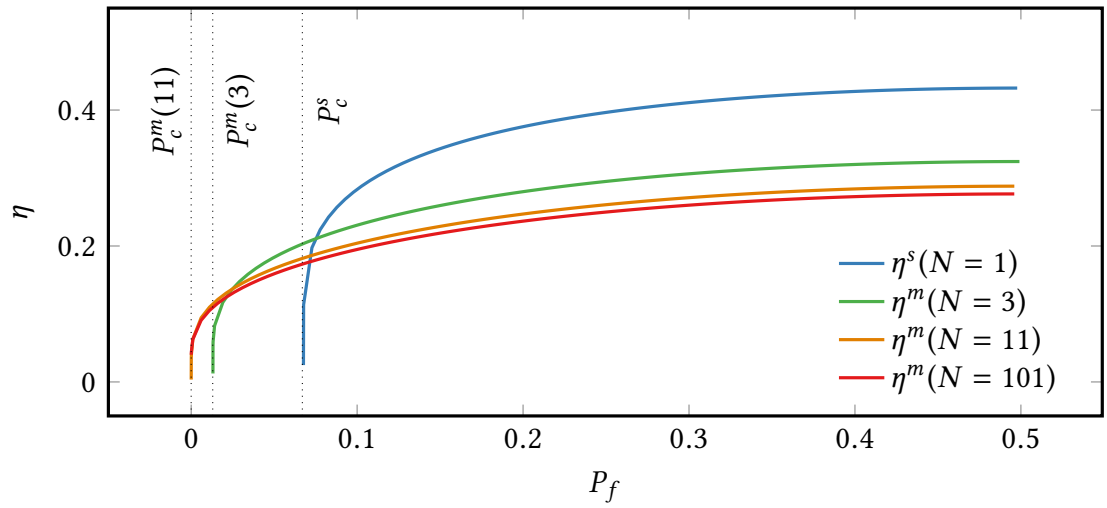

Figure 3.28: *Comparison between the erasure efficiency of a SUB and symmetric MLBs for different N and instantaneously switched $\Delta\epsilon$ using the Fermi-rates unit model and fixed erasure time $\tau = 2$. The vertical dotted lines correspond to the minimal final macroscopic probabilities P_c^s and $P_c^m(N)$, where for $N = 11, 101$ the probabilities are too close to zero to be distinguished.*

relation between the macroscopic efficiencies is inverted. The inset in panel a) that depicts the relative erasure efficiency between a SUB and a symmetric MLB, $(\eta_s - \eta_m)/\eta_s$, reveals that the range of small erasure-error probabilities over which this holds true is increased for the optimal protocol compared to the fixed-energy protocol. Comparing furthermore panels a) and b) also shows that this range of small erasure-error probabilities is increasing with decreasing erasure time τ and increasing N.

In the limit of low- and high dissipation, analytical results can be obtained for the optimal protocol, where we will focus on the Arrhenius-rates unit model. The high-dissipation limit corresponds to an erasure duration τ that approaches the minimal erasure time τ_c. Thus, the parameter K in Eq. (A.71), which represents the degree of irreversibility, is diverging in this limit. Using Eqs. (A.73) and (A.74), the parameter K can be expressed as

$$K^{high}(p_f, \tau) = \frac{1}{2}\frac{1-2p_f}{\tau + \ln(2p_f)}. \tag{3.165}$$

Plugging Eq. (3.165) into Eq. (A.75), one has for the heat dissipation

$$\beta\, q^{high}(p_f, \tau) = \left(\frac{1}{2} - p_f\right) \ln\left[\frac{1}{2}\frac{1-2p_f}{\tau + \ln(2p^f)}\right]. \tag{3.166}$$

We verify that $K, q \to \infty$ as $\tau \to \tau_c = -\ln(2p_f)$ and that Eq. (3.166) also represents the solution for a Fermi-rates unit. In the low-dissipation limit, the erasure duration τ is large

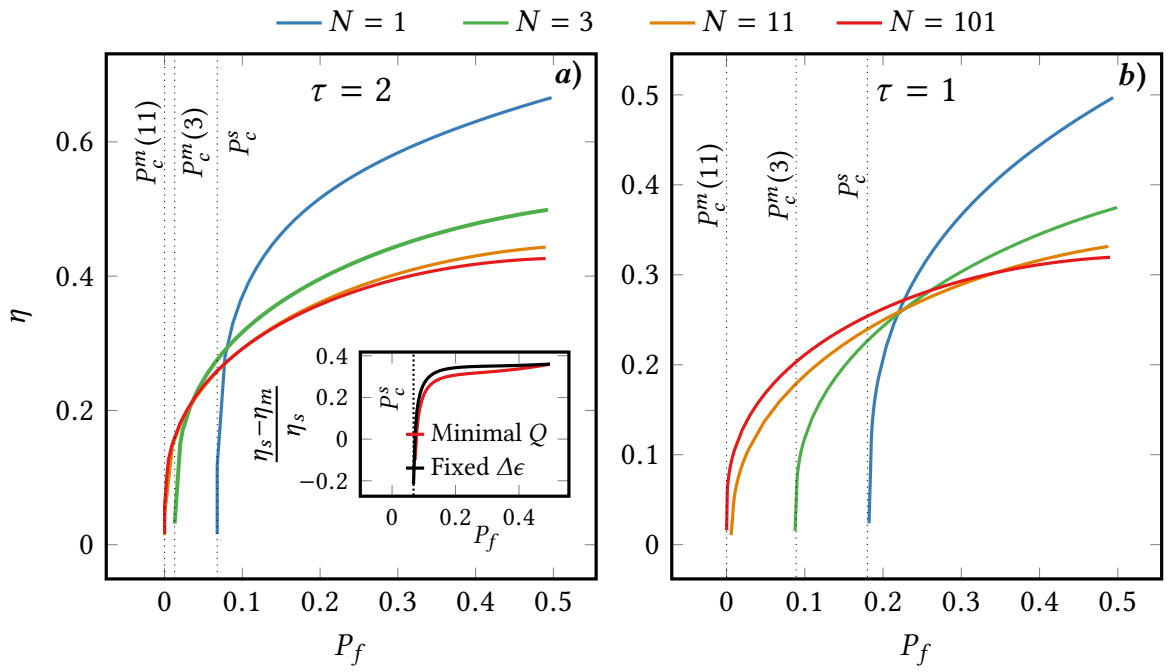

Figure 3.29: *Comparison between the erasure efficiency of a SUB and of a symmetric MLB based on the Arrhenius model for different N under the optimal erasure protocol with the erasure time $\tau = 2$ [panel a)] and $\tau = 1$ [panel b)]. The vertical dotted lines correspond to the minimal final macroscopic P_c^s and probabilities $P_c^m(N)$, where for $N = 11, 101$ the probabilities are too close to zero to be distinguished. The inset in panel a) shows the erasure efficiency of a SUB compared to a symmetric MLB one, $(\eta_s - \eta_m)/\eta_s$ for $N = 101$ using the optimal and the fixed-energy protocol.*

compared to τ_c, hence the parameter K is small. Using Eqs. (A.73) (A.74), the parameter K can be expressed as

$$K^{low}(p_f, \tau) = 8\left[\frac{1-\sqrt{2p_f}}{2\tau + \ln(2p_f)}\right]^2. \tag{3.167}$$

Plugging Eq. (3.167) into Eq. (A.75), one has for the heat dissipation

$$\beta\, q^{low}(p_f, \tau) = \frac{2\left(1-\sqrt{2p_f}\right)^2}{\tau} - \Delta s(1/2, p_f). \tag{3.168}$$

Significantly, the first term in Eq. (3.168), which could be interpreted as the irreversible dissipation, is consistent with the low dissipation assumption made in Refs. [44, 161, 162]. Here, the expression for the heat in Eq. (3.168) differs from that of a Fermi-rates unit at low dissipation. The heat dissipated by a SUB in the high (q^{high}) and low-dissipation limit (q^{low})

are the same as those of the microscopic unit and therefore given by Eqs. (3.166) and (3.168), respectively. The heat dissipated by a symmetric MLB in the high (Q^{high}) and low-dissipation limit (Q^{low}) are readily derived using Eq. (3.127) with results from Eqs. (3.166) and (3.168). For small erasure ($p_f \to 1/2$), the expression for the heat in the low-dissipation limit in Eq. (3.168) simplifies to

$$q^{low} \approx 2\left(1+\frac{1}{\tau}\right)\left(\frac{1}{2}-p_f\right)^2. \tag{3.169}$$

Using the approximation for the majority-logic decoding in the limit of small erasure and large N in Eq. (3.129), the ratio κ between the erasure efficiency of a symmetric MLB and a SUB

$$\kappa \equiv \frac{\eta_m(P_f,N,\tau)}{\eta_s(p_f,\tau)} \approx \frac{2}{\pi}, \tag{3.170}$$

is independent of the erasure duration τ and final erasure error.

Fig. 3.30a) compares the dissipated heat of a SUB and a symmetric MLB for different erasure durations $\log\tau$ using the optimal protocol applied to a given erasure process from $P_i = 1/2$ to P_f.

Additionally, the approximate solutions for the low- and high-dissipation limit in Eqs. (3.166) and (3.168) are overlaid. As can be seen, the symmetric MLB under the optimal protocol is still more efficient in the fast erasure region, where the erasing duration τ approaches the minimal erasure time $\tau_c^s(p_f)$ of a SUB. Except for extremely fast erasure processes, the full numerical solution of the heat dissipated by the symmetric MLB is in excellent agreement with the low-dissipation approximation in Eq. (3.169). This suggests that the calculation of the dissipated heat of a symmetric MLB built upon microscopic Arrhenius-rate units under the optimal erasure protocol can be simplified by using the more convenient Eq. (3.169) instead of the numerically more involving procedure elaborated in Appendix A.5.

In panel 3.30b) the ratio between the efficiency of a symmetric MLB and a SUB are compared for the same erasure process. In agreement with the observation made in panel a) that the heat dissipation in the fast erasure region is significantly reduced for a symmetric MLB, we note in panel b) that the erasure efficiency of a symmetric MLB is considerably higher than that of a SUB ($\kappa > 1$) in this region. It is interesting to notice that even though the approximate expression for the efficiency ratio in Eq. (3.170) is theoretically only valid in asymptotic limit $P_f \to 1/2$, the full numerical solutions are in good agreement with the approximate solution ($2/\pi$) for a large range of values for P_f.

We therefore conclude that with the aid of symmetric majority-logic decoding, the energetically optimized erasure processes can be accelerated, be performed more precise at a lower cost. Hence, for large-N symmetric MLBs the speed-precision-efficiency trade-off is significantly lifted. We remark that the optimization procedure of the Fermi-rates unit model is demonstrated in Appendix A.5 and was already studied in Ref. [95]. The results on the performance of a symmetric MLB built upon Fermi-rates units are qualitatively similar

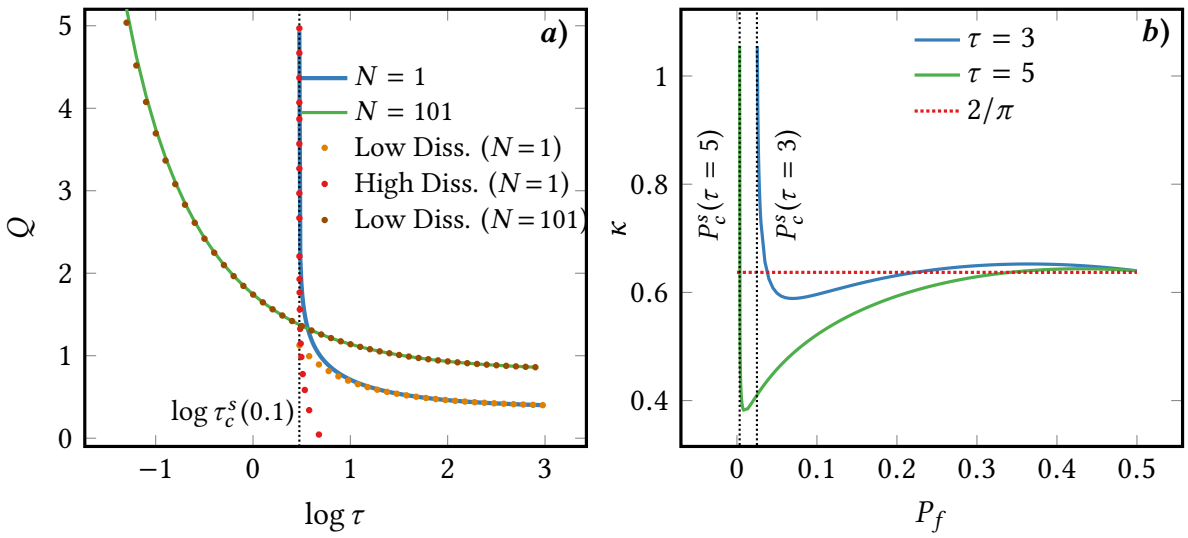

Figure 3.30: *a) Comparison between the dissipated heat of a SUB and a symmetric MLB ($N = 101$) based on Arrhenius-rates units as a function of erasure duration $\log \tau$ for the optimal erasure protocol with $P_f = 0.1$. In addition to the full numerical solution of Eq.* (A.75)*, the analytic low- and high-dissipation solutions in Eqs.* (3.168) *and* (3.166) *are displayed. The vertical dotted line corresponds to the logarithm of the critical time $\log \tau_c^s(0.1)$ of a single microscopic unit for this specific erasure process. b) Ratio between the erasure efficiencies of a symmetric MLB ($N = 101$) and a SUB as a function of the final macroscopic probability P_f for the optimal protocol with the erasure times $\tau = 3, 5$. The red dotted line corresponds to the small-erasure and large-N approximation given by Eq.* (3.170) *and the vertical black dotted lines represent the minimal erasure error $P_c^s(\tau)$.*

to those for the Arrhenius-rates unit model discussed above and thus omitted. It should however be emphasized that the advantages of employing the symmetric majority-logic decoding for information processing are preserved under a change between these two different microscopic models.

Chapter 4

Coarse-Graining in Stochastic Thermodynamics

4.1 Introduction

In Sec. 3.2.1 we established a thermodynamic description for average values across scales - microscopic, mesoscopic and mean field - in the three-state model. In fact, the different thermodynamic descriptions can be extended towards fluctuations and for more general many-body systems, if the assumption of all-to-all interactions is retained. This is achieved by considering a system made up of N identical and all-to-all interacting units with q discrete states which is coupled to multiple heat reservoirs and subjected to generic autonomous and non-autonomous external forces:

At the *microscopic* level, the system is characterized by microstates which correspond to the many-body states that define the state of each of the units. Owing to the all-to-all interactions, the microscopic stochastic dynamics ruled by a microscopic Markov master equation can be exactly coarse grained to a *mesoscopic* stochastic dynamics governed by a mesoscopic Markov master equation, where the system is specified by the global occupation numbers of each of the unit states. At both levels, the transition rates satisfy the local detailed balance condition and are thus thermodynamically consistent, whereby the mesoscopic transition rates differ from the microscopic ones by an entropic factor. The latter accounts for the internal structure of the mesostates reflecting that the microstates insisde the mesostates are energetically indistinguishable. It will be demonstrated that the stochastic thermodynamics is invariant under this dynamically exact coarse-graining, if the microstates inside each mesostate are equiprobable.

In order to compute fluctuations at macroscopic scales, $N \to \infty$, (*i.e.* fluctuations that

are extensive in and thus scale exponentially with the system size [163]), we employ a path integral representation of the stochastic dynamics known as Martin-Siggia-Rose formalism [164, 165]. Employing this formalism, we find that, as in the q-state clock models, the mesoscopic stochastic dynamics exactly reduces to a nonlinear mean-field rate equation for the deterministic mesoscopic state variables which correspond to the most likely values of the occupation of each unit state. Again, the nonlinear rates satisfy local detailed balance and the deterministic mean-field thermodynamics is thus thermodynamically consistent. In the deterministic macroscopic limit the fluctuating Shannon entropy (2.63) vanishes identically and only the aforementioned internal entropy due to the equienergetic microstates inside the mesostates remains. Remarkably, in spite of being deterministic, the internal entropy takes the form of a Shannon entropy for the deterministic occupation in the macroscopic limit. The macroscopic fluctuations will be shown to satisfy the detailed fluctuation theorem in Eq. (2.105), which will also be derived at the microscopic and mesoscopic level, and are thus also thermodynamically consistent. The methodology to calculate macroscopic fluctuations is employed in a semi-analytically solvable Ising model.

Furthermore, as a complement to the study of underdamped many-body systems with all-to-all interactions, we also want to investigate coarse-graining in underdamped systems. For this purpose, we will consider two coupled underdamped particles as a mininmal underdamped model that contains the necessary key ingredients. Among the many coarse-graining schemes preserving thermodynamic consistency that have been proposed in the literature [99–111], we want to apply three different methods that have proven instrumental for underdamped stochastic dynamics governed by master equations:

First, the most straightforward approach where a subset of states is explicitly coarse-grained and the effective thermodynamics is defined for that reduced dynamics as one formally would for the full dynamics [112, 113]. Next, an approach based on splitting the full system in two parts resulting in effective second laws for each parts which are modified by a term describing the transfer of mutual information between each parts. This approach provides a convenient framework to describe how a Maxwell demon [166] mechanism can produce an information flow that is consumed by the system to drive processes against their spontaneous direction [114–117]. Finally, the so-called Hamiltonian of mean force approach which introduces a notion of energy for a system strongly coupled to its environment [118–120].

In particular, we will compare the effective thermodynamic descriptions resulting from these coarse-graining schemes for two distinct physical limits: The effective thermodynamics based on marginalization and the Hamiltonian of mean force become equivalent and capture the correct global thermodynamics in the limit of time-scale separation. In this limit, the second particle has a much faster relaxation dynamics compared to the first one so that it instantaneously relaxes to a local equilibrium with respect to the coordinates of the first particle. Conversely, the thermodynamics based on the slow part of the bipartite structure does not agree with the full thermodynamics. The mismatch corresponds to the entropic contribution due to the coupling of the second particle. Physically, in the limit of time-scale separation the coarse-grained particle becomes part of the heat reservoir. Moreover, in the

limit where one particle has an exceedingly large mass compared to the other one, we will find that the former becomes a work source acting on the latter. In that case, the effective thermodynamics emerging from the first two coarse-graining schemes, marginalization and bipartite structure, again captures the correct global thermodynamics (at least up to a trivial macroscopic friction term in the work source). In contrast, we will show that the Hamiltonian is incompatible with that limit. These theoretical predictions will be confirmed using an analytically tractable model made up of two linearly coupled harmonic oscillators.

4.2 Many-Body Systems with All-To-All Interactions

We start by revisiting the coarse-graining applied on the all-to-all interacting many-body systems in Sec. 3.2. In the following, the thermodynamics under this coarse-graining will be extended to the fluctuating level for a more general setup of all-to-all interacting systems.

4.2.1 Stochastic Dynamics

4.2.1.1 Microscopic Description

Let us consider a system that consists of N all-to-all interacting identical and classical units that consist of q states i with energies $\epsilon_i(\lambda_t)$ that are varying in time according to a known protocol λ_t of an external driving. The system is coupled with multiple heat reservoirs $\nu = 1, 2, \dots, L$ at inverse temperatures $\beta^{(\nu)}$. Each unit is assumed to be fully connected, *i.e.* any state of a given unit can be reached within a finite number of steps from all other states of that unit, so that the global system is irreducible. Moreover, we suppose that all units are subjected to generic nonconservative forces $f_{ij}^{(\nu)}$. Depending on whether a transition is aligned with or acting against the nonconservative force, the latter fosters or represses the transition from state j to i. For generality, the force is assumed to be different depending on which heat reservoir ν the system is exchanging energy with during the transition from j to i. Until explicitly states otherwise, we will take N to be finite in the following.

The many-body system is unambiguously characterized by a microstate

$$\alpha = (\alpha_1, \dots, \alpha_i, \dots, \alpha_N), \quad \alpha_i = 1, 2, \dots, q. \tag{4.1}$$

The system energy consists of the state occupation of the units and the interactions between them. For all-to-all interactions, we readily determine the energy of the system in a microstate α as follows,

$$e_\alpha(\boldsymbol{\lambda}_t) = \sum_{i=1}^{q} \Big\{ N_i(\alpha)\, \epsilon_i(\lambda_t) + \frac{u_i(\lambda_t')}{2N}\, N_i(\alpha)\big[N_i(\alpha) - 1\big] + \sum_{j<i} \frac{u_{ij}(\lambda_t')}{N} N_i(\alpha)\, N_j(\alpha) \Big\}, \tag{4.2}$$

where $u_i(\lambda'_t)/N$ and $u_{ij}(\lambda'_t)/N$ denote the pair potential of units occupying the same or different single-unit states, respectively. These interactions can be tuned by an external driving according to a known protocol λ'_t, hence $\boldsymbol{\lambda}_t = \left(\lambda_t, \lambda'_t\right)^\top$. Moreover, $N_i(\alpha)$ refers to the number of units N_i occupying the single-unit state i for a given microstate α.

The stochastic jump process is governed by an irreducible Markovian master equation which describes the time evolution of the microscopic probability p_α for the system to be in the microstate α as follows,

$$\dot{p}_\alpha(t) = \sum_{\alpha'} w_{\alpha\alpha'}(\boldsymbol{\lambda}_t)\, p_{\alpha'}(t)\,, \tag{4.3}$$

with the microscopic rates $w_{\alpha\alpha'}(\boldsymbol{\lambda}_t)$ for transitions from α' to α that in general depend on the current value of the driving parameter $\boldsymbol{\lambda}_t$. We note that probability conservation is ensured by the stochastic property of the transition rate matrix, $\sum_\alpha w_{\alpha\alpha'}(\boldsymbol{\lambda}_t) = 0$. The transition from α' to α is induced by one of the L heat reservoirs, thus

$$w_{\alpha\alpha'}(\boldsymbol{\lambda}_t) = \sum_{\nu=1}^{L} w^{(\nu)}_{\alpha\alpha'}(\boldsymbol{\lambda}_t). \tag{4.4}$$

Here, for simplicity we assume the additivity property of the microscopic rates, on which the validity of the following does not rely on. For a more general case than in Eq. (4.4) we refer to Ref. [133]. The microscopic transition rates that specify the heat reservoir satisfy the microscopic local detailed balance condition separately,

$$\frac{w^{(\nu)}_{\alpha\alpha'}(\boldsymbol{\lambda}_t)}{w^{(\nu)}_{\alpha'\alpha}(\boldsymbol{\lambda}_t)} = \mathrm{e}^{-\beta^{(\nu)}\left[e_\alpha(\boldsymbol{\lambda}_t) - e_{\alpha'}(\boldsymbol{\lambda}_t) - f^{(\nu)}_{\alpha\alpha'}\right]}, \tag{4.5}$$

which in turn ensures the thermodynamic consistency of the system. Here, $\boldsymbol{f}^{(\nu)}_{\alpha\alpha'}$ is the element of the nonconservative force vector $\boldsymbol{f}^{(\nu)}$ that is equal to $f^{(\nu)}_{ij}$, if the microscopic transition from $\alpha' \longrightarrow \alpha$ corresponds to a single-unit transition from $j \longrightarrow i$. The microscopic local detailed balance condition (4.5) that constrains the asymmetric part of the microscopic transition rates. We also need to make the crucial assumption that the symmetric part of the microscopic transition rates (*e.g.* kinetic prefactor in Arrhenius rates) does not depend on the kinetics and is thus not a function of the microstates.

If the transition rates are kept constant, $\boldsymbol{\lambda}_t = \boldsymbol{\lambda}$, the dynamics will relax into a unique stationary state, $\partial_t p^s_\alpha(\boldsymbol{\lambda}) = 0$. If furthermore all heat reservoirs have the same inverse temperature, $\beta^{(\nu)} = \beta\ \forall \nu$, and the nonconservative forces vanish, $\boldsymbol{f}^{(\nu)} = 0\ \forall \nu$, the stationary distribution coincides with the equilibrium one which satisfies the microscopic detailed balance condition,

$$w_{\alpha\alpha'}(\boldsymbol{\lambda})\, p^{eq}_{\alpha'}(\boldsymbol{\lambda}) = w_{\alpha'\alpha}(\boldsymbol{\lambda})\, p^{eq}_{\alpha}(\boldsymbol{\lambda}). \tag{4.6}$$

The local detailed balance (4.5) implies that the microscopic equilibrium distribution assumes the canonical form,

$$p^{eq}_\alpha(\boldsymbol{\lambda}) = \mathrm{e}^{-\beta\left[e_\alpha(\boldsymbol{\lambda}) - a^{eq}(\boldsymbol{\lambda})\right]}, \tag{4.7}$$

with the microscopic equilibrium free energy

$$a^{eq}(\boldsymbol{\lambda}) = -\frac{1}{\beta} \ln \sum_{\alpha} \mathrm{e}^{-\beta e_{\alpha}(\boldsymbol{\lambda})} . \tag{4.8}$$

4.2.1.2 Mesoscopic Description

The microscopic state space grows exponentially with the number of units, $||\alpha|| = q^N$. Yet, owing to the all-to-all interactions, there are equi-energetic microstates that are characterized by the same values for the occupation numbers N_i. This means that the energetics (4.2) and thus the microscopic transition rates (4.4) do not depend on which particular unit is in a given state but only on the global occupation numbers, *i.e.* the number of units occupying each state of the unit. As a consequence, the microscopic dynamics can be marginalized into a mesoscopic one, where the mesostate $\boldsymbol{N} \equiv (N_1, N_2, \ldots, N_q)$ now identifies the state of the system. We denote by $\alpha_{\boldsymbol{N}}$ the equienergetic microstates α inside a mesostate $\boldsymbol{N}$, that is microstates for which the relation

$$e_{\alpha_{\boldsymbol{N}}}(\boldsymbol{\lambda}_t) = E_{\boldsymbol{N}}(\boldsymbol{\lambda}_t), \tag{4.9}$$

holds. The number $\Omega_{\boldsymbol{N}}$ of microstates which belong to a mesostate is given by

$$\Omega_{\boldsymbol{N}} = \binom{N}{N_1}\binom{N-N_1}{N_2}\cdots\binom{N-N_1-\ldots-N_q}{N_q} = \frac{N!}{\prod_{i=1}^{q} N_i!} . \tag{4.10}$$

We introduce the mesoscopic probability to observe the mesostate $\boldsymbol{N}$

$$P_{\boldsymbol{N}}(t) \equiv \sum_{\alpha_{\boldsymbol{N}}} p_{\alpha_{\boldsymbol{N}}}(t) , \tag{4.11}$$

and we have for the conditional probability to find the system in a microstate $\alpha_{\boldsymbol{N}}$ that belongs to that mesostate,

$$\mathbb{P}_{\alpha_{\boldsymbol{N}}}(t) = \frac{p_{\alpha_{\boldsymbol{N}}}(t)}{P_{\boldsymbol{N}}(t)}, \tag{4.12}$$

for which because of probability conservation holds that

$$\sum_{\alpha_{\boldsymbol{N}}} \mathbb{P}_{\alpha_{\boldsymbol{N}}}(t) = 1. \tag{4.13}$$

With Eqs. (4.9), (4.11) and (4.13) the microscopic master equation (4.3) can be exactly coarse-grained as follows,

$$\partial_t P_{\boldsymbol{N}}(t) = \sum_{\boldsymbol{N}'} \sum_{\alpha_{\boldsymbol{N}}} \sum_{\alpha'_{\boldsymbol{N}'}} w_{\alpha_{\boldsymbol{N}},\alpha'_{\boldsymbol{N}'}}(\boldsymbol{\lambda}_t)\ \mathbb{P}_{\alpha'_{\boldsymbol{N}'}}(t)\ P_{\boldsymbol{N}'}(t) = \sum_{\boldsymbol{N}'} W_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t)\ P_{\boldsymbol{N}'}(t) , \tag{4.14}$$

with the mesoscopic transition rates $W_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t) = \Omega_{\boldsymbol{N},\boldsymbol{N}'}\; w_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t)$. The quantity $\Omega_{\boldsymbol{N},\boldsymbol{N}'}$ takes into account that only those microstates $\alpha_{\boldsymbol{N}}$ and $\alpha'_{\boldsymbol{N}'}$ contribute to the sum in Eq. (4.14) which are connected to each other. This amounts to determine how many microstates α belong to the mesostate $\boldsymbol{N}$ under the constraint that they are connected to microstates α' belonging to the mesostate $\boldsymbol{N}'$. The combinatorial problem is readily solved by noting that the occupation number that is decremented during the transition corresponds to the wanted quantity, *i.e.*

$$\Omega_{\boldsymbol{N},\boldsymbol{N}'} = \sum_{i=1}^{q} N'_i\, \delta_{N'_i,N_i+1}\,, \tag{4.15}$$

where $N_i + 1$ is understood as $(N_i + 1) \bmod q$. It is easy to verify that the stochastic property of the transition rate matrix is preserved by the coarse-graining, $\sum_{\boldsymbol{N}} W_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t) = 0$. The mesoscopic transition rates are still consisting of multiple contributions due to the different heat reservoirs,

$$W_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t) = \sum_{\nu=1}^{L} W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t), \tag{4.16}$$

that separately preserve the microscopic local detailed balance relation (4.5) at the mesoscopic level,

$$\frac{W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t)}{W^{(\nu)}_{\boldsymbol{N}'\boldsymbol{N}}(\boldsymbol{\lambda}_t)} = \mathrm{e}^{-\beta^{(\nu)}\left[A^{(\nu)}_{\boldsymbol{N}}(\boldsymbol{\lambda}_t) - A^{(\nu)}_{\boldsymbol{N}'}(\boldsymbol{\lambda}_t) - f^{(\nu)}_{\boldsymbol{N},\boldsymbol{N}'}\right]}, \tag{4.17}$$

with the notation $\boldsymbol{f}^{(\nu)}_{\boldsymbol{N},\boldsymbol{N}'}$ that is defined as $\boldsymbol{f}^{(\nu)}_{\alpha,\alpha'}$ in Eq. (4.5). Here, we introduced the free energy of a mesostate

$$A^{(\nu)}_{\boldsymbol{N}}(\boldsymbol{\lambda}_t) = E_{\boldsymbol{N}}(\boldsymbol{\lambda}_t) - \frac{1}{\beta^{(\nu)}}\, S^{int}_{\boldsymbol{N}}, \tag{4.18}$$

and used the Boltzmann entropy

$$S^{int}_{\boldsymbol{N}} = \ln \Omega_{\boldsymbol{N}}\,, \tag{4.19}$$

along with the relation

$$\frac{\Omega_{\boldsymbol{N}}}{\Omega_{\boldsymbol{N}'}} = \frac{\Omega_{\boldsymbol{N},\boldsymbol{N}'}}{\Omega_{\boldsymbol{N}',\boldsymbol{N}}}, \tag{4.20}$$

which can be seen by using Eqs. (4.10) and (4.15).

If the transition rates are kept constant, $\boldsymbol{\lambda}_t = \boldsymbol{\lambda}$, the dynamics will reach a unique stationary state, $\partial_t P^s_{\boldsymbol{N}}(\boldsymbol{\lambda}) = 0$. If furthermore all heat reservoirs have the same inverse temperature, $\beta^{(\nu)} = \beta\ \forall \nu$, and the nonconservative forces vanish, $\boldsymbol{f}^{(\nu)} = 0\ \forall \nu$, the stationary

distribution coincides with the equilibrium one which satisfies the mesoscopic detailed balance condition,

$$W_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda})\, P^{eq}_{\boldsymbol{N}'}(\boldsymbol{\lambda}) = W_{\boldsymbol{N}'\boldsymbol{N}}(\boldsymbol{\lambda})\, P^{eq}_{\boldsymbol{N}}(\boldsymbol{\lambda}). \tag{4.21}$$

and because of Eq. (4.17) assumes the canonical form,

$$P^{eq}_{\boldsymbol{N}}(\boldsymbol{\lambda}) = \mathrm{e}^{-\beta[A_{\boldsymbol{N}}(\lambda) - A^{eq}(\boldsymbol{\lambda})]}, \tag{4.22}$$

with the mesoscopic equilibrium free energy

$$A^{eq}(\boldsymbol{\lambda}) = -\frac{1}{\beta} \ln \sum_{\boldsymbol{N}} \mathrm{e}^{-\beta A_{\boldsymbol{N}}(\boldsymbol{\lambda})}\,. \tag{4.23}$$

The marginalization of the equienergetic microstates significantly reduces the complexity of the system since the mesoscopic state space asymptotically grows like a power law,

$$||\boldsymbol{N}|| = \sum_{N_1=0}^{N} \sum_{N_2=0}^{N_1} \cdots \sum_{N_{q-1}=0}^{N_{q-2}} 1 \overset{N\to\infty}{\sim} \frac{N^{q-1}}{(q-1)!}, \tag{4.24}$$

as opposed to the exponential growth of the microscopic state space.

Since it will be useful further below, we now prove that for a stationary mesoscopic distribution, all microstates that belong to the respective mesostates are uniformly distributed. For finite systems, the stationary microscopic probabilities can be determined via the spanning-tree formula [150]. A spanning tree, $\mathcal{T}(G)$ of a graph G consists only of edges that are also edges of G and contains all vertices (microstates α) of G. Further, a spanning tree $\mathcal{T}'(G)$ is connected and contains no circuits. We write $\mathcal{T}^{(\mu)}_{\alpha}(G)$ for the μth spanning tree rooted in α, that is a tree with branches that are directed towards the vertex α. The spanning-tree formula reads

$$p^s_{\alpha}(\boldsymbol{\lambda}) = \frac{\sum\limits_{\mu} \mathcal{T}^{(\mu)}_{\alpha}(G)}{\sum\limits_{\alpha}\sum\limits_{\mu} \mathcal{T}^{(\mu)}_{\alpha}(G)} = \frac{\sum\limits_{\mathcal{T}_{\alpha}(G)} \prod\limits_{\substack{\alpha',\alpha'' \text{ such that} \\ \text{current is directed to } \alpha}} w_{\alpha'\alpha''}(\boldsymbol{\lambda})}{\sum\limits_{\alpha}\sum\limits_{\mathcal{T}_{\alpha}(G)} \prod\limits_{\substack{\alpha',\alpha'' \text{ such that} \\ \text{current is directed to } \alpha}} w_{\alpha'\alpha''}(\boldsymbol{\lambda})}\,. \tag{4.25}$$

First we note that the transition rates do not depend on the microstates α and α' belonging to the same pair of mesostates $(\boldsymbol{N}, \boldsymbol{N}')$, *i.e.* $w_{\alpha_{\boldsymbol{N}}\alpha'_{\boldsymbol{N}'}}(\boldsymbol{\lambda}) = \text{const}\ \ \forall\, \alpha_{\boldsymbol{N}}, \alpha'_{\boldsymbol{N}'}$. Secondly, the number of possible transitions for any microstate belonging to a given mesostate is a constant such that the number of spanning trees rooted in $\alpha_{\boldsymbol{N}}$ is constant for all $\alpha_{\boldsymbol{N}}$. It therefore holds that all microstates belonging to the same mesostate are equally probable,

$$p^s_{\alpha_{\boldsymbol{N}}}(\boldsymbol{\lambda}) = \frac{\sum\limits_{\mathcal{T}_{\alpha_{\boldsymbol{N}}}(G)} \prod\limits_{\substack{\alpha'_{\boldsymbol{N}},\alpha''_{\boldsymbol{N}} \text{ such that} \\ \text{current is directed to } \alpha_{\boldsymbol{N}}}} w_{\alpha'_{\boldsymbol{N}},\alpha''_{\boldsymbol{N}}}(\boldsymbol{\lambda})}{\sum\limits_{\boldsymbol{N}}\sum\limits_{\alpha_{\boldsymbol{N}}}\sum\limits_{\mathcal{T}_{\alpha_{\boldsymbol{N}}}(G)} \prod\limits_{\substack{\alpha'_{\boldsymbol{N}},\alpha''_{\boldsymbol{N}} \text{ such that} \\ \text{current is directed to } \alpha_{\boldsymbol{N}}}} w_{\alpha'_{\boldsymbol{N}},\alpha''_{\boldsymbol{N}}}(\boldsymbol{\lambda})} = \text{const}\ \ \forall\, \alpha_{\boldsymbol{N}}, \tag{4.26}$$

and we therefore find with Eq. (4.12) that

$$\mathbb{P}^{s}_{\alpha_{\boldsymbol{N}}} = \frac{1}{\Omega_{\boldsymbol{N}}}, \qquad p^{s}_{\alpha_{\boldsymbol{N}}}(\boldsymbol{\lambda}) = \frac{P^{s}_{\boldsymbol{N}}(\boldsymbol{\lambda})}{\Omega_{\boldsymbol{N}}}. \tag{4.27}$$

We stress that the global state at the mesoscopic level need not to be at equilibrium for the last two equations to hold.

We demonstrated that for thermodynamically consistent and discrete many-body systems with all-to-all interactions with no nontrivial kinetic dependencies there is an exact coarse-graining of the microscopic stochastic dynamics characterized by many-body states towards a mesoscopic stochastic dynamics that is fully characterized by the global occupation of the different unit states. It is however *a priori* not ensured that the thermodynamic structures built on top of these Markov process using stochastic thermodynamics are equivalent. This issue is investigated in the following section.

4.2.2 Stochastic Thermodynamics

4.2.2.1 Trajectory Definitions Revisited

After having established the stochastic dynamics at microscopic and mesoscopic scales, the following is devoted to formulating the stochastic thermodynamic quantities across these scales. To this end, we first introduce the fluctuating quantities at the level of a single trajectory defined in Eq. (2.57) and depicted in Fig. 2.3. The elementary trajectory observables have already been generically defined in Sec. 2.3.2.1. For better readability, we briefly state them again for the model introduced in the preceding section.

The stochastic energy is defined as

$$e[\boldsymbol{m}_{(\tau)}, t] = \sum_{\alpha} e_{\alpha}(\boldsymbol{\lambda}_t)\, \delta_{\alpha,\boldsymbol{m}_{(\tau)}(t)}, \tag{4.28}$$

and its time-derivative can be decomposed as follows,

$$\mathrm{d}_t\, e[\boldsymbol{m}_{(\tau)}, t] = \dot{q}[\boldsymbol{m}_{(\tau)}, t] + \dot{w}[\boldsymbol{m}_{(\tau)}, t], \tag{4.29}$$

with the stochastic heat and work currents

$$\begin{aligned}
\dot{q}[\boldsymbol{m}_{(\tau)}, t] &= \sum_{\nu=1}^{L}\sum_{j=1}^{M} \delta(\nu-\nu_j)\delta(t-\tau_j)\left[e_{\alpha_j}(\boldsymbol{\lambda}_{\tau_j}) - e_{\alpha_{j-1}}(\boldsymbol{\lambda}_{\tau_j}) - f^{(\nu_j)}_{\alpha_j,\alpha_{j-1}}\right] \\
&= \sum_{\nu=1}^{L} -\sum_{j=1}^{M} \delta(t-\tau_j)\delta(t-\tau_j)\,\frac{1}{\beta^{(\nu_j)}}\,\ln\frac{w^{(\nu_j)}_{\alpha_j,\alpha_{j-1}}(\boldsymbol{\lambda}_{\tau_j})}{w^{(\nu_j)}_{\alpha_{j-1},\alpha_j}(\boldsymbol{\lambda}_{\tau_j})} = \sum_{\nu=1}^{L}\dot{q}^{(\nu)}[\boldsymbol{m}_{(\tau)}, t]
\end{aligned} \tag{4.30}$$

$$\dot{w}[\boldsymbol{m}_{(\tau)}, t] = \sum_{\alpha}\left[\dot{\boldsymbol{\lambda}}_t \cdot \nabla_{\boldsymbol{\lambda}_t}\, e_{\alpha}(\boldsymbol{\lambda}_t)\right]\delta_{\alpha,\boldsymbol{m}_{(\tau)}(t)}\Big|_{\boldsymbol{m}_{(\tau)}(t)} + \sum_{\nu=1}^{L}\sum_{j=1}^{M}\delta(\nu-\nu_j)\delta(t-\tau_j)\, f^{(\nu_j)}_{\alpha_j,\alpha_{j-1}}$$

$$= \dot{w}_{\boldsymbol{\lambda}}[\boldsymbol{m}_{(\tau)}, t] + \sum_{\nu=1}^{L} \dot{w}_f^{(\nu)}[\boldsymbol{m}_{(\tau)}, t], \tag{4.31}$$

where we introduced the notation $\nabla_{\boldsymbol{\lambda}_t} = \left(\partial_{\lambda_t}, \partial_{\lambda'_t}\right)^\top$. It will be proven instrumental to split the fluctuating work current into the contribution $\dot{w}_{\boldsymbol{\lambda}}[\boldsymbol{m}_{(\tau)}, t]$ from the nonautonomous driving and the dissipative contribution $\sum_{\nu=1}^{L} \dot{w}_f^{(\nu)}[\boldsymbol{m}_{(\tau)}, t]$ due to the nonconservative forces.

Next, the stochastic system entropy is defined as follows

$$s[\boldsymbol{m}_{(\tau)}, t] = -\sum_{\alpha} \ln p_\alpha(t)\ \delta_{\alpha, \boldsymbol{m}_{(\tau)}(t)}, \tag{4.32}$$

and its time-derivative can be split as

$$\mathrm{d}_t\, s[\boldsymbol{m}_{(\tau)}, t] = \dot{s}_e[\boldsymbol{m}_{(\tau)}, t] + \dot{\sigma}[\boldsymbol{m}_{(\tau)}, t], \tag{4.33}$$

with the stochastic entropy flow

$$\dot{s}_e[\boldsymbol{m}_{(\tau)}, t] = \sum_{\nu=1}^{L} - \sum_{j=1}^{M} \delta(\nu - \nu_j)\delta(t - \tau_j)\ \ln \frac{w^{(\nu_j)}_{\alpha_j, \alpha_{j-1}}(\boldsymbol{\lambda}_{\tau_j})}{w^{(\nu_j)}_{\alpha_{j-1}, \alpha_j}(\boldsymbol{\lambda}_{\tau_j})} = \sum_{\nu=1}^{L} \beta^{(\nu)}\, \dot{q}^{(\nu)}[\boldsymbol{m}_{(\tau)}, t], \tag{4.34}$$

and the stochastic entropy production rate

$$\dot{\sigma}[\boldsymbol{m}_{(\tau)}, t] = -\left.\frac{\partial_t p_\alpha(t)}{p_\alpha(t)}\right|_{\boldsymbol{m}_{(\tau)}} + \sum_{\nu=1}^{L} \sum_{j=1}^{M} \delta(\nu - \nu_j)\delta(t - \tau_j)\ \ln \frac{w^{(\nu_j)}_{\alpha_j, \alpha_{j-1}}(\boldsymbol{\lambda}_{\tau_j})\, p_{\alpha_{j-1}}(t)}{w^{(\nu_j)}_{\alpha_{j-1}, \alpha_j}(\boldsymbol{\lambda}_{\tau_j})\, p_{\alpha_j}(t)}. \tag{4.35}$$

It will prove useful to also consider the time-integrated stochastic first law

$$\Delta e[\boldsymbol{m}_{(\tau)}, t] \equiv \sum_{\nu=1}^{L} \delta e^{(\nu)}[\boldsymbol{m}_{(\tau)}, t] = \delta q[\boldsymbol{m}_{(\tau)}, t] + \delta w[\boldsymbol{m}_{(\tau)}, t], \tag{4.36}$$

with the time-integrated fluctuating energy current

$$\delta e^{(\nu)}[\boldsymbol{m}_{(\tau)}, t] = \int_0^t \mathrm{d}t' \sum_{j=1}^{M} \delta(\nu - \nu_j)\delta(t - \tau_j)[e_{\alpha_j} - e_{\alpha_{j-1}}]. \tag{4.37}$$

and the fluctuating heat and work

$$\begin{aligned} \delta q[\boldsymbol{m}_{(\tau)}, t] &= \sum_{\nu=1}^{L} - \int_0^t \mathrm{d}t' \sum_{j=1}^{M} \delta(\nu - \nu_j)\delta(t - \tau_j)\, \frac{1}{\beta^{(\nu_j)}}\ \ln \frac{w^{(\nu_j)}_{\alpha_j, \alpha_{j-1}}(\boldsymbol{\lambda}_{\tau_j})}{w^{(\nu_j)}_{\alpha_{j-1}, \alpha_j}(\boldsymbol{\lambda}_{\tau_j})} = \\ &= \sum_{\nu=1}^{L} \underbrace{\left(\delta e^{(\nu)}[\boldsymbol{m}_{(\tau)}, t] - \delta w_f^{(\nu)}[\boldsymbol{m}_{(\tau)}, t] \right)}_{\delta q_f^{(\nu)}[\boldsymbol{m}_{(\tau)}, t]} \end{aligned} \tag{4.38}$$

$$\delta w[\boldsymbol{m}_{(\tau)},t] = \underbrace{\int_0^t \mathrm{d}t' \sum_\alpha \left[\dot{\boldsymbol{\lambda}}_{t'} \cdot \nabla_{\boldsymbol{\lambda}_{t'}} e_\alpha(\boldsymbol{\lambda}_{t'})\right] \delta_{\alpha,\boldsymbol{m}_{(\tau)}(t')}}_{\delta w_{\boldsymbol{\lambda}}[\boldsymbol{m}_{(\tau)},t]} + \sum_{\nu=1}^{L} \underbrace{\int_0^t \mathrm{d}t' \sum_{j=1}^{M} \delta(\nu-\nu_j)\delta(t-\tau_j) \boldsymbol{f}^{(\nu)}_{\alpha_j,\alpha_{j-1}}}_{\delta w_f^{(\nu)}[\boldsymbol{m}_{(\tau)},t]}. \tag{4.39}$$

Using Eqs. (4.35) and (4.38), the entropy production can be written as follows

$$\begin{aligned}\delta\sigma[\boldsymbol{m}_{(\tau)},t] &= -\ln\frac{p_{\alpha_M}(t)}{p_{\alpha_0}(0)} + \sum_{\nu=1}^{L}\int_0^t \mathrm{d}t' \sum_{j=1}^{M} \delta(\nu-\nu_j)\delta(t'-\tau_j)\,\ln\frac{w^{(\nu_j)}_{\alpha_j,\alpha_{j-1}}(\boldsymbol{\lambda}_{\tau_j})}{w^{(\nu_j)}_{\alpha_{j-1},\alpha_j}(\boldsymbol{\lambda}_{\tau_j})} \\ &= -\ln\frac{p_{\alpha_M}(t)}{p_{\alpha_0}(0)} - \sum_{\nu=1}^{L}\beta^{(\nu)}\delta q^{(\nu)}[\boldsymbol{m}_{(\tau)},t].\end{aligned} \tag{4.40}$$

4.2.2.2 Generating Function Techniques Revisited

Microscopic Description

In the preceding section we introduced all relevant fluctuating thermodynamic quantities and now use the generating function techniques presented in Sec. 2.3.2.2 to determine if their statistics is invariant under the dynamically exact coarse-graining in Eq. (4.14). To this end, we first recall the equations of motion (2.85) for the generating functions associated with the change $\delta o[\boldsymbol{m}_{(\tau)},t]$ of the fluctuating microscopic observable o along a trajectory $\boldsymbol{m}_{(\tau)}$ conditioned to be in a microstate α at time t,

$$\dot{g}_\alpha(\gamma_o,t) = \sum_{\alpha'} w_{\alpha\alpha'}(\gamma_o,\boldsymbol{\lambda}_t)\, g_{\alpha'}(\gamma_o,t), \tag{4.41}$$

with the biased microscopic transition rates

$$w_{\alpha\alpha'}(\gamma_o,\boldsymbol{\lambda}_t) = -\gamma_o\,\dot{o}_\alpha\,\delta_{\alpha,\alpha'} + \sum_{\nu=1}^{L} \mathrm{e}^{-\gamma_o\, o^{(\nu)}_{\alpha\alpha'}(\boldsymbol{\lambda}_t)}\, w^{(\nu)}_{\alpha\alpha'}(\boldsymbol{\lambda}_t). \tag{4.42}$$

For state functions, $\delta o[\boldsymbol{m}_{(\tau)},t] = \Delta o[\boldsymbol{m}_{(\tau)},t] = o[\boldsymbol{m}_{(\tau)},t] - o[\boldsymbol{m}_{(\tau)},0]$, the last equation reduces to [cf. Eq. (2.86)]

$$g(\gamma_o,t) = \sum_{\alpha,\alpha'} \mathrm{e}^{-\gamma_o\,[o_\alpha(t)-o_{\alpha'}(0)]}\, p_\alpha(t)\, p_{\alpha'}(0). \tag{4.43}$$

Using Eqs. (4.28) and (4.32), we have for the microscopic generating functions associated with the stochastic state-like observables energy and entropy,

$$g(\gamma_e,t) = \sum_{\alpha,\alpha'} \mathrm{e}^{-\gamma_e\,[e_\alpha(\boldsymbol{\lambda}_t)-e_{\alpha'}(\boldsymbol{\lambda}_0)]}\, p_\alpha(t)\, p_{\alpha'}(0) \tag{4.44}$$

$$g(\gamma_s, t) = \sum_{\alpha,\alpha'} \mathrm{e}^{\gamma_s \ln \frac{p_\alpha(t)}{\ln p_{\alpha'}(0)}}\, p_\alpha(t)\, p_{\alpha'}(0). \tag{4.45}$$

Moreover, substituting Eqs. (4.30), (4.31), (4.34) and (4.35) into Eq. (4.41), we obtain for the microscopic generating functions associated with the currents

$$\dot{g}_\alpha(\gamma_q, t) = \sum_{\nu=1}^{L} \sum_{\alpha'} \mathrm{e}^{-\gamma_q \left[e_\alpha(\boldsymbol{\lambda}_t) - e_{\alpha'}(\boldsymbol{\lambda}_t) - f^{(\nu)}_{\alpha\alpha'}\right]}\; w^{(\nu)}_{\alpha\alpha'}(\boldsymbol{\lambda}_t)\; g_{\alpha'}(\gamma_q, t) \tag{4.46}$$

$$\dot{g}_\alpha(\gamma_w, t) = -\gamma_w\, \dot{\boldsymbol{\lambda}}_t \cdot \left[\nabla_{\boldsymbol{\lambda}_t}\, e_\alpha(\boldsymbol{\lambda}_t)\right]\; g_\alpha(\gamma_w, t) + \sum_{\nu=1}^{L} \sum_{\alpha'} \mathrm{e}^{-\gamma_w f^{(\nu)}_{\alpha\alpha'}}\; w^{(\nu)}_{\alpha\alpha'}(\boldsymbol{\lambda}_t)\; g_{\alpha'}(\gamma_w, t) \tag{4.47}$$

$$\dot{g}_\alpha(\gamma_{s_e}, t) = \sum_{\nu=1}^{L} \sum_{\alpha'} \mathrm{e}^{\gamma_{s_e} \ln \frac{w^{(\nu)}_{\alpha\alpha'}(\boldsymbol{\lambda}_t)}{\ln w^{(\nu)}_{\alpha'\alpha}(\boldsymbol{\lambda}_t)}}\; w^{(\nu)}_{\alpha\alpha'}(\boldsymbol{\lambda}_t)\; g_{\alpha'}(\gamma_{s_e}, t) \tag{4.48}$$

$$\dot{g}_\alpha(\gamma_\sigma, t) = \gamma_\sigma\, \frac{\partial_t p_\alpha(t)}{p_\alpha(t)}\, g_\alpha(\gamma_\sigma, t) + \sum_{\nu=1}^{L} \sum_{\alpha'} \mathrm{e}^{-\gamma_\sigma \ln \frac{w^{(\nu)}_{\alpha\alpha'}(\boldsymbol{\lambda}_t) p_{\alpha'}(t)}{w^{(\nu)}_{\alpha'\alpha}(\boldsymbol{\lambda}_t) p_\alpha(t)}}\; w^{(\nu)}_{\alpha\alpha'}(\boldsymbol{\lambda}_t)\; g_{\alpha'}(\gamma_\sigma, t). \tag{4.49}$$

Mesoscopic Description

We rewrite the microscopic generating function (2.73) as follows

$$g_{\alpha_{\boldsymbol{N}}}(\gamma_o, t) = P_{\boldsymbol{N}}(t)\, \mathbb{P}_{\alpha_{\boldsymbol{N}}}(t)\, \langle \mathrm{e}^{-\gamma_o\, \delta o[\boldsymbol{m}_{(\tau)}, t]} \rangle_{\alpha_{\boldsymbol{N}}}, \tag{4.50}$$

and define the mesososcopic generating function

$$G_{\boldsymbol{N}}(\gamma_O, t) \equiv \sum_{\alpha_{\boldsymbol{N}}} g_{\alpha_{\boldsymbol{N}}}(\gamma_o, t) = P_{\boldsymbol{N}}(t) \sum_{\alpha_{\boldsymbol{N}}} \mathbb{P}_{\alpha_{\boldsymbol{N}}}(t)\, \langle \mathrm{e}^{-\gamma_o\, \delta o[\boldsymbol{m}_{(\tau)}, t]} \rangle_{\alpha_{\boldsymbol{N}}}, \tag{4.51}$$

where $\langle\cdot\rangle_{\alpha_{\boldsymbol{N}}}$ and $\langle\cdot\rangle_{\boldsymbol{N}}$ denote ensemble averages over all trajectories that are in the microstate α belonging to a given mesostate $\boldsymbol{N}$ and over all those that are in mesostate $\boldsymbol{N}$ at time t, respectively. Moreover, O denotes a mesoscopic observable defined along a trajectory in propagating in the mesoscopic state space, $\boldsymbol{M}_{(\tau)}$, which we write as $O[\boldsymbol{M}_{(\tau)}, t]$ in the following. Since the trajectory observables $o = e, q, w, s_e$ do not depend on microscopic information [cf. Eqs. (4.28)–(4.31)], we have for those observables that $o[\boldsymbol{m}_{(\tau)}, t] = O[\boldsymbol{M}_{(\tau)}, t]$ and Eq. (4.51) closes as follows

$$\begin{aligned} G_{\boldsymbol{N}}(\gamma_O, t) &= P_{\boldsymbol{N}}(t) \sum_{\alpha_{\boldsymbol{N}}} \mathbb{P}_{\alpha_{\boldsymbol{N}}}(t)\, \langle \mathrm{e}^{-\gamma_O\, \delta O[\boldsymbol{M}_{(\tau)}, t]} \rangle_{\boldsymbol{N}} \\ &= P_{\boldsymbol{N}}(t)\, \langle \mathrm{e}^{-\gamma_O\, \delta O[\boldsymbol{M}_{(\tau)}, t]} \rangle_{\boldsymbol{N}}, \quad O = E, Q, W, S_e. \end{aligned} \tag{4.52}$$

Thus, the microscopic generating function for the energy (4.44) in mesoscopic representation reads

$$G(\gamma_E, t) = \sum_{\boldsymbol{N},\boldsymbol{N}'} \mathrm{e}^{-\gamma_E\, [E_{\boldsymbol{N}}(\boldsymbol{\lambda}_t) - E_{\boldsymbol{N}'}(\boldsymbol{\lambda}_0)]}\, P_{\boldsymbol{N}}(t)\, P_{\boldsymbol{N}'}(0) = g(\gamma_e, t), \tag{4.53}$$

and from the microscopic equation of motion for the generating function (4.41) we get

$$\dot{G}_{\boldsymbol{N}}(\gamma_O, t) = \sum_{\boldsymbol{N}'} W_{\boldsymbol{N}\boldsymbol{N}'}(\gamma_O, \boldsymbol{\lambda}_t)\, G_{\boldsymbol{N}'}(\gamma_O, t), \tag{4.54}$$

with the mesoscopic biased generator

$$W_{\boldsymbol{N}\boldsymbol{N}'}(\gamma_O, \boldsymbol{\lambda}_t) = -\gamma_O \dot{O}_{\boldsymbol{N}}(\boldsymbol{\lambda}_t)\, \delta_{\boldsymbol{N},\boldsymbol{N}'} + \sum_{\nu=1}^{L} \mathrm{e}^{-\gamma_O\, O^{(\nu)}_{\boldsymbol{N},\boldsymbol{N}'}(\boldsymbol{\lambda}_t)}\, W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t), \tag{4.55}$$

for $O = E, Q, W, S_e$. More explicitly, Eqs. (4.46), (4.47) and (4.48) can be rewritten in mesoscopic representation as follows

$$\dot{G}_{\boldsymbol{N}}(\gamma_Q, t) = \sum_{\nu=1}^{L} \sum_{\boldsymbol{N}'} \mathrm{e}^{-\gamma_Q \left[E_{\boldsymbol{N}}(\boldsymbol{\lambda}_t) - E_{\boldsymbol{N}'}(\boldsymbol{\lambda}_t) - f^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}\right]}\, W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t)\, G_{\boldsymbol{N}'}(\gamma_Q, t) \tag{4.56}$$

$$\dot{G}_{\boldsymbol{N}}(\gamma_W, t) = -\gamma_W\, \dot{\boldsymbol{\lambda}}_t \cdot \left[\nabla_{\boldsymbol{\lambda}_t} E_{\boldsymbol{N}}(\boldsymbol{\lambda}_t)\right] G_{\boldsymbol{N}}(\gamma_W, t) + \sum_{\nu=1}^{L} \sum_{\boldsymbol{N}'} \mathrm{e}^{-\gamma_W f^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}}\, W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t)\, G_{\boldsymbol{N}'}(\gamma_W, t) \tag{4.57}$$

$$\dot{G}_{\boldsymbol{N}}(\gamma_{S_e}, t) = \sum_{\nu=1}^{L} \sum_{\boldsymbol{N}'} \mathrm{e}^{\gamma_{S_e}\left[\ln \frac{W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t)}{W^{(\nu)}_{\boldsymbol{N}'\boldsymbol{N}}(\boldsymbol{\lambda}_t)} - \left(S^{int}_{\boldsymbol{N}} - S^{int}_{\boldsymbol{N}'}\right)\right]}\, W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t)\, G_{\boldsymbol{N}'}(\gamma_{S_e}, t). \tag{4.58}$$

It is easy to verify that $\sum_{\boldsymbol{N}} \dot{G}_{\boldsymbol{N}}(\gamma_O, t) = \sum_\alpha \dot{g}_\alpha(\gamma_o, t)$ for $O = E, Q, W, S_e$ and $o = e, q, w, s_e$. Thus, we find that the statistics of the stochastic first law in microscopic representation (4.29) is invariant under coarse-graining.

Conversely, the stochastic system entropy (4.32) and stochastic entropy production rate (4.35) are functions of the microscopic ensemble probability. The corresponding equation for the mesoscopic generating function (4.51) would, in general, not be closed and the stochastic entropy balance in microscopic representation (4.33) is, in general, not invariant under the coarse-graining. Though, there are two generic cases for which an exact coarse-graining is possible.

First, for the choice of a microscopic initial condition, $p^*_{\alpha'}(0) = P^*_{\boldsymbol{N}'}(0)/\Omega_{\boldsymbol{N}}$, where all microstates are uniformly distributed inside the respective mesostates according to Eq. (4.27). The local equilibrium is preserved at all times since the Hamiltonian (4.2) and thus the microscopic transition rates (4.4) do not discriminate between the equienergetic microstates inside the mesostate. For such an initial condition, the mesoscopic generating functions associated with the system entropy and entropy production rate read, respectively

$$G(\gamma_S, t) = \sum_{\boldsymbol{N},\boldsymbol{N}'} \mathrm{e}^{\gamma_S\left[\ln \frac{P_{\boldsymbol{N}}(t)}{P^*_{\boldsymbol{N}'}(0)} - \left(S^{int}_{\boldsymbol{N}} - S^{int}_{\boldsymbol{N}'}\right)\right]}\, P_{\boldsymbol{N}}(t) P^*_{\boldsymbol{N}'}(0) = g(\gamma_s, t) \tag{4.59}$$

$$\dot{G}_{\boldsymbol{N}}(\gamma_\Sigma, t) = \gamma_\Sigma \frac{\partial_t P_{\boldsymbol{N}}(t)}{P_{\boldsymbol{N}}(t)}\, G_{\boldsymbol{N}}(\gamma_\Sigma, t) + \sum_{\nu=1}^{L} \sum_{\boldsymbol{N}'} \mathrm{e}^{-\gamma_\Sigma \ln \frac{W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t) P_{\boldsymbol{N}'}(t)}{W^{(\nu)}_{\boldsymbol{N}'\boldsymbol{N}}(\boldsymbol{\lambda}_t) P_{\boldsymbol{N}}(t)}}\, W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t)\, G_{\boldsymbol{N}'}(\gamma_\Sigma, t), \tag{4.60}$$

with $\sum_{\boldsymbol{N}} \dot{G}_{\boldsymbol{N}}(\gamma_\Sigma, t) = \sum_\alpha \dot{g}_\alpha(\gamma_\sigma, t)$.

Secondly, according to Eq. (4.27), if the mesoscopic system is in a stationary state the microscopic probabilities inside a mesostate are also stationary and thus uniformly distributed regardless of any possible nonuniform distribution at intial times. Consequently, the stationary mesoscopic generating functions associated with the system entropy and entropy production rate become, respectively

$$G^s(\gamma_S, \boldsymbol{\lambda}) = \sum_{\boldsymbol{N},\boldsymbol{N}'} \mathrm{e}^{\gamma_S \left[\ln \frac{P^s_{\boldsymbol{N}}(\boldsymbol{\lambda})}{P^s_{\boldsymbol{N}'}(\boldsymbol{\lambda})} - \left(S^{int}_{\boldsymbol{N}} - S^{int}_{\boldsymbol{N}'}\right) \right]} P^s_{\boldsymbol{N}}(\boldsymbol{\lambda}) P^s_{\boldsymbol{N}'}(\boldsymbol{\lambda}) = g^s(\gamma_s, \boldsymbol{\lambda}) \tag{4.61}$$

$$\dot{G}^s_{\boldsymbol{N}}(\gamma_\Sigma, \boldsymbol{\lambda}) = \sum_{\nu=1}^{L} \sum_{\boldsymbol{N}'} \mathrm{e}^{-\gamma_\Sigma \ln \frac{W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}) P^s_{\boldsymbol{N}'}(\boldsymbol{\lambda})}{\ln W^{(\nu)}_{\boldsymbol{N}'\boldsymbol{N}}(\boldsymbol{\lambda}) P^s_{\boldsymbol{N}}(\boldsymbol{\lambda})}} W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda})\, G^s_{\boldsymbol{N}'}(\gamma_\Sigma, \boldsymbol{\lambda}), \tag{4.62}$$

with $\sum_{\boldsymbol{N}} \dot{G}^s_{\boldsymbol{N}}(\gamma_\Sigma, \boldsymbol{\lambda}) = \sum_\alpha \dot{g}^s_\alpha(\gamma_\sigma, \boldsymbol{\lambda})$. Hence we conclude that the statistics of the stochastic entropy balance (4.33) is invariant under the coarse-graining, if one considers initial conditions which are uniform within each mesostate, or for systems in stationary states. In fact, Eqs. (4.27), (4.59) and (4.60) represent a potential strategy to infer the entropy fluctuations in the mesoscopic state space at finite-time: Before starting the actual measurement, the non-autonomous driving is switched off and the system is reaching a unique stationary state. The system can now be non-autonomously driven out of its steady state during the measurement while the finite-time expressions for the mesoscopic generating functions in Eqs. (4.59) and (4.60) are still valid.

Comparing Eqs. (4.44), (4.46) and (4.47) with Eqs. (4.53), (4.56) and (4.57), we note that the evolution of the generating functions associated with the first-law observables, that is energy, heat and work, have the same form in microscopic and mesoscopic representation. In contrast, the mesoscopic generating functions associated with the entropies do not have the same form as the microscopic ones but also contain the internal entropy S^{int}. This is due to the coarse-grained degrees of freedom that give rise to Boltzmann entropies (4.19) assigned to the mesostates. Physically, the conditions for the invariance of the stochastic entropy balance [Eqs. (4.59) and (4.60) or (4.61) and (4.62)] can be understood as follows. If the microscopic degrees of freedom inside the mesostates are not equiprobable, there are microscopic currents that can not be grasped at the mesoscopic level and which only vanish identically if uniform probability distributions inside the mesostates are achieved.

So far, we have established two descriptions of the stochastic thermodynamics at the microscopic and mesoscopic level. These two formulations are equivalent for the stochastic first law. In case of the stochastic entropy balance, the microscopic and mesoscopic thermodynamics coincide under the condition that the microstates inside each mesostate are equiprobable. The thermodynamics consistency at each level is ensured by the respective local detailed balance conditions in Eqs. (4.5) and (4.17). Alternatively, the thermodynamic consistency is also encoded by the so-called detailed fluctuation theorem for the stochastic entropy production derived in Eq (2.105). In the following, we will discuss this symmetry of the fluctuations of the entropy production across scales as it will be of importance further below.

4.2.2.3 Detailed Fluctuation Theorems Across Scales

Let us consider a forward and backward process as depicted in Fig. 2.4. In this case, the system starts from a state that is at equilibrium with respect to the reference reservoir $\nu = 1$,

$$p^{eq}_{\alpha_0}(\boldsymbol{\lambda}_0) = \mathrm{e}^{-\beta^{(1)}[e_{\alpha_0}(\boldsymbol{\lambda}_0) - a^{eq}(\boldsymbol{\lambda}_0)]}\,. \tag{4.63}$$

The system then evolves under the driven microscopic Markov process according to the forward protocol $\boldsymbol{\lambda}_{t'}$, $t' \in [0, t]$. For the backward process the system is initially prepared in the final equilibrium state of the forward process

$$p^{eq}_{\alpha_M}(\boldsymbol{\lambda}_t) = \mathrm{e}^{-\beta^{(1)}[e_{\alpha_M}(\boldsymbol{\lambda}_t) - a^{eq}(\boldsymbol{\lambda}_t)]}\,, \tag{4.64}$$

and subsequently evolves under the time-reversed driven microscopic Markov process according to the backward protocol $\tilde{\boldsymbol{\lambda}}_{t'} = \boldsymbol{\lambda}_{t-t'}$, $t' \in [0, t]$. Then, according to Eq. (2.108), the following symmetry of the microscopic generating function holds

$$g\big(\gamma_\lambda, \{\gamma_f^{(\nu)}\}, \{\gamma_e^{(\nu)}\}, t\big) = \tilde{g}\big(1-\gamma_\lambda, \{1-\gamma_f^{(\nu)}\}, \{1-\gamma_e^{(\nu)}\}, t\big)\ \mathrm{e}^{-\beta^{(1)}\Delta a_1^{eq}(\boldsymbol{\lambda})}, \tag{4.65}$$

which implies the microscopic finite-time detailed fluctuation theorem

$$\frac{p\Big(\beta^{(1)}\delta w_{\boldsymbol{\lambda}}\,,\ \{\delta j_f^{(\nu)}\}\,,\ \{\delta j_e^{(\nu)}\}\Big)}{\tilde{p}\Big(-\beta^{(1)}\delta w_{\boldsymbol{\lambda}}\,,\ -\{\delta j_f^{(\nu)}\}\,,\ -\{\delta j_e^{(\nu)}\}\Big)} = \mathrm{e}^{\beta^{(1)}\left[\delta w_{\boldsymbol{\lambda}} - \Delta a_1^{eq}\right] + \sum_{\nu=1}^{L}\left[\beta^{(\nu)}\delta w_f^{(\nu)} + \left[\beta^{(1)} - \beta^{(\nu)}\right]\delta e^{(\nu)}\right]}, \tag{4.66}$$

where we used the compact notation for the microscopic time-integrated currents introduced in Eq. 2.106. Moreover, $\Delta a_1^{eq} = a_1^{eq}(\boldsymbol{\lambda}_t) - a_1^{eq}(\boldsymbol{\lambda}_0)$ refers to the change in global microscopic equilibrium free energy with respect to the reservoir $\nu = 1$ along the forward process that only depends on the initial and final value of the driving protocol and thus does not fluctuate.

Analogously, we can define the forward and backward process as define above also in the mesoscopic state space. In this case, the equilibrum distributions for the forward and, in reversed order for the backward trajectory, read, respectively

$$P^{eq}_{\boldsymbol{N}_0}(\boldsymbol{\lambda}_0) = \mathrm{e}^{-\beta^{(1)}[A_{N_0}(\boldsymbol{\lambda}_0) - A^{eq}(\boldsymbol{\lambda}_0)]} \tag{4.67}$$

$$P^{eq}_{\boldsymbol{N}_M}(\boldsymbol{\lambda}_t) = \mathrm{e}^{-\beta^{(1)}[A_{N_M}(\boldsymbol{\lambda}_t) - A^{eq}(\boldsymbol{\lambda}_t)]}\,. \tag{4.68}$$

Crucially, all fluctuating quantities appearing in the microscopic detailed fluctuation theorem (4.66) are invariant under the dynamically exact coarse-graining (4.14). Consequently, the symmetry for the microscopic generating function (4.65) is also exhibited at the mesoscopic level,

$$G(\boldsymbol{\gamma}, t) = \tilde{G}(\tilde{\boldsymbol{\gamma}}, t)\ \mathrm{e}^{-\beta^{(1)}\Delta A_1^{eq}(\boldsymbol{\lambda})}, \tag{4.69}$$

where $\Delta A_1^{eq} = A_1^{eq}(\boldsymbol{\lambda}_t) - A_1^{eq}(\boldsymbol{\lambda}_0)$. Moreover, for brevity we introduced the notation

$$\boldsymbol{\gamma} \equiv \gamma_{\Lambda}, \{\gamma_F^{(\nu)}\}, \{\gamma_E^{(\nu)}\}, \quad \tilde{\boldsymbol{\gamma}} \equiv 1 - \gamma_{\Lambda}, \{1 - \gamma_F^{(\nu)}\}, \{1 - \gamma_E^{(\nu)}\}, \tag{4.70}$$

where $\{\delta J_{\boldsymbol{F}}^{(\nu)}\} \equiv \big(\beta^{(1)}\delta W_f^{(1)}, \ldots, \beta^{(L)}\delta W_f^{(L)}\big)$ is the mesoscopic time-integrated autonomous work currents and $\{\delta J_E^{(\nu)}\} \equiv \big([\beta^{(1)} - \beta^{(2)}]\delta E^{(2)}, \ldots, [\beta^{(1)} - \beta^{(L)}]\delta E^{(L)}\big)$ is the mesoscopic time-integrated energy currents. Thus, the detailed fluctuation theorem (4.71) also holds at the mesoscopic level,

$$\frac{P\Big(\beta^{(1)}\delta W_{\boldsymbol{\lambda}}\,,\,\{\delta J_{\boldsymbol{F}}^{(\nu)}\}\,,\,\{\delta J_E^{(\nu)}\}\Big)}{\tilde{P}\Big(-\beta^{(1)}\delta W_{\boldsymbol{\lambda}}\,,\,-\{\delta J_{\boldsymbol{F}}^{(\nu)}\}\,,\,-\{\delta J_E^{(\nu)}\}\Big)} = \mathrm{e}^{\beta^{(1)}[\delta W_{\boldsymbol{\lambda}} - \Delta A_1^{eq}] + \sum_{\nu=1}^{L}\left\{\beta^{(\nu)}\delta W_f^{(\nu)} + [\beta^{(1)} - \beta^{(\nu)}]\delta E^{(\nu)}\right\}}, \tag{4.71}$$

where $P\Big(\beta^{(1)}\delta W_{\boldsymbol{\lambda}}\,,\,\{\delta J_F^{(\nu)}\}\,,\,\{\delta J_E^{(\nu)}\}\Big)$ is the probability to observe a mesoscopic nonautonomous work $\beta^{(1)}\delta W_{\boldsymbol{\lambda}}$, the mesoscopic time-integrated autonomous work currents $\{\delta J_{\boldsymbol{F}}^{(\nu)}\}$ and the mesoscopic time-integrated energy currents $\{\delta J_E^{(\nu)}\}$ along the forward process in the mesoscopic state space. Conversely, $\tilde{P}\Big(-\beta^{(1)}\delta W_{\boldsymbol{\lambda}}\,,\,\{-\delta J_{\boldsymbol{F}}^{(\nu)}\}\,,\,\{-\delta J_E^{(\nu)}\}\Big)$ is the probability to observe a mesoscopic nonautonomous work $-\beta^{(1)}\delta W_{\boldsymbol{\lambda}}$, the mesoscopic time-integrated autonomous work currents $\{-\delta J_F^{(\nu)}\}$ and the mesoscopic time-integrated energy currents $\{-\delta J_E^{(\nu)}\}$ along the time-reversed backward process in the mesoscopic state space.

We want to stress that the existence of the detailed fluctuation theorem for the entropy production across scales (4.66), (4.71) ensures that the thermodynamics formulated at each of these levels is consistent. We will make use of this result further below when we formulate the fluctuations at the macroscopic level, that is fluctuations that scale exponentially with the system size N.

4.2.2.4 Microscopic And Mesoscopic First And Second Law

Before turning to the macroscopic limit, for completeness, we want to formulate the thermodynamics at the ensemble level on microscopic and mesoscopic scales and hereby, because of their importance, focus on the laws of thermodynamics. Using Eq. (2.76) and Eqs. (4.44) – (4.47) or Eqs. (4.53) – (4.57), we arrive at the microscopic or mesoscopic first law of thermodynamics, respectively,

$$\mathrm{d}_t e(t) = \dot{q}(t) + \dot{w}(t), \quad \mathrm{d}_t E(t) = \dot{Q}(t) + \dot{W}(t), \tag{4.72}$$

with the average internal energy that is equivalent at microscopic and mesoscopic scale,

$$e(t) = \sum_{\alpha} e_{\alpha}(\boldsymbol{\lambda}_t)\, p_{\alpha}(t) = \sum_{\boldsymbol{N}} E_{\boldsymbol{N}}(\boldsymbol{\lambda}_t)\, P_{\boldsymbol{N}}(t) = E(t). \tag{4.73}$$

and with the equivalent microscopic and mesoscopic heat currents

$$\begin{aligned}\dot{q}(t) &= \sum_{\nu=1}^{L} \sum_{\alpha,\alpha'} \left[e_\alpha(\boldsymbol{\lambda}_t) - e_{\alpha'}(\boldsymbol{\lambda}_t) - f^{(\nu)}_{\alpha\alpha'} \right] \; w^{(\nu)}_{\alpha\alpha'}(\boldsymbol{\lambda}_t)\, p_{\alpha'}(t) \\ &= \sum_{\nu=1}^{L} \sum_{\boldsymbol{N},\boldsymbol{N}'} \left[E_{\boldsymbol{N}}(\boldsymbol{\lambda}_t) - E_{\boldsymbol{N}'}(\boldsymbol{\lambda}_t) - f^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'} \right] \; W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t)\, P_{\boldsymbol{N}'}(t) = \dot{Q}(t)\end{aligned} \tag{4.74}$$

as well as the equivalent microscopic and mesoscopic work currents

$$\begin{aligned}\dot{w}(t) &= \sum_{\alpha} \left[\dot{\boldsymbol{\lambda}}_t \cdot \left[\nabla_{\boldsymbol{\lambda}_t} E_\alpha(\boldsymbol{\lambda}_t) \right] \; p_\alpha(t) + \sum_{\nu=1}^{L} \sum_{\alpha'} f^{(\nu)}_{\alpha\alpha'} \; w^{(\nu)}_{\alpha,\alpha'}(\boldsymbol{\lambda}_t)\, p_{\alpha'}(t) \right] \\ &= \sum_{\boldsymbol{N}} \left[\dot{\boldsymbol{\lambda}}_t \cdot \left[\nabla_{\boldsymbol{\lambda}_t} E_{\boldsymbol{N}}(\boldsymbol{\lambda}_t) \right] P_{\boldsymbol{N}}(t) + \sum_{\nu=1}^{L} \sum_{\boldsymbol{N}'} f^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'} \, W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t)\, P_{\boldsymbol{N}'}(t) \right] = \dot{W}(t).\end{aligned} \tag{4.75}$$

Next, with Eq. (4.59) or (4.61), respectively, we find the equivalence of the average system entropy at microscopic and mesoscopic scale,

$$s(t) = -\sum_{\alpha} p_\alpha(t) \, \ln p_\alpha(t) = \sum_{\boldsymbol{N}} \left[S^{int}_{\boldsymbol{N}} - \ln P_{\boldsymbol{N}}(t) \right] \; P_{\boldsymbol{N}}(t) = S(t) \tag{4.76}$$

$$s^s(\boldsymbol{\lambda}) = -\sum_{\alpha} p^s_\alpha(\boldsymbol{\lambda}) \, \ln p^s_\alpha(\boldsymbol{\lambda}) = \sum_{\boldsymbol{N}} \left[S^{int}_{\boldsymbol{N}} - \ln P^s_{\boldsymbol{N}}(\boldsymbol{\lambda}) \right] \; P^s_{\boldsymbol{N}}(\boldsymbol{\lambda}) = S^s(\boldsymbol{\lambda}). \tag{4.77}$$

Using furthermore Eq. (4.60) or (4.62), respectively, the microscopic and mesoscopic second law of thermodynamics reads

$$\begin{aligned}\dot{\sigma}(t) &= \sum_{\nu=1}^{L} \sum_{\alpha,\alpha'} \ln \frac{w^{(\nu)}_{\alpha\alpha'}(\boldsymbol{\lambda}_t)\, p_{\alpha'}(t)}{w^{(\nu)}_{\alpha'\alpha}(\boldsymbol{\lambda}_t)\, p_\alpha(t)} \; w^{(\nu)}_{\alpha\alpha'}(\boldsymbol{\lambda}_t)\, p_{\alpha'}(t) \\ &= \sum_{\nu=1}^{L} \sum_{\boldsymbol{N},\boldsymbol{N}'} \ln \frac{W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t)\, P_{\boldsymbol{N}'}(t)}{W^{(\nu)}_{\boldsymbol{N}'\boldsymbol{N}}(\boldsymbol{\lambda}_t)\, P_{\boldsymbol{N}}(t)} \; W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda}_t)\, P_{\boldsymbol{N}'}(t) = \dot{\Sigma}(t) \geq 0\end{aligned} \tag{4.78}$$

$$\begin{aligned}\dot{\sigma}^s(\boldsymbol{\lambda}) &= \sum_{\nu=1}^{L} \sum_{\alpha,\alpha'} \ln \frac{w^{(\nu)}_{\alpha\alpha'}(\boldsymbol{\lambda})\, p^s_{\alpha'}(\boldsymbol{\lambda})}{w^{(\nu)}_{\alpha'\alpha}(\boldsymbol{\lambda})\, p^s_\alpha(\boldsymbol{\lambda})} \; w^{(\nu)}_{\alpha\alpha'}(\boldsymbol{\lambda})\, p^s_{\alpha'}(\boldsymbol{\lambda}) \\ &= \sum_{\nu=1}^{L} \sum_{\boldsymbol{N},\boldsymbol{N}'} \ln \frac{W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda})\, P^s_{\boldsymbol{N}'}(\boldsymbol{\lambda})}{W^{(\nu)}_{\boldsymbol{N}'\boldsymbol{N}}(\boldsymbol{\lambda})\, P^s_{\boldsymbol{N}}(\boldsymbol{\lambda})} \; W^{(\nu)}_{\boldsymbol{N}\boldsymbol{N}'}(\boldsymbol{\lambda})\, P^s_{\boldsymbol{N}'}(\boldsymbol{\lambda}) = \dot{\Sigma}^s(\boldsymbol{\lambda}) \geq 0.\end{aligned} \tag{4.79}$$

4.2.3 Macroscopic Theory

4.2.3.1 Macroscopic Fluctuations

Thus far, we have established two equivalent representations of the stochastic dynamics above, the microscopic and mesoscopic representation. We furthermore identified the conditions under which the thermodynamics at these levels coincide. In this section, the question of how to infer the fluctuations in the macroscopic limit, $N \to \infty$, will be addressed. To shed light on this question, we will employ the Martin-Siggia-Rose formalism [164, 165] which equivalently represents the Markovian jump process via a path integral. As will be demonstrated in the following, this path-integral formalism allows to establish a fluctuating description valid at macroscopic scales in the large deviation sense [163], that is for fluctuations that scale exponentially with the number of units N.

For better readability, we defer a detailed presentation of the elementary concepts underlying the construction of the path integral to appendix A.6. The mesoscopic generating function $G(\gamma_O, t)$ associated with a mesoscopic stochastic observable $O[\boldsymbol{M}_{(\tau)}, t]$ within the path integral representation generically reads

$$\begin{aligned} G(\gamma_O, t) &= \int \mathcal{D}[\boldsymbol{N}] \int \mathcal{D}[\boldsymbol{\pi}]\, \mathrm{e}^{\int_0^t \mathrm{d}t' \left[-\boldsymbol{\pi}(t')\cdot\dot{\boldsymbol{N}}(t') + H_{\gamma_O}[\boldsymbol{N}(t'),\boldsymbol{\pi}(t')] - \gamma_O \dot{\boldsymbol{\lambda}}_{t'} \cdot \left[\nabla_{\boldsymbol{\lambda}_{t'}} O_{\boldsymbol{N}(t')}\right] - \gamma_O \mathrm{d}_{t'} O_{\boldsymbol{N}(t')} \right]} P_{\boldsymbol{N}}(0) \\ &\equiv \int \mathcal{D}[\boldsymbol{N}] \int \mathcal{D}[\boldsymbol{\pi}]\, \mathrm{e}^{\mathcal{L}_{\gamma_O}[\boldsymbol{N}(t'),\boldsymbol{\pi}(t')]}, \end{aligned} \tag{4.80}$$

where $\mathcal{D}[X]$ denotes the path-integral measure for the function $\boldsymbol{X}$. The quantity $\boldsymbol{\pi}$ is the conjugated field and can be physically interpreted as the instantaneous counting field for variations in the mesostates $\mathrm{d}\boldsymbol{N}$. Moreover, the biased action functional $\mathcal{L}_{\gamma_O}[\boldsymbol{N}, \boldsymbol{\pi}]$ consists of the kinetic term $-\boldsymbol{\pi}\,\dot{\boldsymbol{N}}$, the biased Hamiltonian that accounts for the current-like contributions to $G(\gamma_O, t)$,

$$H_{\gamma_O}[\boldsymbol{N}(t'), \boldsymbol{\pi}(t')] = \sum_{\nu=1}^{L} \sum_{i,j=1}^{q} \left[\mathrm{e}^{\pi_i(t') - \pi_j(t')}\, \mathrm{e}^{-\gamma_O O_{ij}^{(\nu)}(\boldsymbol{N}(t'))} - 1 \right] W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'}, \boldsymbol{N}(t')), \tag{4.81}$$

of a contribution due to the nonautonomous driving

$$-\gamma_O \int_0^t \mathrm{d}t'\, \dot{\boldsymbol{\lambda}}_{t'} \cdot \left[\nabla_{\boldsymbol{\lambda}_{t'}} O_{\boldsymbol{N}(t')}\right], \tag{4.82}$$

of a state-like contribution

$$-\gamma_O \int_0^t \mathrm{d}t'\, \mathrm{d}_{t'} O_{\boldsymbol{N}}(t') = -\gamma_O [O_{\boldsymbol{N}_M}(t) - O_{\boldsymbol{N}_0}(0)], \tag{4.83}$$

and of the initial condition $\ln P_{\boldsymbol{N}}(0)$. The quantities $O_{ij}^{(\nu)}(\boldsymbol{N})$ and $W_{ij}^{(\nu)}(\boldsymbol{\lambda}, \boldsymbol{N})$ are the change of the fluctuating mesoscopic observable O and the mesoscopic transition rate, respectively, along a jump of the trajectory away from the mesostate $\boldsymbol{N}$ that, at the unit-state level, corresponds to a transition from state j to i induced by the reservoir ν. For vanishing bias, $\gamma_O = 0$, Eq. (4.80) reduces to the path-integral representation of the path probability in the mesoscopic space.

To proceed, we notice that the microscopic stochastic entropy production with bounding Gibbs states (2.103) is invariant under the dynamically exact coarse-graining (4.14). Inserting the latter into the generic path-integral representation for a mesoscopic generating function (4.80), we get

$$G(\boldsymbol{\gamma}, t) = \int \mathcal{D}[\boldsymbol{N}] \int \mathcal{D}[\boldsymbol{\pi}]\, P^{eq}_{\boldsymbol{N}_0}(\boldsymbol{\lambda}_0) \cdot \exp\Bigg\{ \int_0^t \mathrm{d}t' \Bigg(-\gamma_{\Lambda} \beta^{(1)} \dot{\boldsymbol{\lambda}}_{t'} \cdot \big[\nabla_{\boldsymbol{\lambda}_{t'}} E_{\boldsymbol{N}}(\boldsymbol{\lambda}_{t'}) \big] + \sum_{i=1}^{q} \Big\{ -\pi_i(t') \dot{N}_i(t') + \\ + \sum_{\nu=1}^{L} \sum_{j=1}^{q} \Big[\exp\Big\{ \pi_i(t') - \pi_j(t') - \gamma_F^{(\nu)} \beta^{(\nu)} f_{ij}^{(\nu)} - \gamma_E^{(\nu)} [\beta^{(1)} - \beta^{(\nu)}] E_{ij}^{(\nu)}\big(\boldsymbol{N}(t')\big) \Big\} - 1 \Big] W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'}, \boldsymbol{N}(t')) \Big\} \Bigg) \Bigg\}, \tag{4.84}$$

where we used the shorthand notation from Eq. (4.70). We rescale the size-extensive state variables to express them in terms of the size-intensive density $\overline{\boldsymbol{n}} \equiv \boldsymbol{N}/N$. In the macroscopic limit, $N \to \infty$, there is a single trajectory that carries all the weight of all possible paths contributing to the path integral (4.80). This trajectory maximizes the size-intensive action functional, $\max \mathcal{L}_{\boldsymbol{\gamma}}[\overline{\boldsymbol{n}}, \boldsymbol{\pi}] = \mathcal{L}_{\boldsymbol{\gamma}}[\overline{\boldsymbol{n}}^*, \boldsymbol{\pi}^*]$, and its coordinates are therefore determined as follows

$$\frac{\delta \mathcal{L}_{\boldsymbol{\gamma}}[\overline{\boldsymbol{n}}, \boldsymbol{\pi}]}{\delta \boldsymbol{\pi}}\Big|_{\overline{\boldsymbol{n}}^*, \boldsymbol{\pi}^*} = 0, \qquad \frac{\delta \mathcal{L}_{\boldsymbol{\gamma}}[\overline{\boldsymbol{n}}, \boldsymbol{\pi}]}{\delta \overline{\boldsymbol{n}}}\Big|_{\overline{\boldsymbol{n}}^*, \boldsymbol{\pi}^*} = 0, \tag{4.85}$$

where $\overline{\boldsymbol{n}}^* \equiv \lim_{N^* \to \infty} \boldsymbol{N}/N$ is the continuous mean-field density. We consequently obtain via Eq. (4.80) the size-scaled cumulant generating function

$$\mathcal{G}(\boldsymbol{\mathcal{Y}}, t) = \lim_{N \to \infty} \frac{1}{N} \ln G(\boldsymbol{\gamma}, t) = \mathcal{L}_{\boldsymbol{\mathcal{Y}}}[\overline{\boldsymbol{n}}^*, \boldsymbol{\pi}^*], \tag{4.86}$$

where $\boldsymbol{\mathcal{Y}}$ denotes a vector of fields that count the size-extensive observables appearing in $\boldsymbol{\gamma}$.

Crucially, the size-scaled cumulant generating function associated with the entropy production (2.103) satisfies a symmetry that is formally equivalent to the one the mesoscopic generating function (4.69) exhibits, as demonstrated in appendix A.7. Explicitly, we have

$$\mathcal{G}(\boldsymbol{\mathcal{Y}}, t) = \tilde{\mathcal{G}}(\tilde{\boldsymbol{\mathcal{Y}}}, t) - \beta^{(1)} \Delta \mathcal{A}_1^{eq}(\boldsymbol{\lambda}). \tag{4.87}$$

The last equation immediately stipulates the existence of a finite-time detailed fluctuation theorem in the spirit of Eq. (4.71) that asymptotically holds in the macroscopic limit,

$$\lim_{N \to \infty} \frac{1}{N} \ln \frac{P\Big(\beta^{(1)} \delta W_{\boldsymbol{\lambda}},\, \{\delta J_F^{(\nu)}\},\, \{\delta J_E^{(\nu)}\}\Big)}{\tilde{P}\Big(-\beta^{(1)} \delta W_{\boldsymbol{\lambda}}, \{-\delta J_F^{(\nu)}\}, \{-\delta J_E^{(\nu)}\}\Big)} = \beta^{(1)} \big[\delta \mathcal{W}_{\boldsymbol{\lambda}} - \Delta \mathcal{A}_1^{eq} \big] + \sum_{\nu=1}^{L} \big[\beta^{(\nu)} \delta \mathcal{W}_f^{(\nu)} + \big[\beta^{(1)} - \beta^{(\nu)} \big] \delta \mathcal{E}^{(\nu)} \big], \tag{4.88}$$

where $\mathcal{A}_1^{eq}(\boldsymbol{\lambda}) = \lim_{N\to\infty} A_1^{eq}(\boldsymbol{\lambda})/N$, $\delta\mathcal{W}_{\boldsymbol{\lambda}} = \lim_{N\to\infty} \delta W_{\boldsymbol{\lambda}}/N$, $\delta\mathcal{W}_f^{(\nu)} = \lim_{N\to\infty} \delta W_f^{(\nu)}/N$ and $\delta\mathcal{E}^{(\nu)} = \lim_{N\to\infty} \delta E^{(\nu)}/N$ are the size-intensive equilibrium free-energy with respect to the reference reservoir $\nu = 1$, the size-intensive nonautonomous work current, the size-intensive autonomous work current and the size-intensive energy current, respectively, in the macroscopic limit.

The existence of the finite-time detailed fluctuation theorem (4.88) is an important result as it ensures the thermodynamic consistency of the path-integral approach at macroscopic scales, *i.e.* for fluctuations that are extensive in the system size N.

4.2.3.2 Mean-Field Description

Dynamics

We proceed by formulating the dynamics and thermodynamics in the macroscopic mean-field limit, where the system behaves deterministically. First we note that for an unbiased dynamics, $\boldsymbol{\gamma} = 0$, that the extremal values of the auxiliary field are $\boldsymbol{\pi}^* = 0$. Thus, the action functional (4.84) needs only to be maximized with respect to the density resulting into the following Hamiltonian equations of motion

$$\left.\frac{\delta\mathcal{L}[\overline{\boldsymbol{n}},\boldsymbol{\pi}]}{\delta\boldsymbol{\pi}}\right|_{\overline{\boldsymbol{n}}^*,\boldsymbol{\pi}^*=0} = 0 \;\Rightarrow\; \dot{\overline{\boldsymbol{n}}} = \left.\frac{\delta H[\overline{\boldsymbol{n}},\boldsymbol{\pi}]}{\delta\boldsymbol{\pi}}\right|_{\boldsymbol{\pi}=0}. \tag{4.89}$$

The Hamiltonian equations of motion correspond to the mean-field equation governing the deterministic dynamics of the most likely occupation (mean-field) density and read explicitly,

$$\partial_t \overline{n}_i(t) = \sum_{j=1}^{q} k_{ij}(\boldsymbol{\lambda}_t)\,\overline{n}_j(t)\,, \qquad \sum_{i=1}^{q} \overline{n}_i(t) = 1\,, \tag{4.90}$$

with the mean-field transition rate matrix

$$k_{ij}(\boldsymbol{\lambda}_t) = \sum_{\nu=1}^{L} k_{ij}^{(\nu)}(\boldsymbol{\lambda}_t), \tag{4.91}$$

that is stochastic, $\sum_i k_{ij}(\boldsymbol{\lambda}_t) = 0$, and whose contributions corresponding to the different heat reservoirs obey the mean-field local detailed balance

$$\frac{k_{ij}^{(\nu)}(\boldsymbol{\lambda}_t)}{k_{ji}^{(\nu)}(\boldsymbol{\lambda}_t)} = \mathrm{e}^{-\beta^{(\nu)}\left\{\epsilon_i(\lambda_t)-\epsilon_j(\lambda_t)+[u_i(\lambda_t')\,\overline{n}_i-u_j(\lambda_t')\,\overline{n}_j]+\sum_{k\neq i} u_{ik}(\lambda_t')\overline{n}_k(t)-\sum_{k\neq j} u_{jk}(\lambda_t')\overline{n}_k(t)-f_{ij}^{(\nu)}\right\}}. \tag{4.92}$$

It is remarkable that the evolution equation for the mean-field density takes the form of a nonlinear master equation. We furthermore note that because of probability conservation the mean-field equation is $q-1$ dimensional.

First And Second Law

Analogously to Sec. 4.2.2.4, we now want to formulate the first and second law in the macroscopic mean-field limit. Following a similar procedure as for the derivation of Eq. (4.90), we obtain from Eqs. (4.80) and (4.81) for the mean-field energy,

$$\mathcal{E}(t) = \sum_{i=1}^{q} \mathcal{E}_i(\boldsymbol{\lambda}_t)\, \overline{n}_i(t), \quad \mathcal{E}_i(\boldsymbol{\lambda}_t) = \epsilon_i(\lambda_t) + u_i(\lambda_t') + \sum_{k\neq i} u_{ik}(\lambda_t')\overline{n}_k(t), \tag{4.93}$$

whose time-derivative constitutes the first law in the macroscopic limit,

$$\mathrm{d}_t \mathcal{E}(t) = \dot{\mathcal{Q}}(t) + \dot{\mathcal{W}}(t), \tag{4.94}$$

with the mean-field heat and work current,

$$\dot{\mathcal{Q}}(t) = \sum_{\nu=1}^{L}\sum_{i,j=1}^{q}\Big[\mathcal{E}_i(\boldsymbol{\lambda}_t) - \mathcal{E}_j(\boldsymbol{\lambda}_t) - f_{ij}^{(\nu)}\Big] k_{ij}^{(\nu)}(\boldsymbol{\lambda}_t)\overline{n}_j(t) = \sum_{\nu=1}^{L}\sum_{i,j=1}^{q} -\frac{1}{\beta^{(\nu)}} \ln\left[\frac{k_{ij}^{(\nu)}(\boldsymbol{\lambda}_t)}{k_{ij}^{(\nu)}(\boldsymbol{\lambda}_t)}\right] k_{ij}^{(\nu)}(\boldsymbol{\lambda}_t)\, \overline{n}_j(t) \tag{4.95}$$

$$\dot{\mathcal{W}}(t) = \sum_{i=1}^{q}\left\{\dot{\boldsymbol{\lambda}}_t \cdot \nabla_{\boldsymbol{\lambda}_t}\, \mathcal{E}_i(\boldsymbol{\lambda}_t)\, \overline{n}_i(t) + \sum_{\nu=1}^{L}\sum_{j=1}^{q}\Big[\sum_{k\neq i} \overline{n}_i(t)\, u_{ik}(\lambda_t')\, k_{kj}^{(\nu)}(\boldsymbol{\lambda}_t) + f_{ij}^{(\nu)}\, k_{ij}^{(\nu)}(\boldsymbol{\lambda}_t)\Big]\, \overline{n}_j(t)\right\}. \tag{4.96}$$

A closer inspection of Eqs. (4.76) and (4.77) reveals that in the deterministic macroscopic limit the stochastic (Shannon) part of the mesoscopic system entropy vanishes and only the internal entropy of the mesostates (4.19) remains finite. Using the Stirling approximation

$$\ln N! = N\ln N - N + \mathbb{O}(\ln N), \tag{4.97}$$

we find with Eqs. (4.10) and (4.19) that the total internal entropy can be rewritten as follows,

$$S^{int} = \sum_{i=1}^{q}[N_i \ln N - N_i] - \sum_{i=1}^{q}[N_i \ln N_i - N_i] + \mathbb{O}(\ln N) = -\sum_{i=1}^{q} N_i \ln n_i + \mathbb{O}(\ln N), \tag{4.98}$$

where $\mathbb{O}(\ln N)$ gives the order of magnitude of the error made by the approximation. As a result, using Eqs. (4.76) or (4.77) we obtain for the macroscopic entropy, respectively,

$$\mathcal{S}(t) = -\sum_{i=1}^{q} \overline{n}_i(t)\, \ln \overline{n}_i(t) \tag{4.99}$$

$$\mathcal{S}^s(\boldsymbol{\lambda}) = -\sum_{i=1}^{q} \overline{n}_i^s(\boldsymbol{\lambda})\, \ln \overline{n}_i^s(\boldsymbol{\lambda}), \tag{4.100}$$

where we used that $(\ln N)/N \to 0$ as $N \to \infty$. The entropy in deterministic many-body systems therefore originates from the Boltzmann entropies related to the internal structure

of the mesostates. Remarkably, the deterministic macroscopic entropy takes the form of a Shannon entropy for the mean-field density.

Next, Equations (4.78) and (4.99) or Eqs. (4.79) and (4.100) stipulate for the second law in the macroscopic limit, respectively,

$$\dot{\mathcal{S}}_i(t) = \sum_{\nu=1}^{L} \sum_{i,j=1}^{q} \ln \frac{k_{ij}^{(\nu)}(\boldsymbol{\lambda}_t)\, \overline{n}_j(t)}{k_{ji}^{(\nu)}(\boldsymbol{\lambda}_t)\, \overline{n}_i(t)}\, k_{ij}^{(\nu)}(\boldsymbol{\lambda}_t)\, \overline{n}_j(t) = \dot{\mathcal{S}}(t) - \dot{\mathcal{S}}_e(t) \geq 0 \tag{4.101}$$

$$\dot{\mathcal{S}}_i^s(\boldsymbol{\lambda}) = \sum_{\nu=1}^{L} \sum_{i,j=1}^{q} \ln \frac{k_{ij}^{(\nu)}(\boldsymbol{\lambda})\, \overline{n}_j^s(\boldsymbol{\lambda})}{k_{ji}^{(\nu)}(\boldsymbol{\lambda})\, \overline{n}_i^s(\boldsymbol{\lambda})}\, k_{ij}^{(\nu)}(\boldsymbol{\lambda})\, \overline{n}_j^s(\boldsymbol{\lambda}) = -\dot{\mathcal{S}}_e^s(\boldsymbol{\lambda}) \geq 0. \tag{4.102}$$

Hence the microscopic and mesoscopic observables in Eqs. (4.72)–(4.79) converge to the corresponding macroscopic ones in Eq. (4.93)–(4.102) if the macroscopic limit is taken,

$$\lim_{N\to\infty} \frac{1}{N} \dot{O}(t) = \dot{O}(t), \qquad O = E, Q, W, S, S_e, \Sigma, \tag{4.103}$$

where we recall that the mesoscopic representations for $O = S, \Sigma$ are only valid if the microstates inside each mesostate are equiprobable.

This constitutes our main result: For thermodynamically consistent and discrete many-body systems with all-to-all interactions there is an exact coarse-graining (4.14) of the microscopic stochastic dynamics towards a mesoscopic one that is fully characterized by the system occupation. In the macroscopic limit, $N \to \infty$, the stochastic dynamics asymptotically converges to a deterministic and nonlinear macroscopic (mean-field) master equation (4.90). Hence the stochastic dynamics can be equivalently represented across microscopic and mesoscopic scales and asymptotically on macroscopic scales as $N \to \infty$. Furthermore, the thermodynamics can be equivalently formulated at microscopic and mesoscopic scales if the microstates inside each mesostate are equiprobable (4.27). The thermodynamic consistency at each of the two levels is encoded in the respective detailed fluctuation theorem, see Eqs. (4.66) and (4.71). Using a path-integral representation of the stochastic (thermo)dynamics à la Martin-Siggia Rose, the fluctuations which scale exponentially with the system size also satisfy a detailed fluctuation theorem (4.88) and are therefore also thermodynamically consistent.

4.2.4 Example

To illustrate the methodology developed in the preceding Sec. 4.2.3 we consider a semi-analytically solvable autonomous Ising model which exhibits a dynamical phase transition, thus representing a suitable model to demonstrate the utility of the aforementioned methods. To this end, let us consider $N \to \infty$ spins with flat energy landscapes, $\epsilon_1 = \epsilon_2$, that globally interact via a pair potential u/N if they occupy the same spin state $i = 1, 2$. The system is in contact with two heat reservoirs at different inverse temperatures β^h and β^c with $\beta^h < \beta^c$.

According to Eq. (4.90), the mean-field dynamics is governed by the following nonlinear master equation

$$\partial_t \overline{n}_i = -\Big(k_{ji}^{(h)} + k_{ji}^{(c)}\Big)\overline{n}_i + \Big(k_{ij}^{(h)} + k_{ij}^{(c)}\Big)\overline{n}_j = -k_{ji}\,\overline{n}_i + k_{ij}\,\overline{n}_j, \quad i,j = 1,2\,, \tag{4.104}$$

with the mean-field transition rates which we assume to be of Arrhenius form

$$k_{ij}^{(\nu)} = \Gamma\, \exp\Big[-\frac{\beta^{(\nu)}}{2} u(\overline{n}_i - \overline{n}_j)\Big], \quad \nu = c,h, \tag{4.105}$$

with the constant kinetic prefactor Γ that sets the time-scale of the Markov jump process. We note that the mean-field dynamics (4.104) is effectively a one-dimensional equation since we have $\overline{n}_2 = 1 - \overline{n}_1$ because the number of spins are conserved. We can immediately read off the stationary solution $\overline{n}_i^s = 1/2$, $i = 1,2$ for Eq. (4.104). The stability of this symmetric fixed point is encoded in the spectrum of the linearized Jacobian, $\boldsymbol{A}_{ij} \equiv [\partial(\partial_t \overline{n}_i)/\partial \overline{n}_j]|_{\overline{n}_{i,j}=1/2}$, which can be readily determined as follows

$$\lambda_1 = 0, \quad \lambda_2 = -\Gamma\big[4 + u\big(\beta^{(h)} + \beta^{(c)}\big)\big]. \tag{4.106}$$

The zero eigenvalue λ_1 reflects that the rank of the Jacobian is smaller than its dimension due to the constraint $\sum_i \partial_t \overline{n}_i = 0$. More strikingly, the second eigenvalue λ_2 changes its sign for attractive interactions, $u < 0$, at the critical temperatures

$$4 + u\big(\beta_c^{(h)} + \beta_c^{(c)}\big) = 0, \tag{4.107}$$

indicative of a supercritical pitchfork bifurcation that destabilizes the symmetric fixed point into two asymmetric fixed points as can be seen in Fig. 4.1.

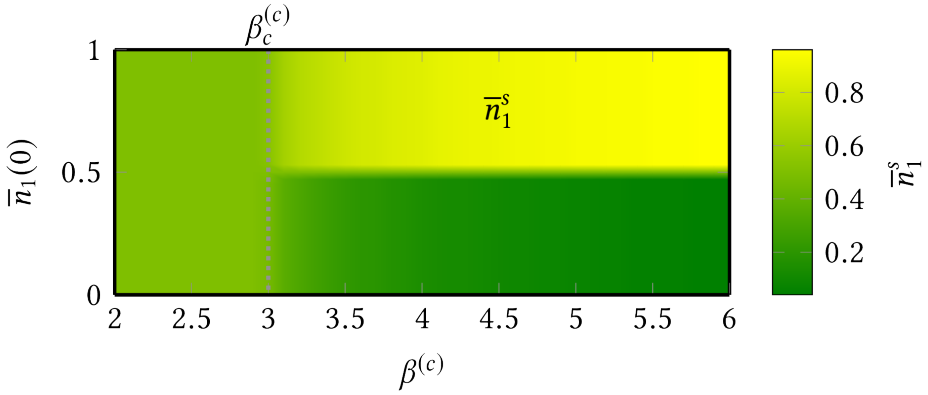

Figure 4.1: *Density plot of the stationary solution $\overline{n}_1^s$ for different temperatures $\beta^{(c)}$ and all physical initial conditions $\overline{n}_1(0)$. We choose the following values for the parameters $\Gamma = 0.1$, $\beta^{(h)} = 1$, $u = -1$ so that $\beta_c^{(c)} = 3$ as indicated by the vertical dotted line.*

This density plot depicts the stationary solution $\overline{n}_1^s$ as a function of all physical initial conditions $\overline{n}_1(0)$ and for different cold temperatures $\beta^{(c)}$ while $\beta^{(h)} \equiv 1$ and $u = -1$ are kept constant. As can be observed, the symmetric fixed point is stable for $\beta^{(c)} < \beta_c^{(c)} = 3$. In contrast, for lower temperatures $\beta^{(c)} > \beta_c^{(c)} = 3$ the symmetric fixed point is unstable and the system dynamics goes to one of the two asymmetric stable fixed points depending on the basin of attraction in which the initial condition lies. These two stable fixed points are related to each other via permutations of their coordinates, in agreement with the invariance of the mean-field Eq. (4.104) under a permutation operation. The phenomenology observed in Fig. 4.1 can be physically seen as follows. In the high-temperature limit the system behaves entropically, thus occupying the symmetric fixed point. Conversely, in the low-temperature limit the system behaves energetically, thus exhibiting two asymmetric fixed points that converge to the two energy ground states $\overline{n}_i = 1, \overline{n}_{i+1} = 0,\ i \bmod 2$ as $\beta \to \infty$. For isothermal systems, Eq. (4.107) implies the critical point $\beta_c = -2/u$. This is in agreement with the q-dependent universal critical temperature, $\beta_c(q) = -q/u$ for isothermal and all-to-all interacting q-state clock models derived in Eq. (3.106). We add that the isothermal system displays a first-order equilibrium phase transition, as can be seen in Fig. 4.2 which shows the equilibrium free energy close to the critical point $\beta_c = -2/u$.

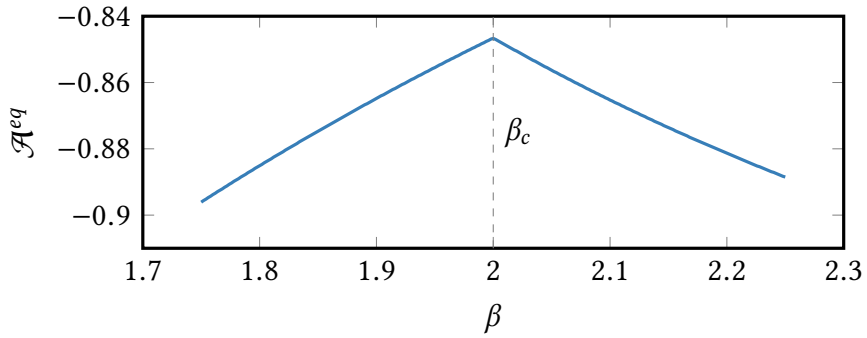

Figure 4.2: *The equilibrium free energy as a function of the temperature close to the critical point $\beta_c = -2/u$ with $u = -1$.*

We now return to the non-isothermal case and consider the fluctuating quantity in Eq. (2.103) that for the autonomous Ising model simplifies to

$$\delta\Sigma[\boldsymbol{m}_{(\tau)}, t] = \sum_{\nu=h,c} \left[\beta^{(h)} - \beta^{(\nu)}\right] \delta E^{(\nu)}[\boldsymbol{m}_{(\tau)}, t],\ \delta E^{(\nu)}[\boldsymbol{m}_{(\tau)}, t] = u \int_0^t \mathrm{d}t' \sum_{j=1}^{M} \delta(\nu - \nu_j)\delta(t' - \tau_j)[n_j - n_{j-1}]. \tag{4.108}$$

According to Eq. (4.71), our model system therefore satisfies a finite-time detailed fluctuation theorem for the time-integrated energy current. Using the path-integral formalism introduced in Sec. 4.2.3, we however observe that analytical progress is difficult at finite time

as it would require to solve the full extremization problem (4.85) which is analytically not possible. Instead, we therefore resort to the stationary case which considerable simplifies the problem of finding the dominant trajectory among all paths contributing to the path integral. The biased Hamiltonian (4.81) in the path-integral formulation of the generating function (4.84) associated with the stochastic observable in the last equation reads

$$\begin{aligned} H_\gamma[\overline{\boldsymbol{n}}, \boldsymbol{\pi}] = k_{21}^{(h)}\,\overline{n}_1\left[\mathrm{e}^{\pi_2-\pi_1}-1\right] + k_{21}^{(c)}\,\overline{n}_1\left[\mathrm{e}^{\pi_2-\pi_1}\mathrm{e}^{\gamma u[\beta^{(h)}-\beta^{(c)}](\overline{n}_2-\overline{n}_1)}-1\right] + \\ + k_{12}^{(h)}\,\overline{n}_2\left(\mathrm{e}^{-(\pi_2-\pi_1)}-1\right) + k_{12}^{(c)}\,\overline{n}_2\left(\mathrm{e}^{-(\pi_2-\pi_1)}\mathrm{e}^{-\gamma u[\beta^{(h)}-\beta^{(c)}](\overline{n}_2-\overline{n}_1)}-1\right). \end{aligned} \tag{4.109}$$

At steady state, the Hamiltonian equations of motion resulting from the extremization of the action functional in Eq. (4.85) read

$$\partial_{\overline{n}_i} H_{\gamma_E^{(c)}}[\overline{\boldsymbol{n}}, \boldsymbol{\pi}, \lambda_{\overline{n}}, \lambda_\pi] = 0, \quad \partial_{\pi_i} H_{\gamma_E^{(c)}}[\boldsymbol{\pi}, \boldsymbol{\pi}, \lambda_{\overline{n}}, \lambda_\pi] = 0, \quad i = 1, 2, \tag{4.110}$$

where we added the Lagrangian multipliers $\lambda_{\overline{n}}$ and λ_π to enforce the spin conservation, $\overline{n}_1 + \overline{n}_2 - 1 = 0$ and $\pi_1 + \pi_2 = 0$. The extremal value for π_1 can be solved analytically,

$$\pi_1^* = \frac{1}{4}\ln\left(\frac{\overline{n}_1}{\overline{n}_1-1}\cdot\frac{\mathrm{e}^{\beta^{(h)}u(2\overline{n}_1-1)}+\mathrm{e}^{\frac{u}{2}(2\overline{n}_1-1)\left[\beta^{(c)}\left(2\gamma_E^{(c)}-1\right)-\beta^{(h)}\left(1+2\gamma_E^{(c)}\right)\right]}}{1+\mathrm{e}^{\frac{\beta^{(c)}-\beta^{(h)}}{2}u(2\overline{n}_1-1)\left(2\gamma_E^{(c)}-1\right)}}\right)+\mathrm{i}\pi(1+2k),\ k\in\mathbb{Z}, \tag{4.111}$$

and the extremal value $\overline{n}_1^*$ is subsequently determined numerically. In the $t\to\infty$ limit, the boundary terms in the action functional become negligible so that the time- and size-scaled cumulant generating function is asymptotically equal to the biased Hamiltonian evaluated at the extremal values $\overline{\boldsymbol{n}}^*$ and $\boldsymbol{\pi}^*$,

$$\mathcal{G}^s\left(\gamma_{\mathcal{E}}^{(c)}\right) = \lim_{t\to\infty}\frac{1}{t}\lim_{N\to\infty}\frac{1}{N}\ln G\left(\gamma_E^{(c)}, t\right) = H^s_{\gamma_{\mathcal{E}}^{(c)}}[\overline{\boldsymbol{n}}^*, \boldsymbol{\pi}^*]. \tag{4.112}$$

The scaled cumulant generating function is plotted in Fig. 4.3a). We choose the values $\beta^{(h)} = 3, \beta^{(c)} = 5, u = -1$ corresponding to the phase where the mean-field dynamics exhibits two asymmetric stable and a symmetric unstable fixed point. Similarly, we observe two asymmetric γ-dependent fixed points $\overline{n}_1^*(\gamma)$ whose coordinates are related to each other via a permutation as well as a symmetric fixed point at $\overline{n}_1^* = 1/2$. The regime around $1/2$ corresponds to the symmetric fixed point and thus to a null observable (4.108). Next, we note that the curve is symmetric with respect to the value $\gamma = 1/2$, thus implying that the scaled cumulant generating function asymptotically satisfies the symmetry relation

$$\mathcal{G}^s\left(\gamma_{\mathcal{E}}^{(c)}\right) = \tilde{\mathcal{G}}^s\left(1-\gamma_{\mathcal{E}}^{(c)}\right), \tag{4.113}$$

which in turn stipulates the existence of a macroscopic steady-state detailed fluctuation theorem for the time-integrated energy current

$$\lim_{t\to\infty}\frac{1}{t}\lim_{N\to\infty}\frac{1}{N}\ln\frac{P\left(\delta J_E^{(c)}\right)}{\tilde{P}\left(-\delta J_E^{(c)}\right)} = \delta\mathcal{J}_{\mathcal{E}}^{(c),s}, \quad \delta\mathcal{J}_{\mathcal{E}}^{(c),s} = [\beta^{(h)}-\beta^{(c)}]\lim_{t\to\infty}\frac{1}{t}\lim_{N\to\infty}\frac{1}{N}\delta E^{(c)}. \tag{4.114}$$

The existence of the steady-state fluctuation theorem is by no means obvious, here. In general, the implicit assumption underlying steady-state fluctuation theorems is that the contribution of the boundary terms related to the initial and final state of each trajectory are subextensive in time and thus negligible in the infinite-time limit. There are however situations where this may not be true, *e.g.* in bistable systems for starting distributions of the forward and backward process that are located in the different basins of attraction. Though, in this model the two $\gamma_\varepsilon^{(c)}$-dependent fixed points are related to each other via permutation of their coordinates and the statistics of the corresponding stationary states are thus identical.

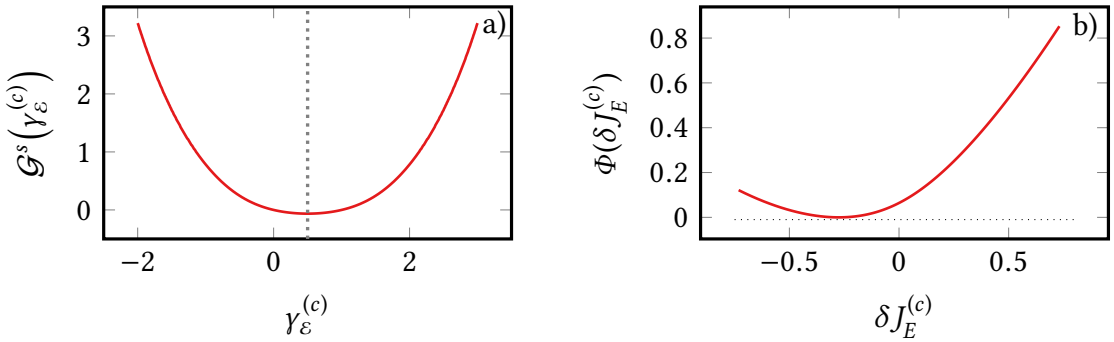

Figure 4.3: *The cumulant generating function* (4.112) *scaled with time t and size N as a function of the counting field $\gamma_\varepsilon^{(c)}$ in a) and the corresponding rate function $\Phi(\delta J_E^{(c)})$ in b). The parameters are chosen as $\beta^{(h)} = 3, \beta^{(c)} = 5, u = -1$ so that for $\gamma_\varepsilon^{(c)} = 0$ the stationary mean-field system is in its energetic phase which has two asymmetric stable fixed points and a symmetric unstable one.*

Figure 4.3b) shows the rate function $\Phi(\delta J_E^{(c)})$ associated with the scaled cumulant generating function $\mathcal{G}^s(\gamma_\varepsilon^{(c)})$ in a). The rate function is defined as [163]

$$\Phi(\delta J_E^{(c)}) = -\lim_{t\to\infty}\frac{1}{t}\lim_{N\to\infty}\frac{1}{N}\ln P(\delta J_E^{(c)}), \tag{4.115}$$

and is related to its corresponding scaled cumulant generating function via a Legendre-Fenchel transformation

$$\Phi(\delta J_E^{(c)}) = \sup_{\delta J_E^{(c)}}[\gamma_\varepsilon^{(c)}\,\delta J_E^{(c)} - \mathcal{G}^s(\gamma_\varepsilon^{(c)})], \qquad \delta J_E^{(c)} = \frac{\partial \mathcal{G}^s(\gamma_\varepsilon^{(c)})}{\partial \gamma_\varepsilon^{(c)}}. \tag{4.116}$$

Here, $P(\delta J_E^{(c)})$ is the probability to observe a change in the energy equal to $\delta J_E^{(c)}$ and sup denotes the supremum. As can be seen in Fig. 4.3, both the scaled cumulant generating and the rate function are convex functions and the latter has a unique minimum equal to zero.

Our thermodynamically consistent framework allows to translate the terminology of nonlinear dynamics, *i.e.* the supercritical pitchfork bifurcation at the critical temperature

(4.107), into the language of nonequilibrium statistical mechanics, *i.e.* a nonequilibrium phase transition at the same critical temperature. For this purpose, we prepare the system in its critical state by setting $\beta^{(h)} = 1, \beta^{(c)} = 3, u = -1$. Fig. 4.4 depicts in a) the scaled cumulant generating function (4.112) with the system being in its critical state. The scaled cumulant generating function exhibits a kink at $\gamma_{\mathcal{E}}^{(c)} = 0$ indicative of a nonequilibrium phase transition. Owing to the symmetry (4.113), the scaled cumulant generating function has another kink at $\gamma_{\mathcal{E}}^{(c)} = 1$. The non-differentiability of the generating function at $\gamma_{\mathcal{E}}^{(c)} = 0$ implies that the rate function in Fig. (4.112) b) would be nonconvex over a finite interval. The Legendre-Fenchel transformation (4.116) yields not the nononvex rate function but its convex envelope $\Phi^{ce}\left(\delta J_E^{(c)}\right)$. Here, the part of the convex envelope that replaces the nonconvex regime of the rate function corresponds to the flat part of the curve in the vicinity of the $\delta J_E^{(c)} = 0$. Thus, we find that the time-integrated energy current distribution in Eq. (4.115) is bimodal, thus also encoding the nonequilibrium phase transition.

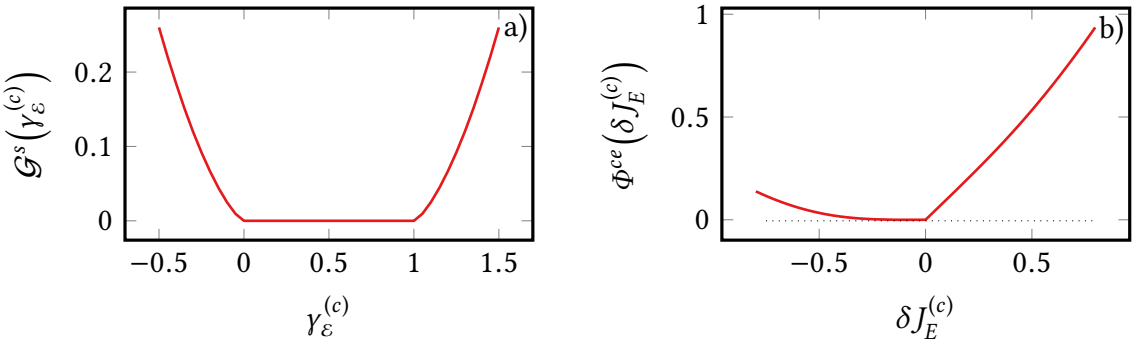

Figure 4.4: *The cumulant generating function* (4.112) *scaled with time t and size N in a) as a function of the counting field $\gamma_{\mathcal{E}}^{(c)}$ and the corresponding convex envelope of the rate function $\Phi^{ce}\left(\delta J_E^{(c)}\right)$ in b). The parameters are chosen as $\beta^{(h)} = 1, \beta^{(c)} = 3, u = -1$ where the unbiased dynamics exhibits a phase transition* (4.107).

4.3 Two Interacting Underdamped Particles

So far, we considered overdamped stochastic systems governed by master equations. As a complement to the study of thermodynamically consistent coarse-graining in underdamped and all-to-all interacting systems, the following is devoted to investigating thermodynamically consistent coarse-graining in the underdamped case.

4.3.1 Stochastic Thermodynamics in Underdamped Systems

4.3.1.1 Single Underdamped Particle

We consider a particle of mass m with the phase-space coordinate $\boldsymbol{\Gamma} = (\boldsymbol{x}, \boldsymbol{v})^\top \in \mathbb{R}^6$, where $\boldsymbol{x} \in \mathbb{R}^3$ and $\boldsymbol{v} \in \mathbb{R}^3$ denote position and velocity of the particle, respectively. The particle moves in a time-dependent potential $u(\boldsymbol{x}, t)$, hence its Hamiltonian reads

$$e(\boldsymbol{\Gamma}, t) = \frac{m}{2}\boldsymbol{v}^2 + u(\boldsymbol{x}, t). \tag{4.117}$$

The particle is furthermore subjected to a generic force $\boldsymbol{f}(\boldsymbol{\Gamma}, t)$. If the force is conservative it derives from a potential, $\boldsymbol{f}(\boldsymbol{x}, t) = -\partial_{\boldsymbol{x}}\hat{u}(\boldsymbol{x}, t)$. In this case, one can define heat and work in different ways depending on whether $\hat{u}(\boldsymbol{x}, t)$ is part of the system Hamiltonian (4.117) or not, see Ref. [167]. Conversely, if the force $\boldsymbol{f}$ is nonconservative, it does not derive from a potential. For generality, and since it will be useful later, we assume that the force may be velocity-dependent, $\boldsymbol{f}(\boldsymbol{\Gamma}, t)$.

The system is coupled to a heat reservoir at inverse temperature β, giving rise to zero-mean delta-correlated Gaussian white noise

$$\langle \boldsymbol{\xi}_i(t) \rangle = 0, \quad \langle \boldsymbol{\xi}_i(t)\boldsymbol{\xi}_j(t') \rangle = 2\xi\, \beta^{-1}\, \delta_{ij}\delta(t - t'), \tag{4.118}$$

for $i, j = 1, 2, 3$. We denote by ξ the friction the particle experiences. Then, the stochastic dynamics of the system is governed by the following Langevin equation

$$\begin{pmatrix} \partial_t \boldsymbol{x} \\ \partial_t \boldsymbol{v} \end{pmatrix} = \begin{pmatrix} \boldsymbol{v} \\ \frac{1}{m}\left[-\partial_{\boldsymbol{x}} u(\boldsymbol{x}, t) + \boldsymbol{f}(\boldsymbol{\Gamma}, t) - \xi\, \boldsymbol{v} + \boldsymbol{\xi}(t)\right] \end{pmatrix}, \tag{4.119}$$

and the equivalent Fokker-Planck equation ruling the time evolution of the probability density $\rho(\boldsymbol{\Gamma}, t)$ reads

$$\partial_t\, \rho = -\nabla \cdot (\boldsymbol{\mu}\rho) + \nabla \cdot \big(\boldsymbol{D} \cdot \nabla\rho\big), \tag{4.120}$$

with the drift and diffusion matrices

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{v} \\ \frac{1}{m}\left[-\partial_{\boldsymbol{x}} u(\boldsymbol{x}, t) + \boldsymbol{f}(\boldsymbol{\Gamma}, t) - \xi\, \boldsymbol{v}\right] \end{pmatrix} \tag{4.121}$$

$$\boldsymbol{D}_{ij} = \frac{\xi\,\delta_{ij}}{\beta m^2} \sum_{n=4}^{6} \delta_{in}\,, \tag{4.122}$$

and the nabla operator $\nabla \equiv (\partial_{\boldsymbol{x}}, \partial_{\boldsymbol{v}})^\top$. The Fokker-Planck Eq. (4.120) can be cast into a continuity equation

$$\partial_t \rho = -\nabla \cdot \boldsymbol{J} = -\nabla \cdot \left(\boldsymbol{L}^{det} + \boldsymbol{L}^{diss}\right) \rho. \tag{4.123}$$

Here, the probability current $\boldsymbol{J}$ is split into a deterministic contribution

$$\boldsymbol{L}^{det} = \begin{pmatrix} \boldsymbol{v} \\ \frac{1}{m}\left[-\partial_{\boldsymbol{x}} u(\boldsymbol{x}, t) + \boldsymbol{f}(\boldsymbol{\Gamma}, t)\right] \end{pmatrix}, \tag{4.124}$$

and a dissipative one

$$\boldsymbol{L}^{diss} = -\frac{\xi}{m^2} \begin{pmatrix} 0 \\ m\boldsymbol{v} + \beta^{-1}\,\partial_{\boldsymbol{v}} \ln \rho \end{pmatrix}. \tag{4.125}$$

The average energy of the particle is

$$E = \int \mathrm{d}\boldsymbol{\Gamma}\; e\,\rho, \tag{4.126}$$

and its rate of change

$$d_t E = \dot{Q} + \dot{W}, \tag{4.127}$$

can be decomposed into a work current

$$\dot{W} = \int \mathrm{d}\boldsymbol{\Gamma}\; \rho\,\partial_t e + \int \mathrm{d}\boldsymbol{\Gamma}\; \rho\,\boldsymbol{f} \cdot \boldsymbol{v}, \tag{4.128}$$

and into a heat current

$$\dot{Q} = \int \mathrm{d}\boldsymbol{\Gamma}\; e\,\partial_t \rho - \int \mathrm{d}\boldsymbol{\Gamma}\; \rho\,\boldsymbol{f} \cdot \boldsymbol{v}. \tag{4.129}$$

Eq. (4.127) constitutes the first law of thermodynamics ensuring energy conservation [126]. Using Eq. (4.123), the heat current can be written as follows

$$\dot{Q} = -\xi \int \mathrm{d}\boldsymbol{\Gamma}\; \rho \left(\boldsymbol{v} + \frac{1}{\beta m}\partial_{\boldsymbol{v}} \ln \rho\right) \boldsymbol{v}\,. \tag{4.130}$$

The nonequilibrium system entropy associated with the particle at $\boldsymbol{\Gamma}$ is defined as [20]

$$s(\boldsymbol{\Gamma}) = -\ln \rho, \tag{4.131}$$

where the ensemble average coincides with the Shannon entropy

$$S = -\int \mathrm{d}\boldsymbol{\Gamma}\ \rho\ \ln\rho. \tag{4.132}$$

Its time-derivative

$$\mathrm{d}_t S = \int \mathrm{d}\boldsymbol{\Gamma}\,[\nabla\cdot\boldsymbol{L}^{diss}]\,\rho + \dot{I}_F = \beta\dot{Q} + \dot{\Sigma} + \dot{I}_F, \tag{4.133}$$

can be split into the entropy flow from the bath to the system, $\beta\dot{Q}$, and the entropy production rate

$$\dot{\Sigma} = \beta\,\xi\int \mathrm{d}\boldsymbol{\Gamma}\ \rho\left(\boldsymbol{v} + \frac{1}{\beta m}\partial_{\boldsymbol{v}}\ln\rho\right)^2 \geq 0, \tag{4.134}$$

whose nonnegativity constitutes the second law of thermodynamics. Since it will be useful later, we introduced the notation

$$\dot{I}_F \equiv \frac{1}{m}\int \mathrm{d}\boldsymbol{\Gamma}\ \rho\,\partial_{\boldsymbol{v}}\cdot\boldsymbol{f}. \tag{4.135}$$

Defining the nonequilibrium free-energy density $f(\boldsymbol{\Gamma}) = e(\boldsymbol{\Gamma}) - \beta^{-1}s(\boldsymbol{\Gamma})$, one has for the average nonequilibrium free energy

$$A = \int \mathrm{d}\boldsymbol{\Gamma}\,\rho\ f = E - \beta^{-1}S. \tag{4.136}$$

Eq. (4.136) allows us to rewrite the work and heat current in Eqs. (4.128) and (4.130) as

$$\begin{aligned}\dot{W} &= \int \mathrm{d}\boldsymbol{\Gamma}\ \rho\,\partial_t f + \int \mathrm{d}\boldsymbol{\Gamma}\rho\,\boldsymbol{f}\cdot\boldsymbol{v}\\ \dot{Q} &= d_t(A + \beta^{-1}S) - \dot{W},\end{aligned} \tag{4.137}$$

and the entropy production rate in Eq. (4.134) as

$$\dot{\Sigma} = \beta(\dot{W} - \mathrm{d}_t A) - \dot{I}_F\ \geq 0. \tag{4.138}$$

The additional term $\dot{I}_F$ in Eqs. (4.133) and (4.138) illustrates that the presence of the velocity-dependent nonconservative force $\boldsymbol{f}$ modifies the thermodynamics as noted in Refs. [168, 169].

4.3.1.2 Special Cases

Standard Stochastic Thermodynamics

Owing to the velocity-dependence of $f(\boldsymbol{\Gamma}, t)$, Eqs. (4.133) and (4.138) constitute a generalized entropy balance and a generalized second law, respectively. The standard thermodynamic formulation

$$\mathrm{d}_t S = \beta\dot{Q} + \dot{\Sigma}, \qquad T\dot{\Sigma} = \dot{W} - \mathrm{d}_t A \geq 0, \tag{4.139}$$

is recovered for velocity-independent or nonconservative Lorentz forces, that is forces that are orthogonal to the velocity, $\partial_{\boldsymbol{v}} \cdot \boldsymbol{f} = 0$. In one dimension, this is only true for velocity-independent forces $\partial_v f = 0$.

Deterministic Limit

The dynamics is deterministic if $\xi = 0$, which physically corresponds to a decoupling of the particle from the thermal reservoir. According to Eq. (4.130), one has $\dot{Q} = 0$ and $\mathrm{d}_t E = \mathrm{d}_t W$. It follows furthermore from Eq. (4.134) that $\dot{\Sigma} = 0$, hence it holds, using Eq. (4.133), that

$$\mathrm{d}_t S = \frac{1}{m} \int \mathrm{d}\boldsymbol{\Gamma}\, \rho\, \partial_{\boldsymbol{v}} \cdot \boldsymbol{f}. \tag{4.140}$$

Again, if $\boldsymbol{f}$ is velocity-independent or a Lorentz force, the deterministic dynamics becomes Hamiltonian and the rate of entropy change is identically zero, $\mathrm{d}_t S = 0$. In this case the second law is a triviality.

Heavy Particle

Finally, we consider the limit where the mass of the particle diverges, $m \to \infty$. We suppose that the conservative force scales with the mass, *i.e.* $\mathbb{O}(\partial_{x_i} u/m) = 1\, \forall i$, to avoid the trivial case of a particle in a flat potential. If ξ and $\boldsymbol{f}$ are finite, so that $\xi/m \to 0$ and $\boldsymbol{f}/m \to 0$, one finds using Eqs. (4.130), (4.133) and (4.134) that

$$\dot{\Sigma} = -\beta \dot{Q} = \beta \xi \boldsymbol{v}_t^2 \geq 0, \quad \mathrm{d}_t S = 0, \tag{4.141}$$

where $\boldsymbol{v}_t$ is the solution of the deterministic Eqs.

$$\partial_t \boldsymbol{x}_t = \boldsymbol{v}_t, \quad \partial_t \boldsymbol{v}_t = -\frac{1}{m}\, \partial_{\boldsymbol{x}} u(\boldsymbol{x}, t)|_{\boldsymbol{x}=\boldsymbol{x}_t}\,. \tag{4.142}$$

According to Eq. (4.141), the heavy particle corresponds to the limit of macroscopic friction.

4.3.1.3 Two Coupled Underdamped Particles

We now consider two particles labeled by $i = 1, 2$ of mass m_i with the phase-space coordinate $\boldsymbol{\Gamma}_i = (\boldsymbol{x}_i, \boldsymbol{v}_i)^\top$, as depicted in Fig. 4.5. The particles move in a time-dependent potential

$$u(\boldsymbol{x}_1, \boldsymbol{x}_2, t) = u_1(\boldsymbol{x}_1, t) + u_2(\boldsymbol{x}_2, t) + u_{12}^{int}(\boldsymbol{x}_1, \boldsymbol{x}_2, t), \tag{4.143}$$

that contains the interaction potential $u_{12}^{int}(\boldsymbol{x}_1, \boldsymbol{x}_2, t)$ and the Hamiltonian therefore reads

$$\begin{aligned} e(\boldsymbol{\Gamma}, t) &= \frac{m_1}{2}\boldsymbol{v}_1^2 + \frac{m_2}{2}\boldsymbol{v}_2^2 + u(\boldsymbol{x}_1, \boldsymbol{x}_2, t) \\ &= \sum_i e_i(\boldsymbol{\Gamma}_i, t) + u_{12}^{int}(\boldsymbol{x}_1, \boldsymbol{x}_2, t)\,, \end{aligned} \tag{4.144}$$

where we denote the bare Hamiltonian of each particle by $e_i(\boldsymbol{\Gamma}_i, t) = m_i \boldsymbol{v}_i^2/2 + u_i(\boldsymbol{x}_i, t)$ with $i = 1, 2$. Moreover, we assume that both particles are subjected to velocity-independent nonconservative forces $\boldsymbol{f}_i(\boldsymbol{x}_i, t)$ [1].

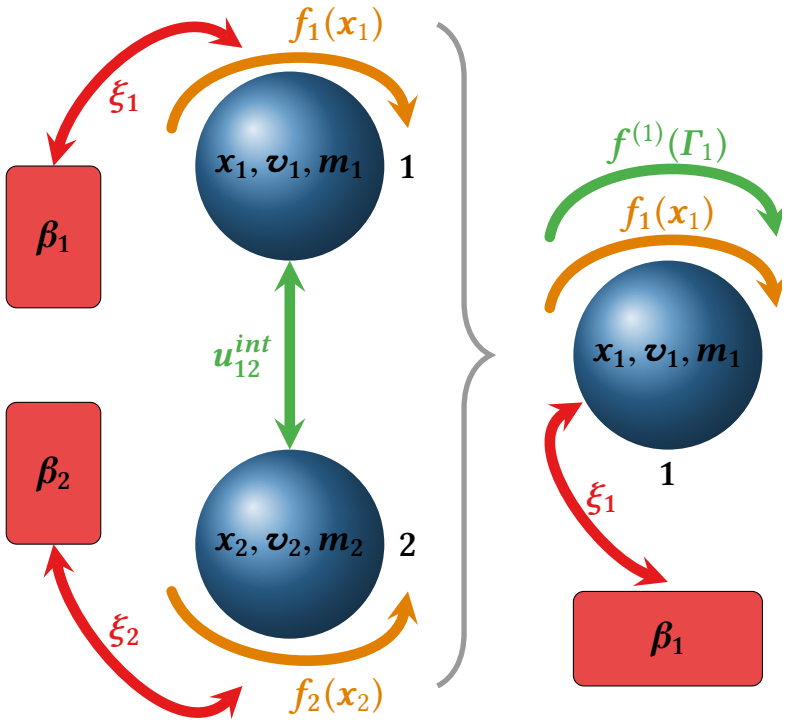

Figure 4.5: *On the left, schematics of the two underdamped and via u_{12}^{int} interacting particles* **1** *and* **2** *that are in contact with heat reservoirs at inverse temperatures β_1 and β_2, respectively, are illustrated. It is furthermore assumed that both particles are subjected to nonconservative forces $\boldsymbol{f}_i(\boldsymbol{x}_i)$. The right depicts the coarse-grained description of solely the first particle in the presence of an additional nonconservative force $\boldsymbol{f}^{(1)}(\boldsymbol{\Gamma}_1)$ that encodes the interaction with the second particle.*

Each of the particles is connected to a heat reservoir at inverse temperature β_i giving rise to uncorrelated zero-mean Gaussian white noise

$$\langle \boldsymbol{\xi}_j^{(i)}(t) \rangle = 0, \quad \langle \boldsymbol{\xi}_j^{(i)}(t) \boldsymbol{\xi}_{j'}^{(i)}(t') \rangle = 2\, \xi_i \beta_i^{-1} \delta_{j,j'}\, \delta(t - t'), \tag{4.145}$$

where ξ_i refers to the friction the particle i experiences. The stochastic dynamics of the

[1] For velocity-dependent nonconservative forces, *e.g.* magnetic forces, the following procedure is analogous to the case of velocity-independent forces. The only formal modification are the additional terms, $\sum_i 1/m_i \int \mathrm{d}\boldsymbol{\Gamma}_i \rho_i \partial_{\boldsymbol{v}_i} \cdot \boldsymbol{f}_i$, that appear in the entropy balance equation (4.156), cf. Eq. (4.133).

two-body system is ruled by the following Langevin equation

$$\begin{pmatrix}\partial_t \boldsymbol{x}_i\\ \partial_t \boldsymbol{v}_i\end{pmatrix}=\begin{pmatrix}\boldsymbol{v}_i\\ \frac{1}{m_i}[-\partial_{\boldsymbol{x}_i} u(\boldsymbol{x}_1,\boldsymbol{x}_2,t)+\boldsymbol{f}_i(\boldsymbol{x}_i,t)-\xi_i\boldsymbol{v}_i+\boldsymbol{\xi}^{(i)}(t)]\end{pmatrix}, \tag{4.146}$$

and the equivalent Fokker-Planck equation governing the time evolution of the probability density $\rho(\boldsymbol{\Gamma},t)$ reads

$$\partial_t \rho = -\nabla\cdot\boldsymbol{J} = -\nabla\cdot\left(\boldsymbol{L}^{det}+\boldsymbol{L}^{diss}\right)\rho, \tag{4.147}$$

with $\nabla=(\partial_{\boldsymbol{x}_1},\partial_{\boldsymbol{v}_1},\partial_{\boldsymbol{x}_2},\partial_{\boldsymbol{v}_2})^\top$. The probability current $\boldsymbol{J}$ can be split into a deterministic part

$$\boldsymbol{L}^{det}=\begin{pmatrix}\boldsymbol{v}_1\\ \frac{1}{m_1}\left[-\partial_{\boldsymbol{x}_1}u(\boldsymbol{x}_1,\boldsymbol{x}_2,t)+\boldsymbol{f}_1(\boldsymbol{x}_1,t)\right]\\ \boldsymbol{v}_2\\ \frac{1}{m_2}\left[-\partial_{\boldsymbol{x}_2}u(\boldsymbol{x}_1,\boldsymbol{x}_2,t)+\boldsymbol{f}_2(\boldsymbol{x}_2,t)\right]\end{pmatrix}, \tag{4.148}$$

and a dissipative one

$$\boldsymbol{L}^{diss}=\begin{pmatrix}0\\ \frac{-\xi_1}{m_1^2}(m_1\boldsymbol{v}_1+\beta_1^{-1}\partial_{\boldsymbol{v}_1}\ln\rho)\\ 0\\ \frac{-\xi_2}{m_2^2}(m_2\boldsymbol{v}_2+\beta_2^{-1}\partial_{\boldsymbol{v}_2}\ln\rho)\end{pmatrix}. \tag{4.149}$$

The average energy of the system is

$$E=\int \mathrm{d}\boldsymbol{\Gamma}\, e\,\rho, \tag{4.150}$$

and the first law of thermodynamics reads

$$\mathrm{d}_t E=\dot{Q}+\dot{W}, \tag{4.151}$$

with the heat and work current

$$\dot{Q}=\int \mathrm{d}\boldsymbol{\Gamma}\, e\,\partial_t\rho-\int \mathrm{d}\boldsymbol{\Gamma}\,\rho\,(\boldsymbol{f}_1\cdot\boldsymbol{v}_1+\boldsymbol{f}_2\cdot\boldsymbol{v}_2) \tag{4.152}$$

$$\dot{W}=\int \mathrm{d}\boldsymbol{\Gamma}\,\dot{e}\,\rho+\int \mathrm{d}\boldsymbol{\Gamma}\,\rho\,(\boldsymbol{f}_1\cdot\boldsymbol{v}_1+\boldsymbol{f}_2\cdot\boldsymbol{v}_2). \tag{4.153}$$

Using the Fokker-Planck Eq. (4.147), we can write the heat current in terms of additive contributions,

$$\dot{Q}=\sum_{i=1}^{2}\dot{q}^{(i)},\ \dot{q}^{(i)}=-\xi_i\int \mathrm{d}\boldsymbol{\Gamma}\rho\left(\boldsymbol{v}_i+\frac{1}{\beta_i m_i}\partial_{\boldsymbol{v}_i}\ln\rho\right)\boldsymbol{v}_i. \tag{4.154}$$

Like in the single-particle case (4.132), the nonequilibrium system entropy is defined as

$$S = -\int \mathrm{d}\boldsymbol{\Gamma}\, \rho \ln \rho, \tag{4.155}$$

and the entropy balance is thus given by

$$\mathrm{d}_t S = \sum_{i=1}^{2} \beta_i\, \dot{q}^{(i)} + \dot{\Sigma}, \tag{4.156}$$

where the non-negative entropy production rate

$$\dot{\Sigma} = \sum_{i=1}^{2} \dot{\sigma}^{(i)},\ \ \dot{\sigma}^{(i)} = \beta_i\, \xi_i \int \mathrm{d}\boldsymbol{\Gamma}\, \rho \left(\boldsymbol{v}_i + \frac{1}{\beta_i m_i} \partial_{\boldsymbol{v}_i} \ln \rho \right)^2 \geq 0, \tag{4.157}$$

constitutes the second law of thermodynamics. In fact, Eq. (4.157) formulates a stronger statement: the additive contributions $\dot{\sigma}^{(i)}$ are separately non-negative.

4.3.2 Coarse graining

4.3.2.1 Effective Dynamics

We now shift our attention to the first particle alone. This formally amounts to integrating the Fokker-Plank Eq. (4.147) over the coordinates of the second particle $\boldsymbol{\Gamma}_2 = (\boldsymbol{x}_2, \boldsymbol{v}_2)$ such that we obtain the marginalized probability distribution of particle one, $\rho_1 \equiv \int \mathrm{d}\boldsymbol{\Gamma}_2\, \rho$, that satisfies the following effective Fokker-Planck equation

$$\partial_t \rho_1 = -\nabla_1 \cdot \boldsymbol{J}_1 = -\nabla_1 \cdot \left(\boldsymbol{L}_1^{det} + \boldsymbol{L}_1^{diss} \right) \rho_1, \tag{4.158}$$

with $\nabla_1 = (\partial_{\boldsymbol{x}_1}, \partial_{\boldsymbol{v}_1})^\top$. The marginal probability current $\boldsymbol{J}_1$ can be split into a deterministic part

$$\boldsymbol{L}_1^{det} = \begin{pmatrix} \boldsymbol{v}_1 \\ \frac{1}{m_1} \left[-\partial_{\boldsymbol{x}_1} u_1(\boldsymbol{x}_1, t) + \boldsymbol{f}_1(\boldsymbol{x}_1, t) + \boldsymbol{f}^{(1)}(\boldsymbol{\Gamma}_1, t) \right] \end{pmatrix}, \tag{4.159}$$

and a dissipative one

$$\boldsymbol{L}_1^{diss} = \begin{pmatrix} 0 \\ \frac{-\xi_1}{m_1^2} (m_1 \boldsymbol{v}_1 + \beta_1^{-1} \partial_{\boldsymbol{v}_1} \ln \rho_1) \end{pmatrix}. \tag{4.160}$$

By comparison with the exact single-particle Fokker-Planck Eq. (4.123), we note that the coarse-graining of the second particle encodes the interaction between the two particles in the effective and nonconservative force imposed on particle one

$$\boldsymbol{f}^{(1)}(\boldsymbol{\Gamma}_1, t) = -\int \mathrm{d}\boldsymbol{\Gamma}_2\, \rho_{2|1}(\boldsymbol{\Gamma}, t)\, \partial_{\boldsymbol{x}_1} u_{12}^{int}(\boldsymbol{x}_1, \boldsymbol{x}_2, t). \tag{4.161}$$

We note that the evolution Eq. (4.158) is not closed since $\boldsymbol{f}^{(1)}$ depends on $\rho_{2|1}$. Thus, solving the effective Fokker-Planck Eq. (4.158) is as difficult as treating the full-Fokker-Planck Eq. (4.147).

Moreover, for specific choices of the interaction potential, the first particle might be considered as an active Brownian particle. In this case, the velocity-dependent nonconservative force (4.161) is interpreted as an additional energy inflow leading to active motion. The latter is described effectively by negative dissipation in the direction of motion with velocity-dependent friction kernels. Some prominent models of active Brownian particles can, for instance, be found in Ref. [170].

4.3.2.2 Effective Thermodynamics

Marginalization

In the following, we attempt to formulate a consistent thermodynamic description for this reduced dynamics. Naively, it is tempting to use as an educated guess the single-particle expressions in Sec. 4.3.1.1 for the reduced dynamics. In this case, the naive entropy balance reads

$$\mathrm{d}_t S_{\mathbf{1}} = \beta_1 \dot{q}^{(1)} + \dot{\Sigma}^{(1)} + \dot{I}_F^{(1)}\,, \tag{4.162}$$

where we use the notation from Eq. (4.135),

$$\dot{I}_F^{(1)} \equiv \frac{1}{m_1} \int \mathrm{d}\boldsymbol{\Gamma}_1\, \rho_1\, \partial_{\boldsymbol{v}_1} \cdot \boldsymbol{f}^{(1)}, \tag{4.163}$$

and denote the single-particle Shannon entropy by

$$S_{\mathbf{1}} = -\int \mathrm{d}\boldsymbol{\Gamma}_1\, \rho_1\, \ln \rho_1, \tag{4.164}$$

which implies for the non-negative effective entropy production rate

$$\dot{\Sigma}^{(1)} = \beta_1 \xi_1 \int \mathrm{d}\boldsymbol{\Gamma}_1\, \rho_1 \left(\boldsymbol{v}_1 + \frac{1}{\beta_1 m_1} \partial_{\boldsymbol{v}_1} \ln \rho_1 \right)^2 \geq 0. \tag{4.165}$$

For reasons that will become clear soon, we however define the effective entropy balance as follows,

$$\mathrm{d}_t S = \beta_1 \dot{Q}^{(1)} + \dot{\Sigma}^{(1)} + \dot{I}_F^{(1)}\,, \tag{4.166}$$

where the effective heat

$$\dot{Q}^{(1)} = \dot{q}^{(1)} + \beta_1^{-1} S_{2|\mathbf{1}}, \tag{4.167}$$

is supplemented by the conditional Shannon entropy

$$S_{2|1} = S - S_1 = -\int \mathrm{d}\boldsymbol{\Gamma}_1\, \rho_1 \int \mathrm{d}\boldsymbol{\Gamma}_2\, \rho_{2|1}\, \ln \rho_{2|1}. \tag{4.168}$$

The difference between the full (4.154) and effective (4.167) heat current can be written as

$$\dot{Q} - \dot{q}^{(1)} = \dot{q}^{(2)} - \beta_1^{-1} S_{2|1}. \tag{4.169}$$

Moreover, the difference between the full (4.157) and the effective entropy production rate (4.165) is given by

$$\dot{\Sigma} - \dot{\Sigma}^{(1)} = \int \mathrm{d}\boldsymbol{\Gamma}_1\, \rho_1\, \dot{\mathbb{\Sigma}}_1, \tag{4.170}$$

with the internal entropy production rate kernel

$$\dot{\mathbb{\Sigma}}_1 = \dot{\mathbb{\Sigma}}_1' + \dot{\mathbb{\Sigma}}_1'', \tag{4.171}$$

that can be split in the following two non-negative contributions

$$\dot{\mathbb{\Sigma}}_1' = \beta_2\, \xi_2 \int \mathrm{d}\boldsymbol{\Gamma}_2\, \rho_{2|1} \left(\boldsymbol{v}_2 + \frac{1}{\beta_2\, m_2} \partial_{\boldsymbol{v}_2} \ln \rho_{2|1} \right)^2 \geq 0 \tag{4.172}$$

$$\dot{\mathbb{\Sigma}}_1'' = \frac{\xi_1}{\beta_1 m_1^2} \int \mathrm{d}\boldsymbol{\Gamma}_2\, \rho_{2|1} \left(\partial_{\boldsymbol{v}_1} \ln \rho_{2|1} \right)^2 \geq 0. \tag{4.173}$$

The first contribution $\dot{\mathbb{\Sigma}}_1'$ is the entropy production rate of the second particle if the coordinates of the first one are fixed, see Eq. (4.157). Conversely, the second contribution $\dot{\mathbb{\Sigma}}_1''$ can be viewed as a contribution to the entropy production rate due to the correlation of the particles as we will see in Eq. (4.181).

An equivalent decomposition to Eq. (4.171) for Markovian master equations was found in Ref. [112]. From the last two equations we deduce that the effective entropy production (rate) always underestimates the physical one

$$\dot{\Sigma} \geq \dot{\Sigma}^{(1)}. \tag{4.174}$$

It is important to note that at this general level it is impossible to fully capture the full thermodynamics solely in terms of properties of the reduced dynamics. The missing contributions require knowledge about the conditional probability $\rho_{2|1}$.

Bipartite System

A second approach to formulate an effective thermodynamics is provided by a bipartite system where the two-particle system is split into two single-particle subsystems. The effective entropic expressions in both subsystems are defined in the same formal way as one

would for a single particle. Subsequently, the sum of the effective entropy balances in both subsystems is compared with the full one of the two-particle system in order to identify the so-called information flows exchanged between the subsystems.

Physically, a bipartite system provides a simple and convenient representation of a Maxwell's demon since the thermodynamic cost of the latter becomes fully accessible [114, 116, 117]. Mathematically, the bipartite structure identifies the non-additive contributions of the full thermodynamic quantities for the two particles. We first note that the additive contributions to the two-particle heat current (4.154) can be rewritten in terms of marginalized probabilities only as follows

$$\dot{q}^{(i)} = -\xi_i \int \mathrm{d}\boldsymbol{\Gamma}_i \rho_i \left(\boldsymbol{v}_i + \frac{1}{\beta_i m_i} \partial_{\boldsymbol{v}_i} \ln \rho_i \right) \boldsymbol{v}_i, \tag{4.175}$$

where the marginal probability ρ_2 is obtained analogously as ρ_1, that is by marginalizing the two-point probability ρ over $\boldsymbol{\Gamma}_1$. Using the last Eq. along with Eqs. (4.129) and (4.154), we see that the following relation holds,

$$\dot{q}^{(i)} = \int \mathrm{d}\boldsymbol{\Gamma}_i \, e_i \, \partial_t \rho_i - \int \mathrm{d}\boldsymbol{\Gamma}_i \; \rho_i \; \boldsymbol{v}_i \cdot \left(\boldsymbol{f}_i + \boldsymbol{f}^{(i)} \right), \tag{4.176}$$

with the nonconservative force $\boldsymbol{f}^{(2)}$

$$\boldsymbol{f}^{(2)}(\boldsymbol{\Gamma}_2, t) = - \int \mathrm{d}\boldsymbol{\Gamma}_1 \, \rho_{1|2}(\boldsymbol{\Gamma}, t) \; \partial_{\boldsymbol{x}_2} u_{12}^{int}(\boldsymbol{x}_1, \boldsymbol{x}_2, t). \tag{4.177}$$

Conversely, the additive contributions $\dot{\sigma}^{(i)}$ to the entropy production rate in Eq. (4.189) can not be represented by marginal distributions only. Therefore, the entropy-balance equations for the subsystems of the bipartite system can not be expressed in terms of its associated degrees of freedom only. We proceed by deriving the non-additive contribution to the entropy and identifying them as the information flow.

To this end, we first define the relative entropy (or Kulback-Leibler divergence) as a statistical measure of the distance between the distributions ρ and $\rho_1\rho_2$ as follows

$$I = D[\rho \,||\, \rho_1\rho_2] = \int \mathrm{d}\boldsymbol{\Gamma} \, \rho \ln \frac{\rho}{\rho_1\rho_2} \geq 0, \tag{4.178}$$

whose non-negativity readily follows from the inequality $\ln \rho \leq \rho - 1$. From Eqs. (4.132) and (4.155) follows that the relative entropy is the non-additive part of the two-particle system entropy, *i.e.*

$$I = S_1 + S_2 - S. \tag{4.179}$$

Physically, this quantity corresponds to the mutual information that is a measure of correlations that quantifies how much one system knows about the other. If I is large, the two

systems are highly correlated, whereas small values of I imply that the two systems know little about each other. The time-derivative of the mutual information

$$\mathrm{d}_t I = \dot{I}^{(2\to1)} + \dot{I}^{(1\to2)}, \tag{4.180}$$

can be split into two directional information flows

$$\dot{I}^{(2\to1)} = \int \mathrm{d}\boldsymbol{\Gamma}_1\, \rho_1 \left(\frac{1}{m_1}\partial_{\boldsymbol{v}_1}\cdot\boldsymbol{f}^{(1)} - \dot{\mathbb{\Sigma}}_1''\right) \tag{4.181}$$

$$\dot{I}^{(1\to2)} = \int \mathrm{d}\boldsymbol{\Gamma}_2\, \rho_2 \left(\frac{1}{m_2}\partial_{\boldsymbol{v}_2}\cdot\boldsymbol{f}^{(2)} - \dot{\mathbb{\Sigma}}_2''\right), \tag{4.182}$$

where we used Eqs. (4.161) and (4.173) in the first equation. In the second equation we used Eq. (4.177) and introduced the integral kernel specifying the difference between the full and the effective entropy production rate for the second particle,

$$\dot{\Sigma} - \dot{\Sigma}^{(2)} = \int \mathrm{d}\boldsymbol{\Gamma}_2\, \rho_2\, \dot{\mathbb{\Sigma}}_2 = \int \mathrm{d}\boldsymbol{\Gamma}_2\, \rho_2 (\dot{\mathbb{\Sigma}}_2' + \dot{\mathbb{\Sigma}}_2''), \tag{4.183}$$

with

$$\dot{\mathbb{\Sigma}}_2' = \beta_1\,\xi_1 \int \mathrm{d}\boldsymbol{\Gamma}_1\, \rho_{1|2} \left(\boldsymbol{v}_1 + \frac{1}{\beta_1\, m_1}\partial_{\boldsymbol{v}_1}\ln\rho_{1|2}\right)^2 \geq 0 \tag{4.184}$$

$$\dot{\mathbb{\Sigma}}_2'' = \frac{\xi_2}{\beta_2 m_2^2} \int \mathrm{d}\boldsymbol{\Gamma}_1\, \rho_{1|2} \big(\partial_{\boldsymbol{v}_2}\ln\rho_{1|2}\big)^2 \geq 0. \tag{4.185}$$

The directional information flows can be interpreted as follows: When $\dot{I}^{(i\to j)} > 0$, the dynamics of particle j increases the mutual information and thus the correlations between the two particles. In other words, j is learning about i and vice versa. Conversely, $\dot{I}^{(i\to j)} < 0$ corresponds to decreasing correlations between the two particles due to the evolution of particle j, which can be interpreted as either information erasure or the conversion of information into energy [114]. We furthermore point out that a positive directional information flow indicates that its force contribution

$$\dot{I}_F^{(i\to j)} \equiv \frac{1}{m_j}\int \mathrm{d}\boldsymbol{\Gamma}_j\, \rho_j\, \partial_{\boldsymbol{v}_j}\cdot\boldsymbol{f}^{(j)}, \tag{4.186}$$

dominates its entropic part

$$\dot{I}_S^{(i\to j)} \equiv -\int \mathrm{d}\boldsymbol{\Gamma}_j\, \rho_j\, \dot{\mathbb{\Sigma}}_j'', \tag{4.187}$$

since the latter is non-positive according to Eq. (4.173). Various other interpretations of these mutual information flows have been discussed in the literature [171–176].

An inspection of Eq. (4.163) reveals that the force contribution of the information flow, $\dot{I}_F^{(i\to j)}$, is the additional term that enters in the effective entropy balance due to the velocity-dependent nonconservative force $\boldsymbol{f}^{(j)}$,

$$\mathrm{d}_t S_j = \beta_j \dot{q}^{(j)} + \dot{\Sigma}^{(j)} + \dot{I}_F^{(i\to j)}. \tag{4.188}$$

Using Eq. (4.187), we furthermore find that the difference between the effective (4.165) and the additive contribution to the two-particle entropy production rate (4.157) corresponds to the entropic part of the information flow,

$$\dot{I}_S^{(i\to j)} = \dot{\Sigma}^{(j)} - \dot{\sigma}^{(j)}. \tag{4.189}$$

The last two equations stipulate the following effective entropy balance equation for particle j,

$$\mathrm{d}_t S_j = \beta_j \dot{q}^{(j)} + \dot{\sigma}^{(j)} + \dot{I}^{(i\to j)}. \tag{4.190}$$

It is important to note that Eq. (4.190) states that the directional information flows are the non-additive quantities entering in the effective entropy balance. We emphasize that Eq. (4.190) is the underdamped Fokker-Planck analogue of the result found for master equations in Ref. [114]. Moreover, using Eqs. (4.172) and (4.189), it holds that

$$\int \mathrm{d}\boldsymbol{\Gamma}_i \, \rho_i \, \dot{\Sigma}_i' = \dot{\sigma}^{(j)} = \dot{\Sigma}^{(j)} - \dot{I}_S^{(i\to j)}, \tag{4.191}$$

which because of Eq. (4.157) implies that

$$\dot{\Sigma} = \dot{\Sigma}^{(1)} + \dot{\Sigma}^{(2)} - \dot{I}_S^{(2\to 1)} - \dot{I}_S^{(1\to 2)}, \tag{4.192}$$

An identical result for bipartite master equations was found in Ref. [177] and recently for the more general case of systems undergoing a quantum dynamics formulated in terms of a density matrix, where the generator is additive with respect to the reservoirs [178].

Hamiltonian of Mean Force

Finally, we present a third approach to define an effective thermodynamics for the reduced dynamics of particle **1** in Fig. 4.5, where we set $\beta_{1,2} = \beta$ and $\boldsymbol{f}_2 = 0$. For reasons that will become clear soon, we furthermore consider an explicitly time-independent bare Hamiltonian of the second particle $\partial_t e_2 = \partial_t u_2 = 0$. As we will see, for this approach only a specific class of initial conditions can be considered.

The key concept is the so-called Hamiltonian of mean force, originally utilized in equilibrium thermostatics [179], which defines an effective energy for particle **1** that accounts for the strong coupling [119] to the second particle **2**. Using it, this approach attempts to overcome the problem identified in the context of Eq. (4.166) that there is *a priori* no systematic way to embed the global energetics into the reduced dynamics.

The marginal of the global (Gibbs) equilibrium distribution over the second particle can be expressed as

$$\rho_1^{hmf} = \int \mathrm{d}\boldsymbol{\Gamma}_2\, \rho^{eq} = \int \mathrm{d}\boldsymbol{\Gamma}_2\, \mathrm{e}^{-\beta(e-A^{eq})} = \mathrm{e}^{-\beta(H^{hmf}-A^{eq}_{hmf})}, \tag{4.193}$$

where we introduced the effective free energy A^{eq}_{hmf} of particle one which is defined as the difference between the full equilibrium free energy

$$A^{eq} = -\frac{1}{\beta}\,\ln \int \mathrm{d}\boldsymbol{\Gamma}\; \mathrm{e}^{-\beta e}, \tag{4.194}$$

and that of the second particle

$$A^{eq}_2 = -\frac{1}{\beta}\,\ln \int \mathrm{d}\boldsymbol{\Gamma}_2\; \mathrm{e}^{-\beta e_2}, \tag{4.195}$$

that is $A^{eq}_{hmf} = A^{eq} - A^{eq}_2$. Consequently the Hamiltonian of mean force is defined as

$$H^{hmf} \equiv e_1 - \beta^{-1}\,\ln\langle \mathrm{e}^{-\beta u^{int}_{12}}\rangle^{eq}_2. \tag{4.196}$$

We denote by $\langle\cdot\rangle^{eq}_2$ and $\langle\cdot\rangle^{eq}$ an ensemble average over the equilibrium distribution of particle two, $\rho^{eq}_2 = \exp[-\beta(e_2 - A^{eq}_2)]$, and over the global equilibrium distribution, respectively.

The conditional equilibrium distribution $\rho^{eq}_{2|1}$ is obtained by dividing the global (Gibbs) equilibrium distribution by the marginal one in Eq. (4.193)

$$\rho^{eq}_{2|1} = \frac{\rho^{eq}}{\rho^{hmf}_1} = \mathrm{e}^{-\beta\left(e-A^{eq}_{2|1}\right)}, \tag{4.197}$$

where the free-energy landscape of particle one for a conditionally equilibrated particle two is

$$A^{eq}_{2|1} = e_1 - \beta^{-1}\,\ln\langle \mathrm{e}^{-\beta u^{int}_{12}}\rangle^{eq}_2 + A^{eq}_2 = H^{hmf} + A^{eq}_2. \tag{4.198}$$

It is noteworthy that $A^{eq}_{2|1}$ is parametrically time-dependent, whereas A^{eq}_2 has no time-dependence due to the choice of a time-independent Hamiltonian e_2. Eq. (4.198) shows that up to A^{eq}_2, the Hamiltonian of mean force is equal to the free energy that the locally equilibrated second particle generates for given coordinates of the first particle.

Furthermore, we note the standard equilibrium identities

$$A^{eq}_{2|1} = E^{eq}_{2|1} - \beta^{-1} S^{eq}_{2|1}, \tag{4.199}$$

$$E^{eq}_{2|1} = \partial_\beta(\beta A^{eq}_{2|1}) = \int \mathrm{d}\boldsymbol{\Gamma}_2\, \rho^{eq}_{2|1}\, e \tag{4.200}$$

$$S^{eq}_{2|1} = \beta^2 \partial_\beta A^{eq}_{2|1} = -\int \mathrm{d}\boldsymbol{\Gamma}_2\, \rho^{eq}_{2|1}\, \ln \rho^{eq}_{2|1}, \tag{4.201}$$

which, using Eq. (4.196), can be rewritten as

$$E^{eq}_{2|1} = \partial_\beta\big[\beta\big(H^{hmf} + A^{eq}_2\big)\big] \tag{4.202}$$

$$S^{eq}_{2|1} = \beta^2\,\partial_\beta\big(H^{hmf} + A^{eq}_2\big). \tag{4.203}$$

Inspired by [118], we employ the Hamiltonian of mean force (4.196) and its derived quantities in Eqs. (4.202) and (4.203) and average them over arbitrary *nonequilibrium* probabilities for particle one, *i.e.*

$$E^{hmf}(t) = \langle \partial_\beta(\beta\, H^{hmf})\rangle(t), \tag{4.204}$$

and

$$S^{hmf}(t) \equiv S_1(t) + \beta^2\langle \partial_\beta\, H^{hmf}\rangle(t), \tag{4.205}$$

where $\langle\cdot\rangle(t)$ refers to an ensemble average over a generic *nonequilibrium* distribution $\rho(t)$. We note that the definition of the entropy (4.205) also includes the single-particle Shannon entropy of particle one in addition to the contribution that stems from the Hamiltonian of mean force. Choosing a definition of work that coincides with the global one (4.153),

$$W^{hmf}(t) \equiv \int_0^t \mathrm{d}t'\,\Big[\,\langle \mathrm{d}_t e\rangle(t') + \Big(\int \mathrm{d}\boldsymbol{\Gamma}_1\;\rho_1\;\boldsymbol{v}_1\cdot\boldsymbol{f}_1\Big)(t')\Big], \tag{4.206}$$

the first law of thermodynamics imposes the following definition for heat

$$Q^{hmf}(t) = -W(t) + \langle \partial_\beta(\beta\, H^{hmf})\rangle(t) - \langle \partial_\beta(\beta\, H^{hmf})\rangle(0). \tag{4.207}$$

Defining the nonequilibrium free energy to be of the same form as in the standard equilibrium case (4.199),

$$A^{hmf}(t) = E^{hmf}(t) - \frac{S^{hmf}(t)}{\beta} = \langle H^{hmf}\rangle(t) - \frac{S_1(t)}{\beta}, \tag{4.208}$$

we can rewrite the entropy balance

$$\Delta S^{hmf}(t) = \beta Q^{hmf}(t) + \Sigma^{hmf}(t), \tag{4.209}$$

in the form of a second law of thermodynamics as follows

$$\Sigma^{hmf}(t) = \beta\big[W(t) - \Delta A^{hmf}(t)\big] \geq 0. \tag{4.210}$$

In order to prove the non-negativity of this definition for the entropy production [118, 120], an initial condition of the form

$$\rho(0) = \rho_1(0)\,\rho^{eq}_{2|1} = \rho_1(0)\,\mathrm{e}^{-\beta\left(e - H^{hmf} - A^{eq}_2\right)}, \tag{4.211}$$

is required. Indeed, using Eqs. (4.136) and (4.210), we have

$$\Sigma^{hmf}(t) - \Sigma(t) = \beta \left(\Delta A - \Delta A^{hmf}(t) \right). \tag{4.212}$$

Due to the special choice for the initial condition (4.211), Eqs. (4.202) and (4.203) are valid at $t = 0$ so that

$$A(0) - A^{hmf}(0) = A_2^{eq}. \tag{4.213}$$

At later times, Eqs. (4.202) and (4.203) are no longer valid and we need to resort to the definitions (4.205) and (4.204) to obtain

$$A(t) - A^{hmf}(t) = \langle e \rangle(t) - \langle H^{hmf} \rangle(t) + \beta^{-1}[S_1(t) - S(t)]. \tag{4.214}$$

Since the Hamiltonian of mean force can also be expressed as

$$\langle H^{hmf} \rangle(t) = \langle e \rangle(t) + \beta^{-1} \langle \ln \rho_{2|1}^{eq} \rangle - A_2^{eq}, \tag{4.215}$$

we have

$$A(t) - A^{hmf}(t) = A_2^{eq} + \beta^{-1} \left\langle \ln \frac{\rho(t)}{\rho_{2|1}^{eq} \rho_1(t)} \right\rangle, \tag{4.216}$$

and finally arrive at

$$\Sigma^{hmf}(t) - \Sigma(t) = D[\rho(t) \,||\, \rho_{2|1}^{eq} \rho_1(t)] \geq 0. \tag{4.217}$$

Thus, the entropy production based on the Hamiltonian of mean force always overestimates the global two-particle entropy production which, because of Eq. (4.157), proves the inequality in Eq. (4.210). Furthermore, with Eq. (4.174) we obtain the following hierarchies of inequalities

$$\Sigma^{hmf}(t) \geq \Sigma(t) \geq \Sigma^{(1)}(t), \tag{4.218}$$

where the equality signs hold in the limit of time-scale separation, as will be shown further below. The last equation is the Fokker-Planck analogue of the result found for master equations in Ref. [120]. This reference also identifies the conditions under which the *rate* of the entropy production (4.210) is non-negative.

4.3.2.3 Limiting Cases

As already pointed out above, the effective Fokker-Planck Eq. (4.158) is, in general, not closed because of the dependence on the conditional probability $\rho_{2|1}$. With the results of the preceding section at hand, we now study the three different coarse-graining schemes for two limiting cases in which the effective Fokker-Planck equation becomes closed and thus analytically tractable.

Fast-Dynamics Limit: The Heat Reservoir

First, we assume a time-scale separation between the stochastic dynamics of the two particles where particle two evolves much faster than particle one. Hence for fixed coordinates of the first particle, the second generically relaxes towards a nonequilibrium steady state and the stationary conditional probability $\rho^{tss}_{2|1}$ can be determined by solving the fast dynamics for fixed $\boldsymbol{\Gamma}_1$. As a consequence, the effective Fokker-Planck Eq. (4.158) becomes closed and the effective thermodynamics follows from replacing $\rho_{2|1}$ by $\rho^{tss}_{2|1}$ in all expressions in Sec. 4.3.2.2. However, this effective thermodynamics naturally does not match with the full one, as we would neglect hidden degrees of freedom that are out-of-equilibrium.

The latter equilibrate only if $\boldsymbol{f}_2 = 0$ and $\beta_{1,2} = \beta$, that is when the second particle instantaneously equilibrates with respect to each value of the slow coordinates of particle one. Then, the conditional probability is given at any time by the Gibbs distribution [112]

$$\rho^{tss}_{2|1}(\boldsymbol{x}_1, \boldsymbol{\Gamma}_2) \equiv \rho^{eq}_{2|1}(\boldsymbol{x}_1, \boldsymbol{\Gamma}_2) = \mathrm{e}^{-\beta\left(e - A^{eq}_{2|1}\right)}. \tag{4.219}$$

As a result, the effective force $\boldsymbol{f}^{(1)}$ in Eq. (4.161), becomes a velocity-independent force that derives from an effective potential so that

$$\left[-\partial_{\boldsymbol{x}_1} u_1 + \boldsymbol{f}^{(1)} \right]\Big|_{tss} = -\partial_{\boldsymbol{x}_1} A^{eq}_{2|1}, \tag{4.220}$$

where the notation $Z|_{tss}$ corresponds to the conditional probability $\rho_{2|1}$ in the expression Z being substituted by the equilibrium one in Eq. (4.219). Hence in the limit of time-scale separation and local equilibrium, the particle is subjected to the effective potential given by the free-energy landscape of the first particle, $A^{eq}_{2|1}$.

Marginalization. Substituting Eq. (4.219) into Eq. (4.168) and accounting for probability conservation, we get

$$\frac{\mathrm{d}_t S_{2|1}\big|_{tss}}{\beta} = \int \mathrm{d}\boldsymbol{\Gamma}\, \partial_t \rho_1(t)\, \rho^{eq}_{2|1} \left(e - A^{eq}_{2|1}\right) \tag{4.221}$$

$$= \int \mathrm{d}\boldsymbol{\Gamma}\, \partial_t \rho_1(t)\, \rho^{eq}_{2|1}\, e - \int \mathrm{d}\boldsymbol{\Gamma}_1\, \partial_t \rho_1(t)\, A^{eq}_{2|1}. \tag{4.222}$$

With Eqs. (4.176) and (4.198), we note the relation

$$\dot{q}^{(1)}\Big|_{tss} = \int \mathrm{d}\boldsymbol{\Gamma}_1\, \partial_t \rho_1(t)\, A^{eq}_{2|1}, \tag{4.223}$$

from which along with Eq. (4.167) follows that

$$\dot{Q}^{(1)}\Big|_{tss} = \dot{q}^{(1)}\Big|_{tss} + \beta^{-1} \mathrm{d}_t S_{2|1}\big|_{tss} = \dot{Q}\big|_{tss}\,, \tag{4.224}$$

hence clarifying why the effective heat (4.167) was defined to contain the conditional Shannon entropy.

We have therefore proven that in the limit of time-scale separation the effective (4.167) and the global heat current (4.152) coincide and the first law of thermodynamics remains formally the same as in Eq. (4.151),

$$\mathrm{d}_t E|_{tss} = \dot{Q}^{(1)}\Big|_{tss} + \dot{W}|_{tss} = \dot{Q}\big|_{tss} + \dot{W}\big|_{tss}\,. \tag{4.225}$$

Furthermore, in the limit of time-scale separation, the time-dependence of all quantities stems only from the dynamics of particle one and the parametric time-dependence of the Hamiltonian. Equation (4.224) proves that the second law of thermodynamics formally also remains the same as in Eq. (4.157),

$$\dot{\Sigma}^{(1)}\Big|_{tss} = \mathrm{d}_t S|_{tss} - \beta\dot{Q}^{(1)}\Big|_{tss} \tag{4.226}$$

$$= \mathrm{d}_t S|_{tss} - \beta\dot{Q}\big|_{tss} = \dot{\Sigma}\big|_{tss} \geq 0. \tag{4.227}$$

Hence in the limit of time-scale separation, the full thermodynamics of the two particles can be described solely by the reduced dynamics of a single particle that is subjected to the potential $A^{eq}_{2|1}$. Physically, the second particle can be viewed as being part of the heat reservoir the first particle is coupled to.

Bipartite System. Furthermore, substituting (4.219) into Eqs. (4.181), (4.186) and (4.187), gives a vanishing directional information flow from the fast to the slow particle,

$$\dot{I}_F^{(2\to1)}\Big|_{tss} = \dot{I}_S^{(2\to1)}\Big|_{tss} = \dot{I}^{(2\to1)}\Big|_{tss} = 0. \tag{4.228}$$

This means that in the limit of time-scale separation the information flow is completely asymmetric, $\mathrm{d}_t I|_{tss} = \mathrm{d}_t I^{(1\to2)}\big|_{tss}$. From the last equation follows that the additive and effective entropy production rate (4.165) agrees with the global one (4.157),

$$\dot{\sigma}^{(1)}\Big|_{tss} = \dot{\Sigma}^{(1)}\Big|_{tss} = \dot{\Sigma}\big|_{tss}\,, \tag{4.229}$$

which in turn implies that $\dot{\sigma}^{(2)}\big|_{tss} = 0$. Though, there is a mismatch between the effective entropy balance of the slow particle (4.190) and the full entropy balance (4.156) given by the conditional Shannon entropy,

$$\mathrm{d}_t S_1|_{tss} = \beta\dot{q}^{(1)}\Big|_{tss} + \dot{\sigma}^{(1)}\Big|_{tss} = \mathrm{d}_t S|_{tss} - \mathrm{d}_t S_{2|1}\big|_{tss}\,. \tag{4.230}$$

Moreover, the effective entropy balance of the second particle reads

$$\mathrm{d}_t S_2|_{tss} = \beta\ \mathrm{d}_t q^{(2)}\Big|_{tss} + \mathrm{d}_t I^{(1\to2)}\Big|_{tss}\,, \tag{4.231}$$

that can be rewritten as

$$\mathrm{d}_t S_2|_{tss} - \mathrm{d}_t S_{2|1}\big|_{tss} = \mathrm{d}_t I^{(1\to 2)}\Big|_{tss}. \tag{4.232}$$

Equation (4.232) stipulates that the information flow $\mathrm{d}_t I|_{tss} = \mathrm{d}_t I^{(1\to 2)}\big|_{tss}$ from the slow to the fast particle does, in general, not vanish. This is physically plausible since the particles are still correlated. The information flow $\mathrm{d}_t I^{(1\to 2)}\big|_{tss}$ reflects time-varying correlations between the two particles due to the change of the probability distribution of both out-of-equilibrium particles. Consequently, the information flow is zero for a global equilibrium state characterized by $\rho^{eq} = \rho^{eq}_{2|1}\,\rho^{eq}_1$.

Hamiltonian of Mean Force. We now turn to the Hamiltonian of mean force formalism in the limit of time-scale separation and local equilibrium, $\beta_{1,2} = \beta$ and $\boldsymbol{f}_2 = 0$. Further, as done above in the introduction of the Hamiltonian of mean force formalism, we assume that the bare Hamiltonian of the second particle is time-independent, $\partial_t\, e_2 = 0$. Because of Eq. (4.219), the requirement of an initial equilibrium conditional probability distribution (4.211) is fulfilled at all times t. Hence Eqs. (4.202) and (4.203) are valid at any time t and a comparison with Eqs. (4.204) and (4.205), respectively, shows that

$$E^{hmf}(t)\big|_{tss} = E(t)|_{tss} - \partial_\beta(\beta\, A_2^{eq}) \tag{4.233}$$

$$S^{hmf}(t)\big|_{tss} = S(t)|_{tss} - \beta^2\,\partial_\beta\, A_2^{eq}. \tag{4.234}$$

This explains the choice of a time-independent Hamiltonian e_2, since in this case A_2^{eq} has no time-dependence. As a result, the Hamiltonian of mean force definitions of the corresponding *currents* coincide with the global ones,

$$\mathrm{d}_t E^{hmf}(t)\big|_{tss} = \mathrm{d}_t E(t)|_{tss} \tag{4.235}$$

$$\mathrm{d}_t S^{hmf}(t)\big|_{tss} = \mathrm{d}_t S(t)|_{tss}\,. \tag{4.236}$$

Moreover, we conclude that an agreement of the definitions for the time-integrated quantities would be achieved in the limit of time-scale separation, if the Hamiltonian of mean force was defined as $H^{hmf^*} \equiv A^{eq}_{2|1}$ which corresponds to the definition $A^{eq^*}_{hmf} \equiv A^{eq}$. In this case, the equivalence of definitions would still be true for a time-dependent Hamiltonian e_2.

By construction, the definitions of work agree [cf. Eqs. (4.153) and (4.206)], thus it follows from Eq. (4.235) that the definitions of heat *current* also coincide

$$\dot{Q}^{hmf}(t)\big|_{tss} = \mathrm{d}_t E^{hmf}(t)\big|_{tss} - \dot{W}(t)\big|_{tss} = \dot{Q}(t)\big|_{tss}\,. \tag{4.237}$$

Since according to Eqs. (4.235) and (4.237) the entropy production *rates* are also identical,

$$\dot{\Sigma}^{hmf}(t)\big|_{tss} = \mathrm{d}_t S^{hmf}(t)\big|_{tss} - \beta\dot{Q}^{hmf}(t)\big|_{tss} = \dot{\Sigma}(t)\big|_{tss}\,, \tag{4.238}$$

we find that at the differential level the Hamiltonian of mean-force formalism captures the full thermodynamics in the limit of time-scale separation. Furthermore, we have proven that in the limit of time-scale separation all definitions of the entropy production rate in Eqs. (4.157), (4.165) and (4.209) are equivalent, *i.e.*

$$\dot{\Sigma}(t)\Big|_{tss} = \dot{\Sigma}^{(1)}(t)\Big|_{tss} = \dot{\sigma}^{(1)}(t)\Big|_{tss} = \dot{\Sigma}^{hmf}(t)\Big|_{tss}\,. \tag{4.239}$$

Together with Eq. (4.226), this proves the equality signs in Eq. (4.218) in the limit of time-scale separation.

This constitutes our first main result: In the limit of time-scale separation and local equilibrium, the effective thermodynamic descriptions resulting from marginalization and the Hamiltonian of mean force formalism fully capture the full thermodynamics. In contrast, the effective bipartite description does not match with the full thermodynamics since it neglects the correlations between the two particles.

Large-Mass Limit: The Work Source

We proceed by studying the limit of a diverging mass of the second particle, $m_2 \to \infty$, that has already been discussed in Sec. 4.3.1.2. In view of active Brownian motion, this limit is interesting since the heavy second particle could represent a passive cargo, while the light particle may be considered active. Again, in order to avoid any triviality we assume that the potentials scale with the mass m_2 as follows: $\mathbb{O}(\partial_{\boldsymbol{x}_{2_i}} u_2/m_2) = 1\,\forall i$ while $\partial_{\boldsymbol{x}_{2_i}} u^{int}/m_2 \to 0\,\forall i$ as $m_2 \to \infty$. Because of the infinite mass of particle two its motion occurs deterministically such that we can neglect the influence of particle one. Consequently, the marginal probabilities become statistically independent and the conditional distribution reads

$$\rho^{det}_{2|1}(\boldsymbol{\Gamma}_2, t) = \rho^{det}_{2}(\boldsymbol{\Gamma}_2, t) = \delta(\boldsymbol{x}_2 - \boldsymbol{x}_t)\,\delta(\boldsymbol{v}_2 - \boldsymbol{v}_t), \tag{4.240}$$

for all times t including the initial time $t = 0$. Here, $\boldsymbol{x}_t$ and $\boldsymbol{v}_t$ are the solutions of the deterministic equations of motion (4.142). As a result, the effective force (4.161) becomes conservative,

$$\boldsymbol{f}^{(1)}(\boldsymbol{x}_1, t)\Big|_{det} = -\partial_{\boldsymbol{x}_1}\, u^{int}_{12}(\boldsymbol{x}_1, \boldsymbol{x}_2, t)\big|_{\boldsymbol{x}_2 = \boldsymbol{x}_t}\,, \tag{4.241}$$

where the notation $Z|_{tss}$ corresponds to the conditional probability $\rho_{2|1}$ in the expression Z being substituted by the delta-correlated one in Eq. (4.240). Thus, we are dealing with a closed effective Fokker-Planck Eq. (4.158) for the light particle one that is externally driven by the deterministic motion of the heavy second particle.

Marginalization. Since the marginal probabilities are statistically independent, the conditional Shannon entropy (4.168) vanishes,

$$S_{2|1}\big|_{det} = 0 \tag{4.242}$$

such that the definition of the effective heat (4.167) reduces to the naive one (4.162), $\dot{Q}^{(1)}\big|_{det} = \dot{q}^{(1)}\big|_{det}$. Therefore, by inserting Eq. (4.240) into Eq. (4.152), we get

$$\dot{Q}\big|_{det} - \dot{q}^{(1)}\Big|_{det} = \dot{q}^{(2)}\Big|_{det} = -\xi_2\, \boldsymbol{v}_t^2. \tag{4.243}$$

Thus, the first law of thermodynamics remains - up to a macroscopic frictional term related to the heavy particle - formally the same as in Eq. (4.151),

$$\mathrm{d}_t E|_{det} = \dot{q}^{(1)}\Big|_{det} + \dot{W}\big|_{det} - \xi_2\, \boldsymbol{v}_t^2 = \dot{Q}\big|_{det} + \dot{W}\big|_{det}\,. \tag{4.244}$$

Here, the difference is that the time-dependence of all quantities comes from the dynamical time-dependence of particle one alone, the parametric time-dependence of the Hamiltonian and from the deterministic trajectory of the second particle $(\boldsymbol{x}_t, \boldsymbol{v}_t)$. Further, Eq. (4.242) implies that the definitions for the single-particle Shannon entropy (4.164) and the full system entropy agree (4.155),

$$\mathrm{d}_t S_1|_{det} = \mathrm{d}_t S|_{det}\,, \tag{4.245}$$

which, in turn, proves that the effective second law of thermodynamics (4.165) - up to a macroscopic frictional term of the heavy particle - formally also remains the same as in Eq. (4.157),

$$\dot{\Sigma}^{(1)}\Big|_{det} = \mathrm{d}_t S_1|_{det} - \beta_1 \dot{q}^{(1)}\Big|_{det} = \dot{\tilde{\Sigma}}\Big|_{det}\,, \tag{4.246}$$

where $\dot{\tilde{\Sigma}}\Big|_{det} = \dot{\Sigma}\big|_{det} - \beta_2\, \xi_2\, \boldsymbol{v}_t^2$. The effective thermodynamic description for the two particles therefore reduces, up to a simple macroscopic term, to the standard one of a single particle that is subjected to an external driving. Consequently, the physical interpretation of this limit is that the second particle represents a work source that modulates the energy landscape of the first particle according to a protocol $(\boldsymbol{x}_t, \boldsymbol{v}_t)$. If the deterministic particle is furthermore Hamiltonian, $\xi_2 = 0$, the work source is non-dissipative and the effective description coincides with the full one.

Bipartite System. Owing to the statistical independence of the marginal distributions, the mutual information (4.178) and thus the information flow is identically zero,

$$I|_{det} = \dot{I}^{2\to1}\big|_{det} = \dot{I}^{1\to2}\big|_{det} = 0. \tag{4.247}$$

As a result, the effective entropy balance of the light particle coincides with the full one,

$$\mathrm{d}_t S_1|_{det} = \beta_1 \dot{q}^{(1)}\Big|_{det} + \dot{\sigma}^{(1)}\Big|_{det} = \dot{Q}\big|_{det} + \dot{\Sigma}\big|_{det} = \mathrm{d}_t S|_{det}\,, \tag{4.248}$$

while the corresponding effective entropy balance equation for the heavy particle takes the simple macroscopic form

$$\beta_2\, \dot{q}^{(2)}\Big|_{det} = -\dot{\sigma}^{(2)}\Big|_{det} = -\beta_2\, \xi_2\, \boldsymbol{v}_t^2. \tag{4.249}$$

Hamiltonian of Mean Force. The large-mass limit represents a special case of systems away from time-scale separation. Yet, the assumption of a conditional Gibbs state (4.211) is inconsistent with the independent single-particle distributions (4.240). Therefore, the Hamiltonian of mean force formalism and the deterministic limit are incompatible.

We can therefore summarize our second main result: In the deterministic limit, the effective thermodynamics of the first two coarse-graining schemes - marginalization and bipartite structure - are, up to a simple macroscopic frictional term, equivalent to the full thermodynamics. In contrast, the Hamiltonian of mean force formalism is incompatible with the deterministic limit. In fact, the Hamiltonian of mean force thermodynamics only matches with the full one in the limit of time-scale separation. This is not surprising since the Hamiltonian of mean force definitions [cf. Eqs. (4.196) and (4.211)] are motivated by equilibrium thermostatics. Notably, in the time-scale separation limit there is a completely asymmetric information flow from the slow to the fast particle, while in the deterministic limit all information flows vanish.

4.3.3 Two Linearly Coupled Harmonic Oscillators

4.3.3.1 Full Solution

In this section, the results derived above are illustrated for an analytically solvable example. For this purpose, we consider an isothermal version of the setup in Fig. 4.5 in one dimension. Moreover, the Hamiltonian (4.144) is assumed time-independent

$$u(x_1, x_2) = (k_1 x_1^2)/2 + (k_2 x_2^2)/2 + \beta(x_1 x_2), \tag{4.250}$$

and the nonconservative forces $\boldsymbol{f}_i$ taken zero. Consequently, there is no work done on or by the two-particle system, $\mathrm{d}_t E = \mathrm{d}_t Q$. The Fokker-Planck Eq. (4.147) reads

$$\partial_t \rho = -\nabla \cdot (\boldsymbol{\gamma} \cdot \boldsymbol{\Gamma} \rho) + \nabla^\top \cdot \big(\boldsymbol{D} \cdot \nabla \rho\big), \tag{4.251}$$

with $\boldsymbol{\Gamma} = (x_1, v_1, x_2, v_2)^\top$ and $\nabla \equiv (\partial_{x_1}, \partial_{v_1}, \partial_{x_2}, \partial_{v_2})^\top$. The constant drift coefficient and diffusion matrix read, respectively,

$$\boldsymbol{\gamma} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\frac{k_1}{m_1} & -\frac{\xi_1}{m_1} & -\frac{\beta}{m_1} & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{\beta}{m_2} & 0 & -\frac{k_2}{m_2} & -\frac{\xi_2}{m_2} \end{pmatrix} \tag{4.252}$$

$$\boldsymbol{D} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{\xi_1}{\beta m_1^2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\xi_2}{\beta m_2^2} \end{pmatrix}. \tag{4.253}$$

This partial differential equation is supplemented by the initial condition $\rho(0) = \delta(\boldsymbol{\Gamma}(t)-\boldsymbol{\Gamma}(0))$. The solution of this Fokker-Planck equation is given by a Gaussian [180]

$$\rho = \frac{1}{(2\pi)^2\sqrt{\det \boldsymbol{\Upsilon}}} \exp\left[-\frac{1}{2}(\boldsymbol{\Gamma}-\langle\boldsymbol{\Gamma}\rangle)^\top\cdot\boldsymbol{\Upsilon}^{-1}\cdot(\boldsymbol{\Gamma}-\langle\boldsymbol{\Gamma}\rangle)\right], \tag{4.254}$$

where the average values of the coordinates are determined as follows

$$\langle\boldsymbol{\Gamma}\rangle(t) = \mathrm{e}^{\boldsymbol{\gamma} t}\cdot\boldsymbol{\Gamma}(0), \tag{4.255}$$

and the covariance matrix is calculated as

$$\Upsilon_{kl}(t) \equiv 2\sum_{i,j}\frac{1-\mathrm{e}^{-(\lambda_i+\lambda_j)t}}{\lambda_i+\lambda_j} C_{ij}\, u_i^{(k)} u_j^{(l)}. \tag{4.256}$$

Here, we introduced the transformation matrix

$$\boldsymbol{C} = \boldsymbol{V}\cdot\boldsymbol{D}\cdot\boldsymbol{V}^\top, \quad \boldsymbol{V} = \left(\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \boldsymbol{v}^{(3)}, \boldsymbol{v}^{(4)}\right), \tag{4.257}$$

where λ_i and $\boldsymbol{u}^{(i)}$ ($\boldsymbol{v}^{(i)}$) denote the ith eigenvale and right (left) eigenvector of the drift coefficient matrix $\boldsymbol{\gamma}$, respectively, *i.e.*

$$\begin{aligned}\boldsymbol{\gamma}\cdot\boldsymbol{u}^{(i)} &= \lambda_i\,\boldsymbol{u}^{(i)}\\ \boldsymbol{v}^{(i)}\cdot\boldsymbol{\gamma} &= \lambda_i\,\boldsymbol{v}^{(i)},\end{aligned} \tag{4.258}$$

such that the left and right eigenvectors of $\boldsymbol{\gamma}$ constitute an orthonormal dual basis, $\boldsymbol{v}^{(i)}\cdot\boldsymbol{u}^{(j)} = \delta_{ij}$. Substituting Eq. (4.254) into Eqs. (4.154) and (4.157), we obtain for the heat current and the entropy production rate

$$\dot{Q} = \sum_{i=1}^{2}\left[-\xi_i\left(\Upsilon_{2i,2i} + \langle\Gamma_{2i}\rangle^2\right) + \frac{\xi_i}{\beta m_i}\right] = \sum_{i=1}^{2}\dot{q}^{(i)} \tag{4.259}$$

$$\dot{\Sigma} = \sum_{i=1}^{2}\left[\beta\,\xi_i\left(\Upsilon_{2i,2i} + \langle\Gamma_{2i}\rangle^2\right) - 2\frac{\xi_i}{m_i} + \frac{\xi_i}{\beta m_i^2}\Upsilon^{-1}_{2i,2i}\right] = \sum_{i=1}^{2}\dot{\sigma}^{(i)}, \tag{4.260}$$

and because of Eq. (4.156)

$$\mathrm{d}_t S(t) = \sum_{i=1}^{2}\left(\frac{\xi_i}{\beta\, m_i^2}\Upsilon^{-1}_{2i,2i} - \frac{\xi_i}{m_i}\right). \tag{4.261}$$

In the following, the distribution for particle one $\rho_1(t)$ is needed. The latter is readily determined by marginalizing Eq. (4.254) over the coordinates $\boldsymbol{\Gamma}_2$ of the second particle,

$$\rho_1 = \frac{1}{2\pi\sqrt{\det\tilde{\boldsymbol{\Upsilon}}}}\ \exp\left[-\frac{1}{2}(\tilde{\boldsymbol{\Gamma}}-\langle\tilde{\boldsymbol{\Gamma}}\rangle)^\top\cdot\tilde{\boldsymbol{\Upsilon}}^{-1}\cdot(\tilde{\boldsymbol{\Gamma}}-\langle\tilde{\boldsymbol{\Gamma}}\rangle)\right], \tag{4.262}$$

with $\tilde{\boldsymbol{\Gamma}} = (x_1, v_1)^\top$ and the inverse of the marginalized covariance matrix $\tilde{\boldsymbol{\Upsilon}}$ that is given by

$$\begin{aligned}
\tilde{\Upsilon}_{11}^{-1} &= \frac{1}{\left(\Upsilon_{34}^{-1}\right)^2 - \Upsilon_{33}^{-1}\Upsilon_{44}^{-1}}\left[\left(\Upsilon_{14}^{-1}\right)^2\Upsilon_{33}^{-1} - 2\Upsilon_{13}^{-1}\Upsilon_{14}^{-1}\Upsilon_{34}^{-1} + \Upsilon_{11}^{-1}\left(\Upsilon_{34}^{-1}\right)^2 + \left(\Upsilon_{13}^{-1}\right)^2\Upsilon_{44}^{-1} - \Upsilon_{11}^{-1}\Upsilon_{33}^{-1}\Upsilon_{44}^{-1}\right] \\
\tilde{\Upsilon}_{12}^{-1} &= \frac{1}{\left(\Upsilon_{34}^{-1}\right)^2 - \Upsilon_{33}^{-1}\Upsilon_{44}^{-1}}\left[\Upsilon_{14}^{-1}\Upsilon_{24}^{-1}\Upsilon_{33}^{-1} - \Upsilon_{14}^{-1}\Upsilon_{23}^{-1}\Upsilon_{34}^{-1} - \Upsilon_{13}^{-1}\Upsilon_{24}^{-1}\Upsilon_{34}^{-1} + \Upsilon_{12}^{-1}\left(\Upsilon_{34}^{-1}\right)^2 + \Upsilon_{13}^{-1}\Upsilon_{23}^{-1}\Upsilon_{44}^{-1} - \Upsilon_{12}^{-1}\Upsilon_{33}^{-1}\Upsilon_{44}^{-1}\right] \\
\tilde{\Upsilon}_{22}^{-1} &= \frac{1}{\left(\Upsilon_{34}^{-1}\right)^2 - \Upsilon_{33}^{-1}\Upsilon_{44}^{-1}}\left[\left(\Upsilon_{24}^{-1}\right)^2\Upsilon_{33}^{-1} - 2\Upsilon_{23}^{-1}\Upsilon_{24}^{-1}\Upsilon_{34}^{-1} + \Upsilon_{22}^{-1}\left(\Upsilon_{34}^{-1}\right)^2 + \left(\Upsilon_{23}^{-1}\right)^2\Upsilon_{44}^{-1} - \Upsilon_{22}^{-1}\Upsilon_{33}^{-1}\Upsilon_{44}^{-1}\right].
\end{aligned} \tag{4.263}$$

Inserting Eq. (4.262) into Eqs. (4.161) and (4.186), gives the force contribution to the information flow from particle two to one

$$\dot{I}_F^{(2\to1)} = -\frac{\beta}{m_1}\left(\tilde{\Upsilon}_{12}^{-1}\,\Upsilon_{13} + \tilde{\Upsilon}_{22}^{-1}\,\Upsilon_{23}\right), \tag{4.264}$$

which can be seen by noting that

$$-\frac{\beta}{m_1}\int \mathrm{d}\boldsymbol{\Gamma}\,\rho_1\,x_2\,\partial_{v_1}\rho_{2|1} = \frac{\beta}{m_1}\int \mathrm{d}\boldsymbol{\Gamma}\,\rho\,x_2\,\partial_{v_1}\ln\rho_1. \tag{4.265}$$

Moreover, from Eq. (4.165) follows for the effective entropy production rate

$$\dot{\Sigma}^{(1)} = \beta\,\xi_1\left(\tilde{\Upsilon}_{22} + \langle\Gamma_2\rangle^2\right) - 2\frac{\xi_1}{m_1} + \frac{\xi_1}{\beta m_1^2}\tilde{\Upsilon}_{22}^{-1}, \tag{4.266}$$

from which via Eqs. (4.189) and (4.260) we get the entropic contribution to the information flow

$$\dot{I}_S^{(2\to1)} = \beta\,\xi_1\left(\tilde{\Upsilon}_{22} - \Upsilon_{22}\right) + \frac{\xi_1}{\beta m_1^2}\left(\tilde{\Upsilon}_{22}^{-1} - \Upsilon_{22}^{-1}\right). \tag{4.267}$$

Combining the last three equations with Eqs. (4.166), (4.167) and (4.168), yields

$$\dot{I}^{(2\to1)} = \beta\,\xi_1\left(\tilde{\Upsilon}_{22} - \Upsilon_{22}\right) + \frac{\xi_1}{\beta m_1^2}\left(\tilde{\Upsilon}_{22}^{-1} - \Upsilon_{22}^{-1}\right) - \frac{\beta}{m_1}\left(\tilde{\Upsilon}_{12}^{-1}\,\Upsilon_{13} + \tilde{\Upsilon}_{22}^{-1}\,\Upsilon_{23}\right) \tag{4.268}$$

$$\mathrm{d}_t\,S_{2|\mathbf{1}} = \frac{\xi_2}{\beta m_2^2}\Upsilon_{44}^{-1} - \frac{\xi_2}{m_2} - \frac{\beta}{m_1}\left(\tilde{\Upsilon}_{12}^{-1}\,\Upsilon_{13} + \tilde{\Upsilon}_{22}^{-1}\,\Upsilon_{23}\right) - \beta\,\xi_1\left(\tilde{\Upsilon}_{22} - \Upsilon_{22}\right) - \frac{\xi_1}{\beta m_1^2}\left(\tilde{\Upsilon}_{22}^{-1} - \Upsilon_{22}^{-1}\right) \tag{4.269}$$

$$\dot{Q}^{(1)} = \frac{\xi_2}{\beta m_2^2}\Upsilon_{44}^{-1} + \frac{\xi_1}{m_1} - \frac{\xi_2}{m_2} - \frac{\beta}{m_1}\left(\tilde{\Upsilon}_{12}^{-1}\,\Upsilon_{13} + \tilde{\Upsilon}_{22}^{-1}\,\Upsilon_{23}\right) - \beta\,\xi_1\left(\tilde{\Upsilon}_{22} + \langle\Gamma_2\rangle^2\right) - \frac{\xi_1}{\beta m_1^2}\left(\tilde{\Upsilon}_{22}^{-1} - \Upsilon_{22}^{-1}\right). \tag{4.270}$$

4.3.3.2 Fast-Dynamics Limit

Since $f_2 = 0$, the limit of time-scale separation implies that the second particle is at local equilibrium conditioned on the coordinates of particle one. Within time-scale separation, the effective force (4.220) reads

$$f^{(1)} = \beta^2 \frac{x_1}{k_2}, \tag{4.271}$$

and closes the effective Fokker-Planck Eq. (4.158),

$$\partial_t \rho_1 = -\nabla_1 \cdot \left[\left(\boldsymbol{\gamma}_1 \cdot \tilde{\boldsymbol{\Gamma}} \right) \rho_1 \right] + \nabla_1 \cdot \left(\boldsymbol{D}_1 \cdot \nabla_1 \rho_1 \right), \tag{4.272}$$

with $\nabla_1 \equiv (\partial_{x_1}, \partial_{v_1})^\top$. The drift coefficient and the diffusion matrix read

$$\boldsymbol{\gamma}_1 = \begin{pmatrix} 0 & 1 \\ -\frac{k_1}{m_1} - \frac{\beta^2}{k_2 m_1} & -\frac{\xi_1}{m_1} \end{pmatrix}, \quad \boldsymbol{D}_1 = \begin{pmatrix} 0 & 0 \\ 0 & \frac{\xi_1}{\beta m_1^2}, \end{pmatrix}. \tag{4.273}$$

This Fokker-Planck equation implies that we are dealing with a bivariate Ornstein-Uhlenbeck process, thus its solution is given by a bivariate Gaussian [180]

$$\rho_1 = \frac{1}{2\pi \sqrt{\det \tilde{\boldsymbol{\Upsilon}}}} \, \mathrm{e}^{-\frac{1}{2}\left(\tilde{\Gamma} - \langle \tilde{\Gamma} \rangle\right)^\top \cdot \tilde{\Upsilon}^{-1} \cdot \left(\tilde{\Gamma} - \langle \tilde{\Gamma} \rangle\right)}, \tag{4.274}$$

where the covariance matrix $\tilde{\boldsymbol{\Upsilon}}$ is specified by Eq. (4.256) and the averages of the coordinates $\tilde{\boldsymbol{\Gamma}}$ are determined as follows

$$\langle \tilde{\boldsymbol{\Gamma}} \rangle(t) = \mathrm{e}^{\boldsymbol{\gamma}_1 t} \cdot \tilde{\boldsymbol{\Gamma}}(0). \tag{4.275}$$

In the following of this subsection, we employ the numerical values $\xi_1 = 0.8$, $\beta = 0.05$, $k_1 = 1, m_1 = 1$, while we consider three different spring constants k_2 masses m_2 and friction coefficients ξ_2: $(k_2 = 15, m_{2,a} = 5\,,\ \xi_{2,a} = 0.75)$, $(k_2 = 25, m_{2,b} = 7.5\,,\ \xi_{2,b} = 0.25)$ and $(k_2 = 50, m_{2,c} = 10\,,\ \xi_{2,c} = 0.1)$. This choice of parameters corresponds to an increasing separation of the time-scales between the different stochastic dynamics of the two particles. In the order $a - b - c$, the second particle approaches equilibrium conditioned on the coordinates of the first particle: Since the interaction potential scales linearly in the inverse temperature [Eq. (4.250)], we chose a relatively small value for β to implement a weak-coupling condition between the first and second particle - a crucial requisite for the second particle to behave like an ideal heat reservoir [123, 124]. As k_2 and m_2 increases and ξ_2 decreases, the relaxation time-scale of the second particle further shrinks, hence the time-scales of the particles dynamics start to separate, as desired. Moreover, we prepare the initial condition (4.211) with $\rho_1(0) = \delta(\tilde{\boldsymbol{\Gamma}} - \tilde{\boldsymbol{\Gamma}}(0))$ with $\tilde{\boldsymbol{\Gamma}}(0) = (2, 1)^\top$.

Fig. 4.6 depicts in a) the difference between the global $\dot{Q}$ and effective heat current $\dot{Q}^{(1)}$ and in b) the scaled difference between the global $\dot{\Sigma}$ and effective entropy production rate $\dot{\Sigma}^{(1)}$ as a function of time t.

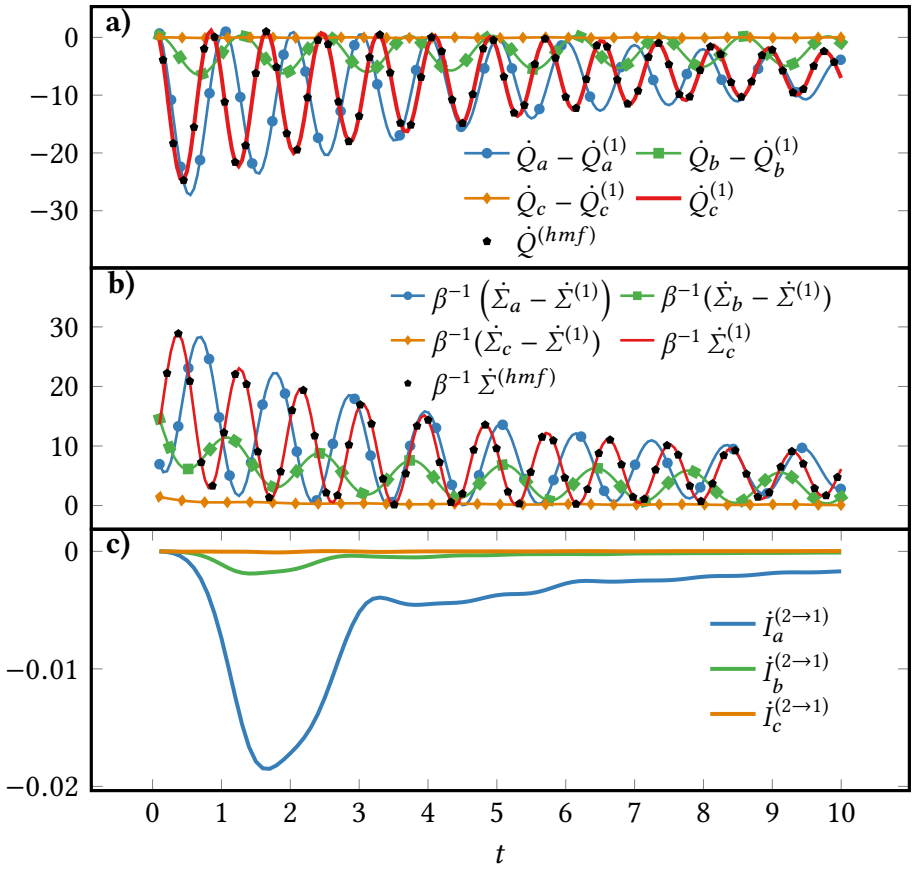

Figure 4.6: *Difference between the full $\dot{Q}$ and effective heat current $\dot{Q}^{(1)}$ in a) and between the scaled full $\beta\,\dot{\Sigma}$ and scaled effective entropy production rate $\beta\,\dot{\Sigma}^{(1)}$ in b) as a function of time t. The information flow $\dot{I}^{(2\to1)}$ is depicted in c). Moreover, the effective quantities based on the Hamiltonian of mean force are overlaid in Figs. a) and b).*

We observe that both the effective heat current and entropy production rate converge to the corresponding full quantities in the limit of time-scale separation. The overall system remains out-of-equilibrium as reflected by finite (effective) heat currents and (effective) entropy production rates of the first particle. Since the corresponding single-particle definition for the heat, $\dot{q}^{(1)}$, does not agree with definition of the effective one [not shown in a)], it follows that the time-derivative of the conditional Shannon entropy, $\mathrm{d}_t S_{2|1}$ remains finite in the limit of time-scale separation. We furthermore note that the effective heat current and entropy production rate are in agreement with the time-derivative of the heat (4.207) and entropy production (4.209) using the Hamiltonian of mean force formalism. Moreover, Fig. 4.6 c) shows that the directional information flow $\dot{I}^{(2\to1)}$ vanishes in the limit of time-scale separation. This in turn implies first that the additive contribution $\dot{\sigma}^{(2)}$ to the full entropy production rate becomes zero while the inverse information flow $\dot{I}^{(1\to2)}$ remains

finite. It furthermore follows from the nonpositivity of $\dot{I}^{(2\to1)}$ that the non-positive entropic contribution $\dot{I}_S^{(2\to1)}$ dominates over the non-negative force contribution $\dot{I}_F^{(2\to1)}$.

4.3.3.3 Large-Mass Limit

In the large-m_2 limit, the effective force (4.241) reads

$$f^{(1)} = -\beta\, x_2|_{x_2=x_t}\,, \tag{4.276}$$

and closes the effective Fokker-Planck Eq. (4.158),

$$\partial_t \rho_1 = -\nabla_1 \cdot \left[\left(\boldsymbol{\gamma}_1 \cdot \tilde{\boldsymbol{\Gamma}} + \boldsymbol{f}^{(1)}\right)\rho_1\right] + \nabla_1 \cdot \left(\boldsymbol{D}_1 \cdot \nabla_1 \rho_1\right). \tag{4.277}$$

The constant drift coefficient, the scaled effective force vector and the diffusion matrix read

$$\boldsymbol{\gamma}_1 = \begin{pmatrix} 0 & 1 \\ -\frac{k_1}{m_1} & -\frac{\xi_1}{m_1} \end{pmatrix},\ \boldsymbol{f}^{(1)} = \begin{pmatrix} 0 \\ -\frac{\beta x_t}{m_1} \end{pmatrix},\ \boldsymbol{D}_1 = \begin{pmatrix} 0 & 0 \\ 0 & \frac{\xi_1}{\beta m_1^2}, \end{pmatrix}. \tag{4.278}$$

This partial differential Eq. is supplemented by the initial condition $\rho_1(0) = \delta(\tilde{\boldsymbol{\Gamma}} - \tilde{\boldsymbol{\Gamma}}(0))$ with $\tilde{\boldsymbol{\Gamma}}(0) = (2,1)^\top$. The averages are determined as follows

$$\langle \tilde{\boldsymbol{\Gamma}} \rangle(t) = \mathrm{e}^{\boldsymbol{\gamma}_1 t} \cdot \tilde{\boldsymbol{\Gamma}}(0) + \int_0^t \mathrm{e}^{\boldsymbol{\gamma}_1 (t-t')} \cdot \boldsymbol{f}^{(1)}(t')\,\mathrm{d}t', \tag{4.279}$$

while the coordinates (x_t, v_t) of the second particle follow the solution of the deterministic equation of motion (4.142),

$$\begin{aligned} x_t &= 2\cos\left(\frac{k_2}{m_2}t\right) + \frac{m_2}{k_2}\sin\left(\frac{k_2}{m_2}t\right) \\ v_t &= \cos\left(\frac{k_2}{m_2}t\right) - 2\frac{k_2}{m_2}\sin\left(\frac{k_2}{m_2}t\right), \end{aligned} \tag{4.280}$$

for the initial condition as chosen above. In the following, we employ the numerical values $\xi_1 = 0.3$, $\xi_2 = 1.5$, $\beta = 1$, $k_1 = 4$, $m_1 = 1$, while we consider three different masses m_2 and constants k_2 such that their ratio remains constant: $(m_{2,a} = 4\,,\ k_{2,a} = 3.8)$, $(m_{2,b} = 40\,,\ k_{2,b} = 38)$ and $(m_{2,c} = 400\,,\ k_{2,c} = 380)$. It is important to note that the set of parameters a, b, and c are chosen such that the ratio of m_2 and k_2 remains constant and thus leaves the determinstic trajectory of the second particle invariant according to Eq. (4.280).

Fig. 4.7 depicts the variances $\boldsymbol{\Upsilon}_{11}$ and $\boldsymbol{\Upsilon}_{33}$ of the positional variables x_1 and x_2, in panels a) and b) respectively. As expected, the fluctuations of the first particle do not exhibit striking qualitative changes since the variance of the second particle vanishes with growing mass m_2. We verify that

$$\boldsymbol{\Upsilon}_{ij} = 0, \quad \forall\, ij \neq \{11, 12, 21, 22\}, \tag{4.281}$$

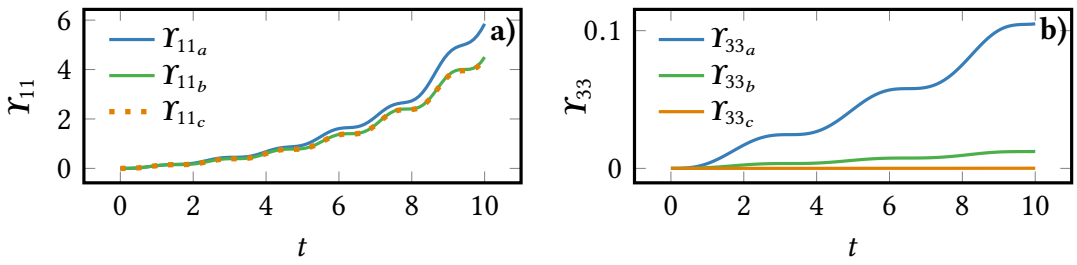

Figure 4.7: *Variance Υ_{11} in a) and Υ_{33} in b) of the positional degrees of freedom x_1 and x_2, respectively, as a function of time t.*

thus confirming that the second particle behaves deterministically in the large-m_2 limit as prescribed by the equations of motion (4.280).

Next, Fig. 4.8 a) shows that the effective heat current, $\dot{Q}^{(1)}$, converges to the full one, $\dot{Q}$, minus the macroscopic dissipation of the heavy particle, $\xi_2 v_t^2$, as m_2 increases.

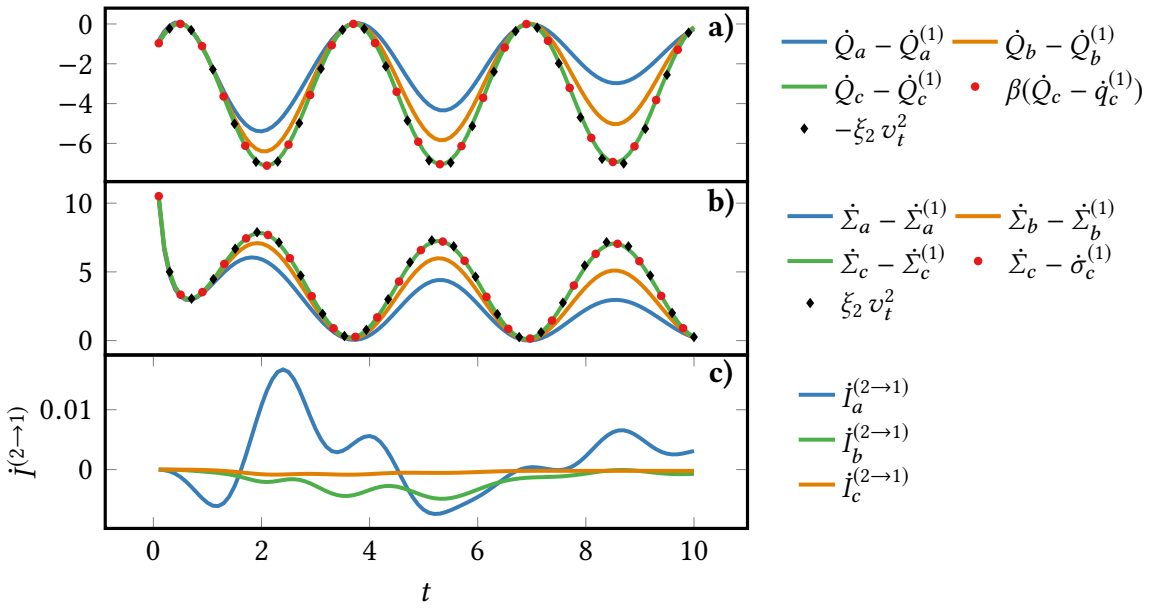

Figure 4.8: *Difference between the full $\dot{Q}$ and effective $\dot{Q}^{(1)}$ heat current in a). Moreover, the heat current associated with the heavy particle, $\dot{q}^{(2)} = -\xi_2 v_t^2$, as well as the difference between $\dot{Q}$ and $\dot{q}^{(1)}$ is overlaid. Fig. b) is analogous to a) but depicting entropy production rates. Information flow $\dot{I}^{(2\to1)}$ is shown in panel c).*

This macroscopic term is naturally non-negative and periodic with the frequency k_2/m_2 due to the choice of a harmonic potential (4.250). Furthermore, Fig. 4.8 b) illustrates the convergence of the effective entropy production $\dot{\Sigma}^{(1)}$ to the full one, $\dot{\Sigma}$, plus the macroscopic

dissipation of the heavy particle with increasing , m_2. Since the single-particle definitions for the heat current, $\dot{q}^{(1)}$, and the entropy production rate, $\dot{\sigma}^{(1)}$, also converge to the full quantities, respectively, it follows that the time-derivative of the conditional Shannon entropy, $\mathrm{d}_t S_{\mathbf{2}|\mathbf{1}}$, and the information flow from the light to the heavy particle, $\dot{I}^{(1\to 2)}$, vanish as m_2 grows. Finally, in Fig. 4.8 c) the directional information flow from the heavy to the light particle $\dot{I}^{(2\to 1)}$ is shown to decrease in modulus with increasing m_2. It is interesting to note that the vanishing directional flow becomes negative if m_2 is sufficiently large. This means that the non-positive entropic contribution converges at a slower rate to zero than the force one does.

Chapter 5

Conclusion and Perspectives

Thermodynamics is the study of energy conversion processes. It is a universal theory in the sense that all energy and matter exchanges have to obey constraints imposed by the laws of thermodynamics. These laws have prevailed throughout the past two centuries, during which the formalism of thermodynamics has changed tremendously from its traditional to its state-of-the-art formulation, known as stochastic thermodynamics. In chapter 2 we briefly reviewed the evolution from a macroscopic and quasistatic thermodynamic - more precisely thermo*static* - theory towards the modern formulation of stochastic thermodynamics. The latter was introduced in detail as it represents the formal pillar on which this dissertation is based on.

Chapter 3 addressed the first main research question of this thesis: How collective effects arising in macroscopic ensembles of microscopic units can affect its thermodynamic performance. For this purpose, we first studied in Sec. 3.2 an open and driven network of interacting three-state units in the framework of stochastic thermodynamics. The stochastic dynamics and ensemble thermodynamics was consistently formulated across different scales, that is at the microscopic, mesoscopic and mean-field level. In the latter case, we discovered a rich phenomenology of this model: a phase characterized by a single stable fixed point in the high-temperature regime, a stable limit cycle indicative of synchronization for intermediate temperatures and a multistability phase at low temperatures. The first and second phase are separated by a Hopf bifurcation, while the second phase transitions into the third phase via an infinite-period bifurcation. This model was shown to be a minimal one as it contains the key ingredients to exhibit synchronization while being thermodynamically consistent at the same time.

Importantly, we resolved the apparent paradox that an irreducible linear Markovian master equation at the microscopic and mesoscopic level, which has a unique stationary state, can asymptotically converge to a nonlinear mean-field equation exhibiting complex

and non-unique solutions at infinite-time. This was achieved by studying the spectrum of the generator of the Markovian dynamics: The two complex-conjugated eigenvalues with the largest finite real part were shown to encode the mean-field dynamics over metastable timescales which increase with the system size N. The predictions based on the spectral analysis were confirmed employing dynamic Monte Carlo simulations.

We formulated the average values of the thermodynamic quantities at the three different scales and then characterized the nonequilibrium phase transitions using the work dissipated by the external driving of the units as a proxy for the total dissipation (entropy production). At the mean-field level, the dissipated work undergoes a first-order phase transition at the Hopf bifurcation followed by a second order one at the infinite-period bifurcation point. Next, we compared the dissipation of a single unit to those of a unit in an interacting network and found that at the mean-field level the dissipation of the units is equal in the first phase. Conversely, compared to a single unit, the dissipation of the interacting unit monotonously drops in the synchronization phase and in the third phase with decreasing temperature. Interestingly, in the presence of interactions and when N is too low to produce a meaningful metastable mean-field dynamics, the average dissipated work in the second (third) phase is lower (higher) than in the mean field ($N \to \infty$). Finally, when operating our system in the mean-field limit as a work-to-work converter, we found that the synchronization phase leads to a significant boost in the power output. The efficiency at maximum power of this far-from-equilibrium machine is surprisingly close to the universal linear-regime prediction.

Subsequently, we generalized the three-state model to a class of isothermal and all-to-all interacting q-state clock models. Using simple thermodynamic arguments, we demonstrated that the high- and low-temperature phase exhibited by the three-state model, a single symmetric and q asymmetric stable fixed points, respectively, are universal for all q. More strikingly, we derived the universal Hopf temperature as a function of q below which the universal high-temperature symmetric fixed point destabilizes into a limit cycle. Though, numerical evidence suggests that only for odd q these limit cycles are stable over time, while for even q they degenerate in the infinite-time limit into the multiple fixed points of the universal low-temperature phase. This classification of the clock models into two classes of universal phenomenology according to the oddness or evenness of q could not be grasped physically, which reflects the nontrivial character of this result. We furthermore investigated the thermodynamic implications of this striking difference in dynamical phenomenology: For systems with odd q the dissipated work undergoes a first-order phase transition at the supercritical Hopf bifurcation followed by a second-order one at the infinite-period bifurcation point, while systems with even q only display the first-order phase transition at the subcritical Hopf bifurcation point. In case of work-to-work conversion processes, the optimal power-efficiency trade-off was shown to be attained in the synchronization regime, that is for odd q.

We furthermore explored the relationship between collective effects and the thermodynamics of information processing in Sec. 3.3. For this purpose, we compared the performance of a single-unit bit and a majority-logic bit that corresponds to an array of binary units under majority-logic decoding. We identified multiple benefits of employing a majority-logic bit

in finite-time information erasure processes: First, the time required to perform a given erasure is reduced. Secondly, for a given erasure duration the resulting erasure error can be narrowed. Finally, for fast erasure processes, the majority-logic unit displays a higher efficiency. A further optimization of the erasure process under majority-logic decoding was observed for the optimal protocol that minimizes the heat dissipation.

Thus, the results presented in chapter 3 suggest that there is a strong analogy between the thermodynamics of energy conversion and information erasure processes. In both cases, collective effects in large macroscopic ensembles - synchronization and decoding schemes - enhance the power-efficiency and precision-speed-efficiency trade-off, respectively, compared to the performance of a single microscopic constituent of that macroscopic ensemble. While we derived these results for several specific models, they may find various applications (*e.g.* interacting molecular motors or coupled quantum dots for the minimal three-state model). We furthermore emphasize that the analytical and computational methods we used are generic in that they can be used on other models.

Motivated by the intrinsic complexity of large many-body systems, chapter 4 studied various methods to consistently coarse-grain systems with many interacting degrees of freedom and can thus be topically considered as the second half of this thesis. Here, we investigated two conceptionally different cases: First, we demonstrated in Sec. 4.2 how to consistently build a stochastic dynamics and thermodynamics description across scales for many-body systems with all-to-all interactions: For this purpose, we generalized the clock models towards a system of N all-to-all interacting identical and classical units consisting of q states. Here, the units are connected to several heat reservoirs and subjected to both autonomous and nonautonomous external forces. We showed that for this class of systems the *microscopic* stochastic dynamics characterized by many-body states can be exactly coarse-grained towards a *mesoscopic* one that is determined by the occupation numbers of the different unit states. The all-to-all interactions give rise to equienergetic many-body states which form the mesostates. Importantly, the coarse-graining significantly reduces the complexity of the many-body system as the growth of the state space changes from an exponential to a power-law one with respect to N. Employing the formalism of stochastic thermodynamics, it was proven that the stochastic first law of thermodynamics is always invariant under the dynamically exact coarse-graining. Conversely, this only holds true for the stochastic entropy balance if the microstates within each mesostate are equiprobable.

We then considered the macroscopic limit, $N \to \infty$. To consistently determine the macroscopic fluctuations (fluctuations that are extensive in and thus scale exponentially with the system size N) the Martin-Siggia-Rose formalism was used. It was demonstrated that the macroscopic fluctuations are thermodynamically consistent as they obey a detailed fluctuation theorem. Detailed fluctuation theorems of the same form were also derived at the microscopic and mesoscopic level, hence proving thermodynamic consistency across scales. Moreover we proved via the path integral representation of the stochastic dynamics that the mesoscopic master equation asymptotically converges to a nonlinear rate equation. The methodology to determine macroscopic fluctuations was demonstrated via a semi-analytically solvable Ising model in contact with two reservoirs and which exhibits a

nonequilibrium phase transition.

Next, we investigated in Sec. 4.3 two interacting and underdamped Brownian particles for which we reconsidered three thermodynamically consistent coarse-graining methods established in the literature for master equations: The observation of only one particle while the other one has been coarse-grained, the partitioning of the two-body system into two single-particle systems exchanging information flows and the Hamiltonian of mean force formalism. We demonstrated that the effective thermodynamics of the first and third approach is equivalent to the correct global thermodynamics in the limit of time-scale separation between the two particles, where the faster evolving particle equilibrates with respect to the coordinates of the more slowly evolving particle. Conversely, we observed a mismatch between the effective and full thermodynamics in the bipartite case, since the entropic contribution due to the coupling of the two particles is not taken into account. Physically, in this limit the faster evolving particle becomes part of the heat reservoir to which the other particle is coupled to. Conversely, if one particle becomes deterministic because of an exceedingly large mass compared to the other particle's mass, it acts as an additional work source on the lighter particle. In this case, the effective thermodynamics of the first two of the aforementioned three approaches agree, up to a simple macroscopic term related to the dissipation of the work source, with the correct global one. The Hamiltonian of mean force formalism however was shown to be incompatible with the large-mass limit. In fact, the same is true for any physical regime outside the time-scale separation limit. This reflects that the Hamiltonian of mean force formalism was originally motivated by and employed in equilibrium thermostatics. These theoretical predictions were confirmed via an analytically tractable model made up of two linearly coupled harmonic oscillators. We remark that the generalization to an arbitrary many-body system, where particle one and two are replaced by two subsets of interacting particles is straightforward and deferred to appendix A.8.

In this doctoral thesis, the results developed during the doctorate were presented. While significant contributions were made to the field of many-body stochastic thermodynamics, the research on this topic is far from complete. Among the many open research questions we want to selectively point out some potential follow-up projects based on the research that was presented in this thesis.

Thermodynamics of Nonequilibrium Phase Transitions

An obvious question is if the phenomenology of the three-state model is retained if the assumption of all-to-all interactions is relaxed. To this end, it should be noted that a closed version of the three-state model with totally irreversible transition rates and all-to-all interactions was investigated as a nonliner dynamics problem in Refs. [85, 86]. Although only the Hopf bifurcation is reported in the latter references, their mean-field solution exhibits in fact the same two phase transitions we observe in the thermodynamically consistent version of the three-state model. Using a tremendous amount of computational resources, it was shown that even for nearest-neighbor interactions on a three-dimensional lattice, the

mean-field characteristics can be reproduced. Thus, it is reasonable to expect that the same holds true in our open version of the three-state model. It might nonetheless be interesting to determine the lower critical dimension for which the mean-field phenomenology can be observed under local interactions.

Another research path worthwhile to pursue is to explore the fluctuating thermodynamics, in particular across the nonequilibrium phase transitions. Numerically, this amounts to extending the dynamic Monte-Carlo simulations to count the statistics of the relevant observables. Here, the path-integral formalism introduced in Sec. (4.2.3) could be a promising strategy to determine the fluctuations in the large N limit. In Sec. 4.2.4 the formalism proved instrumental to semi-analytically compute macroscopic fluctuations in the Ising model. Yet, to which extent analytical progress can be made in the three-state model, in particular in the regime of its non-stationary complex solutions, is *a priori* not clear. Nonetheless, the path-integral formalism represents an alternative numerical method to determine fluctuations in the large N limit.

Of particular interest is to consider non-tightly-coupled energy-converters, that is systems that have more than one net current, in the presence of a synchronous-asynchronous phase transition. At the time of writing, it was unfortunately not possible to find a regime in which the driven three-state model in contact with two heat baths at different temperatures could function like a heat engine *and* still display synchrony. The motivation to build and study such a heat engine converting thermal energy into mechanical energy while being in synchrony is twofold: First such an engine is of higher practical relevance than a system that on the net-level merely dissipates energy taken from the driving into the heat bath. Secondly, since such a system has at least two independent fluctuating currents, the efficiency fluctuations [28] of that energy conversion can be studied. Harnessing critical fluctuations in the vicinity of phase transitions seems to be a promising strategy to enhance the energetic performance of a machine. Indeed, Ref. [181] corroborates that intuition: The authors consider locally interacting discrete rotors on a two-dimensional lattice subdivided into two sublattices each in contact with a heat bath at distinct temperature. At equilibrium, there is a high-temperature disordered and a low-temperature ordered phase separated by a quasi-liquid phase at intermediate temperatures. In the nonequilibrium model a directed rotation of the spin variables can be observed corresponding to a conversion of thermal into mechanical energy. Remarkably, when perturbing the equilibrium in the vicinity of the lower critical temperature, the operation of these autonomous motors is optimal. This suggests that critical fluctuations can indeed be exploited to enhance the performance of a system made up of interacting microscopic machines.

Physical Modeling of Majority-Logic Decoding

There are two possible extensions to the work on the effects of majority-logic decoding on the thermodynamics in information erasure processes. First, it could possibly be interesting to consider a hierarchical system of majority-logic bits and to study if the thermodynamic benefits of majority-logic decoding can be further enhanced by such an iterative approach.

Secondly, as discussed at the end of Sec. 3.3.2.1, the majority-logic decoding can be thought of as a coarse-graining procedure. As a consequence, the logical information stored in the macroscopic bit is, in general, not equal to the sum of microscopic physical information stored in the units contained in the bit. However, the difference between real and physical information was not quantified. In order to corroborate the conclusions made regarding the advantages of majority-logic decoding, these hidden costs need to be accounted for in the various definitions related to the information erasure in the majority-logic bit. A more rigorous approach would possibly be to implement a physical mechanism that mimics the majority-logic decoding, *e.g.* an interacting system of microscopic units that exhibits a phase transition reminiscent of the features of majority-logic decoding. As an educated guess, a two-state model as studied in Sec. 3.2.2 could be used in the context of information erasure. This model is of interest since it has a phase transition that separates a disordered from two ordered phases where all spins tend to occupy one of the two states depending on which state is more populated at initial time.

Coarse-Graining in Stochastic Thermodynamics

In view of applications, the emergence of velocity-dependent nonconservative forces in coarse-grained underdamped Brownian systems found in Sec. 4.3 could give our work significant methodological value for active particles, where the issue of coarse-graining is indeed very important. Active Brownian particles are typically modeled as Brownian particles subjected to nonlinearly velocity-dependent friction. There is thus a practical motivation to identify the conditions, in particular the type of interactions among the Brownian particles, so that the coarse-grained description coincides with typical models of active Brownian particles.

Moreover, to the best of our knowledge, there is no systematic study of an effective fluctuating thermodynamics in underdamped systems. In this context, it would be interesting to explore if coarse-graining methods applied to jump processes, other than the three presented in Sec. 4.3, like for instance the ones proposed in Refs. [29, 182], can also be utilized in underdamped Fokker-Planck systems. These studies could give rise to new strategies for systematically and consistently coarse-graining the thermodynamics of many-body systems.

Appendix A

Appendices

A.1 Derivation of the Fluctuation Theorem (2.105)

In this section we prove the detailed fluctuation theorem (2.105) following a procedure detailed in Ref. [133]. We denote by $p\big(w_\lambda, \{j^\nu_{w_f}\}, \{j^\nu_e\}, m, t\big)$ the joint probability to observe a nonautonomous work contribution w_λ defined in Eq. (2.71) and work and energy current ones equal to the amount $\{j^\nu_{w_f}\}$ and $\{j^\nu_e\}$ defined in Eqs. (2.71) and (2.69), respectively, along a trajectory that is in state m at time t. In the following, we note arrays with bold characters and in case of the generating function $\boldsymbol{g}$ and the associated probability $\boldsymbol{p}$ correspond to different microstates m. According to Eq. (2.75), the microscopic generating function associated with w_λ, $\{j^\nu_{w_f}\}$ and $\{j^\nu_e\}$ is given by

$$\boldsymbol{g}(\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, t) = \int_{-\infty}^{\infty} \prod_\nu \mathrm{d}w_\lambda \; \mathrm{d}j^\nu_{w_f} \; \mathrm{d}j^\nu_e \, \mathrm{e}^{\mathrm{i}\gamma_{w_\lambda} w_\lambda + \mathrm{i}\gamma_{j^\nu_{w_f}} j^\nu_{w_f} + \mathrm{i}\gamma_{j^\nu_e} j^\nu_e} \, \boldsymbol{p}\big(w_\lambda, \{j^\nu_{w_f}\}, \{j^\nu_e\}, t\big), \tag{A.1}$$

and its time evolution is governed by the following biased stochastic dynamics

$$\mathrm{d}_t \, \boldsymbol{g}(\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, t) = \boldsymbol{W}(\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, \lambda_t) \cdot \boldsymbol{g}(\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, t), \tag{A.2}$$

with the biased generator

$$\begin{aligned} W_{mm'}(\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, \lambda_t) = {} & \gamma_{w_\lambda} \left[\dot\lambda_t \cdot \nabla_{\lambda_t} e_m(\lambda_t)\right] \delta_{m,m'} + \\ & + W_{mm'}(\lambda_t) \prod_\nu \mathrm{e}^{\gamma_{j^\nu_{w_f}} j^\nu_{w_{f_{mm'}}} + \gamma_{j^\nu_e} j^\nu_{e_{mm'}}}. \end{aligned} \tag{A.3}$$

We easily verify with Eq. (2.37) that the biased generator satisfies the following symmetry

$$\boldsymbol{W}^{\top}(\gamma_{w_\lambda}, \{\gamma_{j^{\nu}_{w_f}}\}, \{\gamma_{j^{\nu}_{e}}\}, \lambda_t) = \boldsymbol{A}^{-1}(\lambda_t) \cdot \boldsymbol{W}(\gamma_{w_\lambda}, \{1-\gamma_{j^{\nu}_{w_f}}\}, \{1-\gamma_{j^{\nu}_{e}}\}, \lambda_t) \cdot \boldsymbol{A}(\lambda_t), \tag{A.4}$$

with the matrix

$$A_{mm'}(\lambda_t) = \mathrm{e}^{-\beta e_m}\, \delta_{mm'}. \tag{A.5}$$

In this notation, the initial Gibbs state (2.96) can be written as

$$\boldsymbol{g}(\gamma_{w_\lambda}, \{\gamma_{j^{\nu}_{w_f}}\}, \{\gamma_{j^{\nu}_{e}}\}, 0) = \mathbf{p}^{eq}(\lambda_0) = \boldsymbol{A}(\lambda_0) \cdot \mathbf{1}\, \mathrm{e}^{\beta A_1^{eq}(\lambda_0)}, \tag{A.6}$$

where $\mathbf{1}$ refers to a vector whose entries are all unity.

Since it will be useful to proceed later, we proceed by proving a preliminary result. To this end, we consider a generic biased dynamics as in Eq. (A.2)

$$\partial_t\, \boldsymbol{g}(\gamma, t) = \boldsymbol{W}(\gamma, \lambda_t) \cdot \boldsymbol{g}(\gamma, t), \tag{A.7}$$

which has the formal solution

$$\boldsymbol{g}(\gamma, t) = \boldsymbol{U}(\gamma, t) \cdot \boldsymbol{p}(0), \tag{A.8}$$

with the time-evolution operator

$$\boldsymbol{U}(\gamma, t) = T_+\, \mathrm{e}^{\int_0^t \mathrm{d}t'\, \boldsymbol{W}(\gamma, \lambda_{t'})}, \tag{A.9}$$

where T_+ is the time-ordering operator. We define a transformed time-evolution operator

$$\hat{\boldsymbol{U}}(\gamma, t) \;=\; \boldsymbol{B}^{-1}(\lambda_t) \cdot \boldsymbol{U}(\gamma, t) \cdot \boldsymbol{B}(\lambda_0) \tag{A.10}$$

with a generic but invertible operator $\boldsymbol{B}$ and find for its evolution equation

$$\partial_t \hat{\boldsymbol{U}}(\gamma, t) = \left\{ \mathrm{d}_t[\boldsymbol{B}^{-1}(\lambda_t) \cdot \boldsymbol{B}(\lambda_0)] + \boldsymbol{B}^{-1}(\lambda_t) \cdot \boldsymbol{W}(\gamma, \lambda_t) \cdot \boldsymbol{B}(\lambda_t) \right\} \hat{\boldsymbol{U}}(\gamma, t) \equiv \hat{\boldsymbol{W}}(\gamma, \lambda_t) \cdot \hat{\boldsymbol{U}}(\gamma, t), \tag{A.11}$$

which implies for the transformed time-evolution operator

$$\hat{\boldsymbol{U}}(\gamma, t) = T_+\, \mathrm{e}^{\int_0^t \mathrm{d}t'\, \hat{\boldsymbol{W}}(\gamma, \lambda_{t'})}. \tag{A.12}$$

Combining the last three equations, we arrive at the preliminary result

$$\boldsymbol{B}^{-1}(\lambda_t) \cdot \boldsymbol{U}(\gamma, t) \cdot \boldsymbol{B}(\lambda_0) = T_+\, \mathrm{e}^{\int_0^t \mathrm{d}t' \left\{ \mathrm{d}_{t'}[\boldsymbol{B}^{-1}(\lambda_{t'}) \cdot \boldsymbol{B}(\lambda_{t'})] + \boldsymbol{B}^{-1}(\lambda_{t'}) \cdot \boldsymbol{W}(\gamma, \lambda_{t'}) \cdot \boldsymbol{B}(\lambda_{t'}) \right\}}. \tag{A.13}$$

We now return to the specific biased stochastic dynamics considered in Eq. (A.2) and obtain for the generating function

$$\begin{aligned} &g(\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, t) \;= \\ &= \mathbf{1}\cdot \boldsymbol{U}(\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, t)\cdot \boldsymbol{A}(\lambda_0)\cdot \mathbf{1}\, \mathrm{e}^{\beta A_1^{eq}(\lambda_0)} \\ &= \mathrm{e}^{\beta A_1^{eq}(\lambda_t)}\mathbf{1}\cdot \boldsymbol{A}(\lambda_t)\cdot \boldsymbol{A}^{-1}(\lambda_t)\cdot \boldsymbol{U}(\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, t)\cdot \boldsymbol{A}(\lambda_0)\cdot \mathbf{1}\, \mathrm{e}^{-\beta\Delta A_1^{eq}(\lambda)}, \end{aligned} \tag{A.14}$$

where $\boldsymbol{U}(\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, t)$ is the time-evolution operator for that biased stochastic dynamics. Owing to the assumption of initial equilibrium distributions for the forward and backward process, $\boldsymbol{p}^{eq}(\lambda_0)$ and $\boldsymbol{p}^{eq}(\lambda_t)$, we have

$$\mathbf{p}^{eq}(\lambda_t) = \mathbf{1}\cdot \boldsymbol{A}(\lambda_t)\, \mathrm{e}^{\beta A_1^{eq}(\lambda_t)}. \tag{A.15}$$

Substituting the last and Eq. (A.13) into Eq. (A.14) gives

$$\begin{aligned} &g(\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, t) \;= \\ &= \mathbf{p}^{eq}(\lambda_t)\cdot T_+\, \mathrm{e}^{\int_0^t \mathrm{d}t' \left\{\partial_{t'}[\boldsymbol{A}^{-1}(\lambda_{t'})\cdot \boldsymbol{A}(\lambda_{t'})] + \boldsymbol{A}^{-1}(\lambda_{t'})\cdot \boldsymbol{W}(\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, \lambda_{t'})\cdot \boldsymbol{A}(\lambda_{t'})\right\}} \cdot \mathbf{1}\mathrm{e}^{-\beta\Delta A_1^{eq}(\lambda)} \\ &= \mathbf{p}^{eq}(\lambda_t)\cdot T_+\, \mathrm{e}^{\int_0^t \mathrm{d}t'\, \boldsymbol{A}^{-1}(\lambda_{t'})\cdot \boldsymbol{W}(\gamma_{w_\lambda}-1, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, \lambda_{t'})\cdot \boldsymbol{A}(\lambda_{t'})} \cdot \mathbf{1}\, \mathrm{e}^{-\beta\Delta A_1^{eq}(\lambda)} \\ &= \mathbf{p}^{eq}(\lambda_t)\cdot T_+\, \mathrm{e}^{\int_0^t \mathrm{d}t'\, \boldsymbol{W}^\top(\gamma_{w_\lambda}-1, \{1-\gamma_{j^\nu_{w_f}}\}, \{1-\gamma_{j^\nu_e}\}, \lambda_{t'})} \cdot \mathbf{1}\, \mathrm{e}^{-\beta\Delta A_1^{eq}(\lambda)}, \end{aligned} \tag{A.16}$$

where we used Eq. (A.4) in the last equality. Next, we transform the time from t' to $\tilde{t'} = t - t'$ corresponding to a time-reserved process. As a result of this transformation, the time-ordering operator becomes an anti-time-ordering one T_- and the diagonal entries of the biased generator (A.3) become

$$\begin{aligned} W_{mm}(\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, \lambda_{t-\tilde{t'}}) = \gamma_{w_\lambda}\, \dot{\lambda}_{t-\tilde{t'}}\cdot \nabla_{\lambda_{t-\tilde{t'}}} e_m(\lambda_{t-\tilde{t'}}) + W_{mm}(\lambda_{t-\tilde{t'}}) - \\ - \gamma_{w_\lambda}\, \dot{\lambda}_{t-\tilde{t'}}\cdot \nabla_{\lambda_{\tilde{t'}}} e_m(\lambda_{t-\tilde{t'}}) + W_{mm}(\lambda_{t-\tilde{t'}}). \end{aligned} \tag{A.17}$$

Thus, we conclude that

$$\begin{aligned} W_{mm'}(\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, \lambda_{t-\tilde{t'}}) &= W_{mm'}(-\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, \lambda_{t-\tilde{t'}}) \\ &\equiv \tilde{W}_{mm'}(-\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, \tilde{\lambda}_{\tilde{t'}}), \end{aligned} \tag{A.18}$$

where $\tilde{W}_{mm'}(\gamma_{w_\lambda}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, \tilde{\lambda}_{t'})$ is the biased generator of the time-reversed stochastic dynamics that is naturally a function of the time-reversed protocol, $\tilde{\lambda}_{t'} = \lambda_{t-t'}$, $t' \in [0, t]$. Consequently, Eq. (A.16) becomes

$$\begin{aligned} g(\gamma_{w_{\tilde{\lambda}}}, \{\gamma_{j^\nu_{w_f}}\}, \{\gamma_{j^\nu_e}\}, t) &= \tilde{\mathbf{p}}^{eq}(\tilde{\lambda}_0)\cdot T_-\, \mathrm{e}^{\int_0^t \mathrm{d}\tilde{t'}\, \tilde{\boldsymbol{W}}^\top(1-\gamma_{w_\lambda}, \{1-\gamma_{j^\nu_{w_f}}\}, \{1-\gamma_{j^\nu_e}\}, \tilde{\lambda}_{t'})} \cdot \mathbf{1}\, \mathrm{e}^{-\beta\Delta A_1^{eq}(\lambda)} \\ &= \mathbf{1}\cdot T_+\, \mathrm{e}^{\int_0^t \mathrm{d}\tilde{t'}\, \tilde{\boldsymbol{W}}(1-\gamma_{w_\lambda}, \{1-\gamma_{j^\nu_{w_f}}\}, \{1-\gamma_{j^\nu_e}\}, \tilde{\lambda}_{t'})} \cdot \tilde{\mathbf{p}}^{eq}(\tilde{\lambda}_0)\, \mathrm{e}^{-\beta\Delta A_1^{eq}(\lambda)}. \end{aligned} \tag{A.19}$$

In the last equality we applied a global transposition and used the relationship

$$T_{+}\Big(\prod_{i} \boldsymbol{C}(\lambda_{t_i})^{\top}\Big) = \Big(T_{-}\prod_{i} \boldsymbol{C}(\lambda_{t_i})\Big)^{\top}, \tag{A.20}$$

that is valid for a generic operator $\boldsymbol{C}$. Inserting Eq. (A.12) into Eq. (A.19) yields

$$g(\gamma_{w_\lambda}, \{\gamma_{j^{\nu}_{w_f}}\}, \{\gamma_{j^{\nu}_{e}}\}, t) = \mathbf{1}\cdot\tilde{\boldsymbol{U}}(1-\gamma_{w_\lambda}, \{1-\gamma_{j^{\nu}_{w_f}}\}, \{1-\gamma_{j^{\nu}_{e}}\}, t)\cdot\tilde{\mathbf{p}}^{eq}(\tilde{\lambda}_t)\,\mathrm{e}^{-\beta\Delta A_1^{eq}(\lambda)}, \tag{A.21}$$

from which we conclude the following symmetry

$$g(\gamma_{w_\lambda}, \{\gamma_{j^{\nu}_{w_f}}\}, \{\gamma_{j^{\nu}_{e}}\}, t) = \tilde{g}(1-\gamma_{w_\lambda}, \{1-\gamma_{j^{\nu}_{w_f}}\}, \{1-\gamma_{j^{\nu}_{e}}\}, t)\,\mathrm{e}^{-\beta\Delta A_1^{eq}(\lambda)}, \tag{A.22}$$

which via inverse Fourier transformation of the definition (A.1) stipulates the detailed fluctuation theorem

$$\frac{p\Big(\beta^{(1)}\,w_\lambda, \{j^{\nu}_{w_f}\}, \{j^{\nu}_{e}\}\Big)}{\tilde{p}\Big(-\beta^{(1)}\,w_\lambda\,, \{-j^{\nu}_{w_f}\}, \{-j^{\nu}_{e}\}\Big)} = \mathrm{e}^{\beta^{(1)}\left[w_\lambda-\Delta A_1^{eq}\right]+\sum_{\nu=1}^{L}\left[\beta^{(\nu)}w_f^{(\nu)}+\left[\beta^{(1)}-\beta^{(\nu)}\right]e^{(\nu)}\right]}, \tag{A.23}$$

as stated in Eq. (2.105).

A.2 Proof of Supercritical Hopf Bifurcation (3.38)

We shall in the following prove that the Hopf bifurcation is supercritical, *i.e.* results in stable limit cycless. To characterize the limit cycles close to the bifurcation point, we consider the normal form of the Hopf bifurcation. The procedure is detailed in [183]. At first, we transform the two-dimensional system in Eq. (3.31) into a single equation

$$\partial_t z = \lambda(\Delta\beta) z + h(z, z^*, \Delta\beta), \tag{A.24}$$

where z is a complex variable, z^* its complex-conjugate, $\Delta\beta = \beta - \beta_{c_1}$ gives the distance of the inverse temperature to the critical inverse temperature of the Hopf bifurcation and $h = \mathbb{O}\left(||z||^2\right)$ is a smooth function of $(z, z^*, \Delta\beta)$. Such a transformation is achieved by first finding the complex eigenvectors $\boldsymbol{r}$ and $\boldsymbol{v}$ determined by

$$\boldsymbol{J}(0)\boldsymbol{r} = \lambda(0)\boldsymbol{r}, \quad \boldsymbol{J}(0)^\top \boldsymbol{v} = \lambda(0)^* \boldsymbol{v}, \tag{A.25}$$

where the real and non-symmetric Jacobian $\boldsymbol{J}$ resulting from the linearization of Eq. (3.31) is evaluated at the Hopf bifurcation point $\beta = \beta_{c_1}$, yielding

$$\boldsymbol{r} = \left(\frac{1}{2}(-1 + \sqrt{3}\,\mathrm{i}), 1\right)^\top \tag{A.26}$$

$$\boldsymbol{v} = \frac{1}{3 - \sqrt{3}\,\mathrm{i}} \left(1 + \sqrt{3}\,\mathrm{i}, 2\right)^\top . \tag{A.27}$$

If $|\Delta\beta|$ is sufficiently small, the two-dimensional system from Eq. (3.31) can be written as

$$\partial_t \overline{\boldsymbol{n}} = \boldsymbol{J}(\Delta\beta) \cdot \overline{\boldsymbol{n}} + \boldsymbol{F}(\overline{\boldsymbol{n}}, \Delta\beta), \tag{A.28}$$

where $\boldsymbol{H}(\overline{\boldsymbol{n}}, \Delta\beta)$ is a smooth vector function whose components have Taylor expansions in $\overline{\boldsymbol{n}}$ starting with at least quadratic terms, $\boldsymbol{H}_{1,2} = \mathbb{O}\left(||\overline{\boldsymbol{n}}^2\right)$. Using Eq. (A.24) together with the properties $\langle \boldsymbol{v}, \boldsymbol{r} \rangle = 1$ and $\langle \boldsymbol{v}, \boldsymbol{r}^* \rangle = 0$, one can show that

$$h(z, z^*, \Delta\beta) = \langle \boldsymbol{v}(\Delta\beta), \boldsymbol{H}(z\,\boldsymbol{r}(\Delta\beta) + z^* \boldsymbol{r}^*(\Delta\beta), \Delta\beta) \rangle. \tag{A.29}$$

The function h can be formally written as a Taylor series in the two complex variables z and z^*,

$$h(z, z^*, \Delta\beta) = \sum_{k+l \geq 2} \frac{1}{k!l!} \frac{\partial^{k+l}}{\partial z^k \partial {z^*}^l} h_{kl}(\Delta\beta)\, z^k {z^*}^l, \tag{A.30}$$

with

$$h_{kl}(\Delta\beta) = \langle \boldsymbol{v}(\Delta\beta), \boldsymbol{H}(z\boldsymbol{r}(\Delta\beta) + z^* \boldsymbol{r}^*(\Delta\beta), \Delta\beta) \rangle|_{z=0} . \tag{A.31}$$

Moreover, if the function $\boldsymbol{H}(\overline{\boldsymbol{n}}, \Delta\beta)$ from Eq. (A.28) is represented as

$$\boldsymbol{H}(\boldsymbol{x}, 0) = \frac{1}{2}\boldsymbol{A}(\boldsymbol{x}, \boldsymbol{x}) + \frac{1}{6}\boldsymbol{B}(\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{x}) + \mathbb{O}\big(||x||^4\big), \tag{A.32}$$

where $\boldsymbol{A}(\boldsymbol{x}, \boldsymbol{y})$ and $\boldsymbol{B}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u})$ are *symmetric* multilinear vector functions of $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u} \in \mathbb{R}^2$, it follows that

$$g_{20} = \langle \boldsymbol{v}, \boldsymbol{A}(\boldsymbol{r}, \boldsymbol{r})\rangle = 0 \tag{A.33a}$$
$$g_{11} = \langle \boldsymbol{v}, \boldsymbol{A}(\boldsymbol{r}, \boldsymbol{r}^*)\rangle = 0 \tag{A.33b}$$
$$g_{21} = \langle \boldsymbol{v}, \boldsymbol{B}(\boldsymbol{r}, \boldsymbol{r}, \boldsymbol{r})\rangle. \tag{A.33c}$$

In coordinates, one has for these vector functions

$$\boldsymbol{A}_i(\boldsymbol{x}, \boldsymbol{y}) = \sum_{j,k=1}^{2} \left.\frac{\partial^2 \boldsymbol{H}_i(\boldsymbol{x}\boldsymbol{i}, 0)}{\partial \boldsymbol{x}\boldsymbol{i}_j \partial \boldsymbol{x}\boldsymbol{i}_k}\right|_{\boldsymbol{x}\boldsymbol{i}=0} \boldsymbol{x}_j\, \boldsymbol{y}_k, \quad i = 1, 2 \tag{A.34}$$

$$\boldsymbol{B}_i(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}) = \sum_{j,k,l=1}^{2} \left.\frac{\partial^3 \boldsymbol{H}_i(\boldsymbol{x}\boldsymbol{i}, 0)}{\partial \boldsymbol{x}\boldsymbol{i}_j \partial \boldsymbol{x}\boldsymbol{i}_k \partial \boldsymbol{x}\boldsymbol{i}_l}\right|_{\boldsymbol{x}\boldsymbol{i}=0} \boldsymbol{x}_j \boldsymbol{y}_k \boldsymbol{u}_l, \quad i = 1, 2. \tag{A.35}$$

With these expressions at hand, we can determine the first Lyapunov coefficient Λ_1 as

$$\Lambda_1 = \frac{1}{2\,\omega_{lc}^2}\, \mathrm{Re}\left(\mathrm{i}\, g_{20}\, g_{11} + \omega_{lc}\, g_{21}\right), \tag{A.36}$$

where the eigenvalue of the Jacobian (A.25) is decomposed as $\lambda(\Delta\beta)=\sigma(\Delta\beta) + \mathrm{i}\,\omega(\Delta\beta)$, with the limit cycle frequency

$$\omega_{lc} = \lambda(0) = \sqrt{3}\,\Gamma\, \sinh\left(-\frac{3f}{2u}\right). \tag{A.37}$$

For Eq. (A.36) to hold, the two requirements $\omega(0) > 0$ and $\sigma'(0) < 0$ must be met. From Eq. (A.37) and

$$\sigma'(0) = u\,\Gamma\, \cosh\left(\frac{3f}{2u}\right), \tag{A.38}$$

follows that this is requirement is only met for attractive interactions, $u < 0$, which is the condition to observe a Hopf bifurcation (3.38) at all. Collecting results, we finally arrive at

$$\Lambda_1 = -\frac{81}{2}\Gamma \cosh\left(\frac{3f}{2u}\right), \tag{A.39}$$

which is negative for $u < 0$, hence for attractive interactions stable limit cycles emerge at the bifurcation point β_{c_1} as in the main body above.

A.3 Simulation of a Markovian Jump Process

In this section we rigorously derive the dynamic Monte-Carlo techniques employed in this work to numerically simulate the Markov process that the ME (3.14) describes analytically. Although the ME is never solved explicitly, the simulation algorithm is equivalent to the ME. It relies, like the ME itself, only on the fundamental hypothesis that $W_{\boldsymbol{N}\boldsymbol{N}'}\,\mathrm{d}t$ gives the probability, to first order in $\mathrm{d}t$, that there will be a transition from $\boldsymbol{N}'$ to $\boldsymbol{N}$ in the next time interval $\mathrm{d}t$.

Instead of solving the ME numerically, the so-called Gillespie algorithm [146, 147] is based on determining a quantity we shall refer to as the transition probability density $P(\tau,\mu)\,\mathrm{d}\tau$. It is defined as the joint probability at time t that the next transition will occur in the differential time interval $(t+\tau, t+\tau+\mathrm{d}\tau)$ and is a transition from $\boldsymbol{N}'$ to $\boldsymbol{N}$ indexed by μ. As will be demonstrated, this quantity allows us to rigorously construct an algorithm for simulating exactly the time evolution of the probability $P_{\boldsymbol{N}}(t)$ ruled by the ME. The technical details of the algorithm are taken from Ref. [146].

We can calculate the transition probability density

$$P(\tau,\mu) = P_0(\tau)\cdot W_\mu\,\mathrm{d}\tau, \tag{A.40}$$

as the product of the probability at time t that no reaction will occur in the time interval $(t, t+\tau)$, times $W_\mu\,\mathrm{d}\tau$. The probability that more than one reaction occurs is $\mathbb{O}(\mathrm{d}\tau)$ and can therefore be neglected. To determine $P_0(\tau)$, we subdivide the interval $(t, t+\tau)$ into K subintervals of equal length $\Delta t = \frac{\tau}{K}$. The probability, to first order in Δt, that none of the M possible transition occurs in the first subinterval $(t, t+\Delta t)$ is with the fundamental given by

$$\prod_{\nu=1}^{M}\left[1 - W_\nu\Delta t + o(\Delta t)\right] = 1 - \sum_{\nu=1}^{M} W_\nu\,\Delta t + \mathbb{O}(\Delta t). \tag{A.41}$$

We note that this is also the subsequent probability that there is no transition in $(t+\Delta t, t+2\Delta t)$, and then in $(t+2\Delta t, t+3\Delta t)$, and so on. Thus we can write

$$P_0(\tau) = \left[1 - \sum_{\nu=1}^{M} W_\nu\,\Delta t + \big(\Delta t\big)\right]^K = \left[1 - \sum_{\nu=1}^{M} W_\nu\,\frac{\tau}{K} + \mathbb{O}\big(K^{-1}\big)\right]^K. \tag{A.42}$$

In the limit $K\to\infty$, we arrive at

$$P_0(\tau) = \lim_{K\to\infty}\left[1 - \sum_{\nu=1}^{M}\frac{W_\nu\tau}{K} + \frac{\mathbb{O}\big(K^{-1}\big)}{K^2}\right]^K = \mathrm{e}^{-\sum_{\nu=1}^{M} W_\nu\tau}. \tag{A.43}$$

Collecting results, we obtain for the transition probability density

$$P(\tau,\mu) = W_\mu\,\mathrm{e}^{-\sum_{\nu=1}^{M} W_\nu\tau}, \tag{A.44}$$

which is normalized, since

$$\int_0^\infty \mathrm{d}\tau \sum_{\mu=1}^{M} P(\tau,\mu) = \sum_{\mu=1}^{M} W_\mu \int_0^\infty \mathrm{d}\tau\, \mathrm{e}^{-\sum_{\nu=1}^{M} W_\nu \tau} = 1. \tag{A.45}$$

The Gillespie algorithm is a Monte-Carlo method, as it uses uniformly generated pseudo-random numbers r to generate a (pseudo-)random pair (τ,μ) according to the probability density function in Eq. (A.44). While there are several methods to achieve this, we only discuss the so-called *direct method*, as this was also used for all the Monte-Carlo simulations in this work. Here, the idea is to express the joint probability density in the form

$$P(\tau,\mu) = P_1(\tau)\cdot P_2(\mu|\tau), \tag{A.46}$$

where $P_1(\tau)\mathrm{d}\tau$ is the probability that any transition will occur between times $t+\tau$ and $t+\tau+\mathrm{d}\tau$ and $P_2(\mu|\tau)$ is the probability that the next transition will be the one indexed by μ given that it occurs at time $t+\tau$. Consequently, the probability $P_1(\tau)\mathrm{d}\tau$ is obtained by

$$P_1(\tau) = \sum_{\mu=1}^{M} P(\tau,\mu), \tag{A.47}$$

and we find for the conditional probability

$$P_2(\mu|\tau) = \frac{P(\tau,\mu)}{\sum_{\nu=1}^{M} P(\tau,\nu)}. \tag{A.48}$$

Using Eq. (A.44), yields

$$P_1(\tau) = \sum_{\mu=1}^{M} W_\mu\, \mathrm{e}^{-\sum_{\nu=1}^{M} W_\nu \tau} \tag{A.49}$$

$$P_2(\mu|\tau) = \frac{W_\mu}{\sum_{\nu=1}^{M} W_\nu}. \tag{A.50}$$

We easily verify normalization of the distributions and remark that $P_2(\mu|\tau)$ is independent of τ. The crucial idea of the *direct method* is to first generate a random value for τ according to Eq. (A.49) and then to generate a random integer μ according to Eq. (A.50).

To this end, a general Monte-Carlo technique called the *inversion method* is applied. It uses random numbers from the uniform distribution in the unit interval to generate random numbers distributed according to any prescribed probability function $P(x)$. Consider the probability distribution function

$$F(x) = \int_{-\infty}^{x} P(x')\,\mathrm{d}x'. \tag{A.51}$$

A uniformly distributed random number r_1 drawn from the unit interval can be converted into a random number x according to $P(x)$ by taking

$$x = F^{-1}(r_1), \tag{A.52}$$

where F^{-1} is the inverse of the distribution function F. This can be seen by calimit cyclesulating the probability that the x-value will lie between x' and $x' + \mathrm{d}x'$. This probability is equal to the probability that r_1 will lie between $F(x')$ and $F(x' + \mathrm{d}x')$, hence

$$F(x' + \mathrm{d}x') - F(x') = F'(x')\mathrm{d}x' = P(x')\,\mathrm{d}x'. \tag{A.53}$$

It readily follows from Eqs. (A.49) and (A.52) that

$$\tau = \left(\sum_{\mu=1}^{M} W_\mu\right)^{-1} \ln\left(\frac{1}{r_1}\right), \tag{A.54}$$

where we have replaced the random variable $1 - r_1$ by the statistically equivalent random variable r_1. To determine the integer value μ we have to consider the discrete case which is analogous to the continuous case. The distribution function $F(i)$ for the discrete random variable i with the probability density function $P(i)$ reads

$$F(i) = \sum_{i'=-\infty}^{i} P(i'). \tag{A.55}$$

Given a uniformly distributed random number r_2 in the unit interval, an integer i obeying $P(i)$ is found by taking the value for i which satisfies

$$F(i-1) < r_2 \leq F(i), \tag{A.56}$$

since the probability that i takes the value i' is indeed equal to the probability that r_2 lies between $F(i' - 1)$ and $F(i')$

$$F(i') - F(i'-1) = \sum_{i''=-\infty}^{i'} P(i'') - \sum_{i''=-\infty}^{i'-1} P(i'') = P(i'). \tag{A.57}$$

Using Eqs. (A.49) and (A.57), we find as the selection rule for the random integer variable μ

$$\sum_{\nu=1}^{\mu-1} W_\nu < r_2 \sum_{\nu=1}^{M} W_\nu \leq \sum_{\nu=1}^{\mu} W_\nu \tag{A.58}$$

With these theoretical results at hand, the simulation algorithm can be outlined as follows:

Step 1: Initialization of the system in the state $\boldsymbol{N}(0) \equiv \boldsymbol{N}_0$ at time $t = 0$. Set the final time of the simulation t_f and define time marks t_i of equal length.

Step 2: Generation of the random numbers τ and μ according to the joint probability density function $P(\tau, \mu)$ using the Monte-Carlo techniques from above.

Step 3: Update the time $t + \tau$ and the state the system occupies according to the the transition specified by μ.

Step 4: Iteration of Step 2 and 3 until $t > t_i$. The corresponding state is taken to be the representative state the system is occupying at time t_i. Iteration of Step 2,3 and 4 until $t > t_f$.

The Gillespie algorithm is heavily used in cellular chemical kinetics. Owing to its proper account of the discrete, stochastic nature of chemical reactions, it is more accurate than the deterministic reaction-rate equations given by a set of ordinary differential equations. However, this stochastic simulation often turns out to be unpractical as every successive molecular reaction (Monte-Carlo Step) is simulated. An approximate acceleration to the Gillespie algorithm is given by the so-called *tau-leaping*, in which the time steps are fixed by an amount τ and the number of the reactions of the the different molecular species are approximated by Poisson random numbers [184]. For stiff systems[1], τ must be chosen small with respect to the fastest time-scale in the dynamics, which makes even tau-leaping seem too slow. This issue is addressed with the implicit tau-leaping, which mirrors the implicit Euler method, and the slow-scale Gillespie algorithm in which the fastest reactions are skipped over and only the slow reactions are directly simulated using specially modified reaction-rate functions [184].

[1]Stiff systems evolve on well-separated fast and slow time-scales with the fastest dynamic modes being stable.

A.4 The Monotonicity of the Entropy Function (3.140)

The first-order derivative of the function $\mathbb{S}(p,N)$ introduced in Eq. (3.140) with respect to p can be expressed as

$$\frac{\partial \mathbb{S}(p,N)}{\partial p} = N \ln\frac{1-p}{p} - \frac{(1-p)^{\frac{N-1}{2}} p^{\frac{N-1}{2}}}{\mathcal{B}\left(\frac{N+1}{2},\frac{N+1}{2}\right)} \ln\frac{1-h(p,N)}{h(p,N)}, \tag{A.59}$$

where we write the regularized incomplete beta function with symmetric arguments as $h(p,N) \equiv \mathcal{I}_p\left(\frac{N+1}{2},\frac{N+1}{2}\right)$ and introduce the Beta function $\mathcal{B}(a,b) \equiv \int_0^1 t^{a-1}(1-t)^{b-1}\,\mathrm{d}t$. We find for the second-order derivative

$$\begin{aligned}\frac{\partial^2 \mathbb{S}(p,N)}{\partial p^2} = &\frac{(N-1)(1-p)^{\frac{N-3}{2}} p^{\frac{N-3}{2}}(2p-1)\ln\frac{1-h(p,N)}{h(p,N)}}{2\mathcal{B}(\frac{N+1}{2},\frac{N+1}{2})} - \\ &- \frac{N}{p(1-p)}\left[1 - \binom{N}{\frac{N-1}{2}}^2 \frac{N(1-p)^N p^N}{[1-h(p,N)]h(p,N)}\right].\end{aligned} \tag{A.60}$$

We easily verify that the first term on the right-hand side of Eq. (A.60) is non-positive for any p.

In order to determine the sign of the second term on the right-hand side of Eq. (A.60), we first recast the term $[1 - h(p,N)]h(p,N)$ into

$$\begin{aligned}[1-f(p,N)]f(p,N) &= \sum_{i=\frac{N+1}{2}}^{N}\sum_{j=0}^{\frac{N-1}{2}} \binom{N}{i}\binom{N}{j} p^{i+j}(1-p)^{2N-(i+j)} \\ &= \sum_{m=\frac{N+1}{2}}^{N-1}\sum_{i=\frac{N+1}{2}}^{m} \binom{N}{i}\binom{N}{m-i}[p^m(1-p)^{2N-m} + p^{2N-m}(1-p)^m] + \\ &+ \sum_{i=\frac{N+1}{2}}^{N} \binom{N}{i}\binom{N}{N-i} p^N(1-p)^N,\end{aligned} \tag{A.61}$$

where we have introduced the index $m = i + j$. Using the inequality

$$p^s(1-p)^{4n+2-s} + p^{4n+2-s}(1-p)^s \geq 2p^{2n+1}(1-p)^{2n+1}, \tag{A.62}$$

Eq. (A.61) can be transformed into the inequality

$$[1-f(p,N)]\,f(p,N) \geq \left[2\sum_{m=\frac{N+1}{2}}^{N-1}\sum_{i=\frac{N+1}{2}}^{m} \binom{N}{i}\binom{N}{m-i} + \sum_{i=\frac{N+1}{2}}^{N} \binom{N}{i}\binom{N}{N-i}\right] p^N(1-p)^N, \tag{A.63}$$

where the equal sign holds for $p = 1/2$.

Furthermore, using the binomial theorem, we write

$$\begin{aligned} 2^{2N-2} &= \left[\sum_{i=\frac{N+1}{2}}^{N}\binom{N}{i}\right]\left[\sum_{j=0}^{\frac{N-1}{2}}\binom{N}{j}\right] \\ &= \sum_{m=\frac{N+1}{2}}^{N-1}\sum_{i=\frac{N+1}{2}}^{m}\binom{N}{i}\binom{N}{m-i} + \sum_{i=\frac{N+1}{2}}^{N}\binom{N}{i}\binom{N}{N-i} + \sum_{m=N+1}^{\frac{3N-1}{2}}\sum_{i=m-\frac{N-1}{2}}^{N}\binom{N}{i}\binom{N}{m-\frac{N-1}{2}} \\ &= 2\sum_{m=\frac{N+1}{2}}^{N-1}\sum_{i=\frac{N+1}{2}}^{m}\binom{N}{m-i}\binom{N}{i} + \sum_{i=\frac{N+1}{2}}^{N}\binom{N}{i}\binom{N}{N-i}, \end{aligned} \tag{A.64}$$

which is exactly equal to the prefactor in Eq. (A.63). Thus, we arrive at the inequality

$$[1 - f(p,N)]f(p,N) \geq 2^{2N-2}p^N(1-p)^N. \tag{A.65}$$

Substituting Eq. (A.65) into Eq. (A.60), we derive

$$\frac{\partial^2 \mathbb{S}(p,N)}{\partial p^2} \leq 0, \tag{A.66}$$

by using the inequality

$$\frac{2}{\pi} < \frac{4N}{2^{2N}}\binom{N-1}{(N-1)/2}^2 \leq 1. \tag{A.67}$$

Since $\partial\mathbb{S}(p,N)/\partial p$ is a monotonically decreasing function with respect to p, it follows from $\partial\mathbb{S}(p,N)/\partial p|_{p=1/2} = 0$ that $\partial\mathbb{S}(p,N)/\partial p$ is positive (negative) for $p < 1/2$ ($p > 1/2$). Therefore, $\mathbb{S}(p,N)$ is monotonically increasing (decreasing) for $p < 1/2$ ($p > 1/2$). According to Eq. (3.120), $P(p,N)$ is a monotonically increasing function with respect to p and $P(1/2,N) = 1/2$, thus we prove that $\mathbb{S}(P,N)$ is also monotonically increasing (decreasing) for $P < 1/2$ ($P > 1/2$).

A.5 Derivation of the Optimal Erasure Protocol

For Arrhenius rates, the master equation (A.68) can be recast as follows

$$\beta\Delta\epsilon = -\ln\left(\frac{\partial_t p + p}{1-p}\right). \tag{A.68}$$

The heat dissipated by a microscopic Arrhenius unit is given by the functional

$$\beta q = \beta\int_0^\tau \Delta\epsilon(\partial_t p)\mathrm{d}t = \beta\int_{p_i}^{p_f}\Delta\epsilon\,\mathrm{d}p = \int_0^\tau \mathcal{L}(p,\partial_t p)\mathrm{d}t, \tag{A.69}$$

with the explicitly time-independent Lagrangian

$$\mathcal{L}(p,\partial_t p) \equiv -(\partial_t p)\ln\left(\frac{\partial_t p + p}{1-p}\right). \tag{A.70}$$

The minimization of the heat functional amounts to solving the Euler-Lagrange equation

$$\mathcal{L} - (\partial_t p)\frac{\partial\mathcal{L}}{\partial(\partial_t p)} = K, \tag{A.71}$$

that admits the solutions

$$\partial_t p_1 = \frac{K - \sqrt{K^2+4Kp}}{2}, \quad \partial_t p_2 = \frac{K+\sqrt{K^2+4Kp}}{2}, \tag{A.72}$$

where K is constant resulting from the time-integration of the Euler-Lagrange equation.

Since we consider the erasure branch from the initial state $p_i = 1/2$ to the final one $p_f \leq p_i$, we restrict to the solution $\dot{p}_1$. This ordinary differential equation yields the following explicit expression of the erasure duration τ

$$\tau = \int_{p_i}^{p_f}\frac{1}{\partial_t p}\mathrm{d}p = h(p_f) - h(p_i), \tag{A.73}$$

where we defined the function $h(p)$ as

$$h(p) = -\sqrt{1+\frac{4p}{K}} - \ln\left(\sqrt{1+\frac{4p}{K}} - 1\right). \tag{A.74}$$

Substituting Eq. (A.72) into Eq. (A.69), we obtain the following expression for the dissipated heat

$$\beta q = \beta\int_{p_i}^{p_f}\Delta\epsilon\,\mathrm{d}p = H(p_f) - H(p_i), \tag{A.75}$$

where we defined the function $H(p)$ as

$$H(p) = \frac{1}{2}\sqrt{K^2+4Kp} - \ln(1-p) - p\ln\left(\frac{K+2p-\sqrt{K^2+4Kp}}{2(1-p)}\right). \tag{A.76}$$

A.6 Martin-Siggia Rose Formalism in a Nutshell

In the following, we derive the path-integral representation of the Markovian jump process studied in Sec. 4.2. This derivation is mainly adopted from Ref. [185]. The path integral will be constructed via the Martin-Siggia-Rose formalism which considers a stochastic process with Poissonian white noise. For this purpose, let us consider the stochastic occupation number $N_i(t)$ and write its change during a time interval Δt as follows

$$N_i(t+\Delta t) = N_i(t) + \sum_{\nu=1}^{L}\sum_{j=1}^{q}\left[\mathcal{N}_{ij}^{(\nu)}(\boldsymbol{N}(t),\Delta t) - \mathcal{N}_{ji}^{(\nu)}(\boldsymbol{N}(t),\Delta t)\right] \tag{A.77}$$

where $\mathcal{N}_{ij}^{(\nu)}(\boldsymbol{N}(t),\Delta t)$ refers to the number of jumps from j to i induced by the reservoir ν during the time interval Δt given that the system is in state $\boldsymbol{N}$ at time t. If the considered interval Δt is sufficiently small we can consider the probability per unit time for such a transition, $W_{ij}^{(\nu)}(\boldsymbol{\lambda}_t,\boldsymbol{N}(t))$, to be constant during the interval Δt. Then, the number of jumps $\mathcal{N}_{ij}^{(\nu)}(\boldsymbol{N}(t),\Delta t)$ are statistically independent random variables $\mathcal{X}_{ij}^{(\nu)}$ that take nonnegative integer values which Poisson distribution with parameter $W_{ij}^{(\nu)}(\boldsymbol{\lambda}_t,\boldsymbol{N}(t))\,\Delta t$, *i.e.*

$$\prod_{\nu=1}^{L}\prod_{j=1}^{q}\sum_{\mathcal{X}_{ij}^{(\nu)}}\frac{\left[W_{ij}^{(\nu)}(\boldsymbol{\lambda}_t,\boldsymbol{N}(t))\,\Delta t\right]^{\mathcal{X}_{ij}^{(\nu)}}}{\mathcal{X}_{ij}^{(\nu)}\,!}\,\mathrm{e}^{-W_{ij}^{(\nu)}(\boldsymbol{\lambda}_t,\boldsymbol{N}(t))\,\Delta t}\,. \tag{A.78}$$

The path probability $\mathcal{P}[\boldsymbol{m}_{(\tau)}]$ for a trajectory conditioned to be in state $\boldsymbol{N}$ at time $t'=0$ reads

$$\mathcal{P}(\boldsymbol{m}_{(\tau)}) = \left\langle \prod_{t'=0}^{t-1}\delta\left(\sum_i\left[N_i(t'+1)-N_i(t')-\sum_{\nu=1}^{L}\sum_{j=1}^{q}\left[\mathcal{N}_{ij}^{(\nu)}(\boldsymbol{N}(t'),\Delta t')-\mathcal{N}_{ji}^{(\nu)}(\boldsymbol{N}(t'),\Delta t')\right]\right]\right)\right\rangle_{\text{Poisson}} P_{\boldsymbol{N}}(0)\,, \tag{A.79}$$

Using the identity for the delta-distribution

$$\delta(\boldsymbol{N}(t)) = \prod_{i=0}^{q}\int_{-\infty}^{\infty}\mathrm{d}\pi_i(t)\,\mathrm{e}^{-\mathrm{i}\,\boldsymbol{\pi}_i(t)\,\boldsymbol{N}_i(t)}, \tag{A.80}$$

we can express the path probability (A.81) as

$$\begin{aligned}&\mathcal{P}(\boldsymbol{m}_{(\tau)},t) = \\ &= \prod_{i=1}^{q}\prod_{t'=0}^{t-1}\int_{-\infty}^{\infty}\mathrm{d}\pi_i(t'+1)\mathrm{e}^{-\mathrm{i}\pi_i(t'+1)[N_i(t'+1)-N_i(t')]}\,P_{\boldsymbol{N}_0}(0)\cdot\end{aligned}$$

$$\cdot \prod_{\nu=1}^{L}\prod_{j=1}^{q}\sum_{\mathcal{X}_{ij}^{(\nu)}} \frac{\left[W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'},\boldsymbol{N}(t))\,\Delta t'\right]^{\mathcal{X}_{ij}^{(\nu)}}}{\mathcal{X}_{ij}^{(\nu)}\,!}\, \mathrm{e}^{-W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'},\boldsymbol{N}(t'))\,\Delta t'}\, \mathrm{e}^{\mathrm{i}\pi_i(t'+1)\sum_{\nu=1}^{L}\sum_{j=1}^{q}\mathcal{X}_{ij}^{(\nu)}}$$

$$= \prod_{i=1}^{q}\prod_{t'=0}^{t-1}\int_{-\infty}^{\infty} \mathrm{d}\pi_i(t'+1)\mathrm{e}^{-\mathrm{i}\pi_i(t'+1)[N_i(t'+1)-N_i(t')]}\, P_{\boldsymbol{N}_0}(0)\cdot$$

$$\cdot \prod_{\nu=1}^{L}\prod_{j=1}^{q} \mathrm{e}^{W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'},\boldsymbol{N}(t'))\,\Delta t'}\, \mathrm{e}^{W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'},\boldsymbol{N}(t'))\,\Delta t'\,\mathrm{e}^{\mathrm{i}\pi_i(t'+1)}}$$

$$= \prod_{i=1}^{q}\left(\prod_{t'=0}^{t-1}\int_{-\infty}^{\infty}\mathrm{d}\pi_i(t'+1)\right) \mathrm{e}^{\sum_{t'=0}^{t-1}-\mathrm{i}\pi_i(t'+1)[N_i(t'+1)-N_i(t')]+\Delta t'\sum_{\nu=1}^{L}\sum_{j=1}^{q}W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'},\boldsymbol{N}(t'))\left[\mathrm{e}^{\mathrm{i}\pi_i(t'+1)}-1\right]}\, P_{\boldsymbol{N}_0}(0)$$

$$\equiv \int \mathcal{D}\,[\boldsymbol{\pi}]\, \mathrm{e}^{\int_0^t \mathrm{d}t'\left(\boldsymbol{\pi}\dot{\boldsymbol{N}}+H[\boldsymbol{N}(t'),\boldsymbol{\pi}(t')]\right)}\, P_{\boldsymbol{N}_0}(0) = \int \mathcal{D}\,[\boldsymbol{\pi}]\, \mathrm{e}^{\mathcal{L}[\boldsymbol{N}(t'),\boldsymbol{\pi}(t')]}, \tag{A.81}$$

where we defined the path integral measure $\int \mathcal{D} \equiv \prod_{i=1}^{q}\prod_{t'=0}^{t-1}\int_{-\infty}^{\infty}\mathrm{d}\pi_i(t'+1)[\boldsymbol{\pi}]$ under the transformations $\mathrm{i}\pi \to \pi$ and $\sum_{t'}\Delta t' \to \int \mathrm{d}t'$. Moreover, $H[\boldsymbol{N}(t'),\boldsymbol{\pi}(t')]$ and $\mathcal{L}[\boldsymbol{N}(t'),\boldsymbol{\pi}(t')]$ denote the unbiased Hamiltonian and action functional, see Eq. (4.80) and hereinafter.

Let us consider a generic observable O which changes along the trajectory $\boldsymbol{m}_{(\tau)}$ until time t by the amount $\delta O[\boldsymbol{m}_{(\tau)},t]$ that might be split into a time-integrated current-like contribution

$$\sum_{\nu=1}^{L}\sum_{i,j=1}^{q}\int_0^t \mathrm{d}t'\, O_{ij}^{(\nu)}(\boldsymbol{N}(t')), \tag{A.82}$$

in a contribution due to the nonautonomous driving

$$\int_0^t \mathrm{d}t'\,\dot{\boldsymbol{\lambda}}_{t'}\cdot\left[\nabla_{\boldsymbol{\lambda}_{t'}}O_{\boldsymbol{N}(t')}\right], \tag{A.83}$$

and into a state-like contribution

$$\int_0^t \mathrm{d}t'\;\mathrm{d}_{t'}O_{\boldsymbol{N}}(t') = -\gamma_O[O_{\boldsymbol{N}_M}(t)-O_{\boldsymbol{N}_0}(0)]. \tag{A.84}$$

The generating function associated with that change is given by

$$G(\gamma_O,t,\boldsymbol{m}_{(\tau)}) = \mathrm{e}^{\gamma_O\,\delta O[\boldsymbol{m}_{(\tau)},t]}\,\mathcal{P}(\boldsymbol{m}_{(\tau)},t), \tag{A.85}$$

so that, substituting the last equation into (A.81), we obtain

$$G(\gamma_O,t,\boldsymbol{m}_{(\tau)}) =$$

$$
\begin{aligned}
&= \prod_{i=1}^{q} \prod_{t'=0}^{t-1} \int_{-\infty}^{\infty} \mathrm{d}\pi_i(t'+1)\, P_{\boldsymbol{N}_0}(0) \cdot \\
&\cdot \mathrm{e}^{-\mathrm{i}\pi_i(t'+1)[N_i(t'+1)-N_i(t')] - \gamma_O[O_{\boldsymbol{N}(t'+1)} - O_{\boldsymbol{N}(t')}] - \gamma_O[\boldsymbol{\lambda}_{t'+1}-\boldsymbol{\lambda}_{t'}]\left[\nabla_{\boldsymbol{\lambda}_{t'+1}} O_{\boldsymbol{N}(t'+1)} - \nabla_{\boldsymbol{\lambda}_{t'}} O_{\boldsymbol{N}(t')}\right]} \cdot \\
&\cdot \prod_{\nu=1}^{L} \prod_{j=1}^{q} \sum_{\mathcal{X}_{ij}^{(\nu)}} \frac{\left[W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'}, \boldsymbol{N}(t)) \cdot \Delta t'\right]^{\mathcal{X}_{ij}^{(\nu)}}}{\mathcal{X}_{ij}^{(\nu)}\,!}\, \mathrm{e}^{-W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'}, \boldsymbol{N}(t'))\cdot\Delta t'}\, \mathrm{e}^{\mathrm{i}\pi_i(t'+1)\sum_{\nu=1}^{L}\sum_{j=1}^{q} \mathcal{X}_{ij}^{(\nu)} - \gamma_O \mathcal{X}_{ij}^{(\nu)} O_{ij}^{(\nu)}(\boldsymbol{N}(t'))} \\
&= \prod_{i=1}^{q} \prod_{t'=0}^{t-1} \int_{-\infty}^{\infty} \mathrm{d}\pi_i(t'+1)\, P_{\boldsymbol{N}_0}(0) \cdot \\
&\cdot \mathrm{e}^{-\mathrm{i}\pi_i(t'+1)[N_i(t'+1)-N_i(t')] - \gamma_O[O_{\boldsymbol{N}(t'+1)} - O_{\boldsymbol{N}(t')}] - \gamma_O[\boldsymbol{\lambda}_{t'+1}-\boldsymbol{\lambda}_{t'}]\left[\nabla_{\boldsymbol{\lambda}_{t'+1}} O_{\boldsymbol{N}(t'+1)} - \nabla_{\boldsymbol{\lambda}_{t'}} O_{\boldsymbol{N}(t')}\right]} \cdot \\
&\cdot \prod_{\nu=1}^{L} \prod_{j=1}^{q} \mathrm{e}^{W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'}, \boldsymbol{N}(t'))\cdot\Delta t'}\, \mathrm{e}^{W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'}, \boldsymbol{N}(t'))\cdot\Delta t' \cdot \mathrm{e}^{\mathrm{i}\pi_i(t'+1) - \gamma_O O_{ij}^{(\nu)}(\boldsymbol{N}(t'))}} \\
&= \prod_{i=1}^{q} \left(\prod_{t'=0}^{t-1} \int_{-\infty}^{\infty} \mathrm{d}\pi_i(t'+1) \right) \cdot \\
&\cdot \mathrm{e}^{\sum_{t'=0}^{t-1} -\mathrm{i}\pi_i(t'+1)[N_i(t'+1)-N_i(t')] - \gamma_O[O_{\boldsymbol{N}(t'+1)} - O_{\boldsymbol{N}(t')}] - \gamma_O[\boldsymbol{\lambda}_{t'+1}-\boldsymbol{\lambda}_{t'}]\left[\nabla_{\boldsymbol{\lambda}_{t'+1}} O_{\boldsymbol{N}(t'+1)} - \nabla_{\boldsymbol{\lambda}_{t'}} O_{\boldsymbol{N}(t')}\right]} \cdot \\
&\cdot \mathrm{e}^{\Delta t' \sum_{\nu=1}^{L}\sum_{j=1}^{q} W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'}, \boldsymbol{N}(t'))\left[\mathrm{e}^{\mathrm{i}\pi_i(t'+1) - \gamma_O O_{ij}^{(\nu)}(\boldsymbol{N}(t'))} - 1\right]}\, P_{\boldsymbol{N}_0}(0) \\
&\equiv \int \mathcal{D}\,[\boldsymbol{\pi}]\, \mathrm{e}^{\int_0^t \mathrm{d}t' \left[\boldsymbol{\pi}\dot{\boldsymbol{N}} + H_{\gamma_O}[\boldsymbol{N}(t'), \boldsymbol{\pi}(t')] - \gamma_O \dot{\boldsymbol{\lambda}}_{t'} \cdot \left[\nabla_{\boldsymbol{\lambda}_{t'}} O_{\boldsymbol{N}(t')}\right] - \gamma_O \mathrm{d}_{t'} O_{\boldsymbol{N}(t')}\right]}\, P_{\boldsymbol{N}_0}(0) \\
&= \int \mathcal{D}\,[\boldsymbol{\pi}]\, \mathrm{e}^{\mathcal{L}_{\gamma_O}[\boldsymbol{N}(t'), \boldsymbol{\pi}(t')]}. \qquad \text{(A.86)}
\end{aligned}
$$

Integrating over all possible paths, $\int \mathcal{D}[\boldsymbol{N}]$, we finally arrive at the following path-integral representation of the generating function

$$
G(\gamma_O, t) = \int \mathcal{D}[\boldsymbol{N}]\, G(\gamma_O, t, \boldsymbol{m}_{(\tau)}) = \int \mathcal{D}[\boldsymbol{N}] \int \mathcal{D}[\boldsymbol{\pi}]\, \mathrm{e}^{\mathcal{L}_{\gamma_O}[\boldsymbol{N}(t'), \boldsymbol{\pi}(t')]}, \qquad \text{(A.87)}
$$

as stated in Eq. (4.80).

A.7 Derivation of the Fluctuation Theorem (4.87)

We rescale the extensive state variables $\boldsymbol{n} = \boldsymbol{N}/N$ so that $NO_{\boldsymbol{n}} \equiv O(\boldsymbol{N})$. Also, the discrete gradient of an observable along an edge asymptotically becomes a derivative with growing size, $N(\partial_{n_i} - \partial_{n_j})\, O_{\boldsymbol{n}} = O_{ij}(\boldsymbol{N})$. The path-integral representation of the generating function (4.84) then reads

$$G(\boldsymbol{\gamma}, t) = \int \mathcal{D}[\boldsymbol{N}] \int \mathcal{D}[\boldsymbol{\pi}]\, P^{eq}_{\boldsymbol{N}_0}(\boldsymbol{\lambda}_0) \cdot \exp\left[\int_0^t \mathrm{d}t' \left(-N\gamma_{\Lambda}\beta^{(1)} \dot{\boldsymbol{\lambda}}_{t'} \cdot \left[\nabla_{\boldsymbol{\lambda}_{t'}} E_{\boldsymbol{n}}(\boldsymbol{\lambda}_{t'}) \right] + \sum_{i=1}^{q} \Big\{ -\pi_i(t') \dot{N}_i(t') + \right.\right.$$
$$\left.\left. + \sum_{\nu=1}^{L} \sum_{j=1}^{q} \left[\mathrm{e}^{\pi_i(t') - \pi_j(t') - \gamma_F^{(\nu)} \beta^{(\nu)} f_{ij}^{(\nu)} - N\gamma_E^{(\nu)} [\beta^{(1)} - \beta^{(\nu)}] (\partial_{n_i} - \partial_{n_j}) E^{(\nu)}_{\boldsymbol{n}_{t'}}(\boldsymbol{\lambda}_{t'})} - 1 \right] W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'}, \boldsymbol{N}(t')) \Big\} \right) \right]. \tag{A.88}$$

The crucial step of the derivation is to define physically consistent transformation rules to time-reverse the biased stochastic dynamics. Time-reversal transformations of unbiased Langevin dynamics have been investigated in Ref. [186]. For the generating function in question (A.88), we define the time-reversed biased stochastic dynamics as follows

$$\begin{aligned} &\tilde{t}' = t - t', \quad \tilde{\boldsymbol{n}} = \boldsymbol{n}, \quad \tilde{\boldsymbol{\lambda}}_{t'} = \boldsymbol{\lambda}_{t-t'}, \\ &\tilde{\boldsymbol{\pi}} = -\boldsymbol{\pi} + N\beta^{(1)} \nabla_{\boldsymbol{n}} A^{(1)}(\boldsymbol{n}) = -\boldsymbol{\pi} + N\beta^{(1)} \nabla_{\boldsymbol{n}} E(\boldsymbol{n}) - N\nabla_{\boldsymbol{n}} S^{int}(\boldsymbol{n}), \quad \tilde{\boldsymbol{\gamma}} = 1 - \boldsymbol{\gamma}, \end{aligned} \tag{A.89}$$

while reusing the shorthand notation from Eq. (4.70). The definitions of the time-reversed physical quantities in the first line are trivial. Less obvious is the transformation rule of the auxiliary field $\boldsymbol{\pi}$. This transformation rule amounts to inverting the directions of the edges corresponding to a reversion of the Markov dynamics: The change of the sign in front of $\boldsymbol{\pi}$ can be seen by noting that the latter is a counting field for variations in the state variables $\mathrm{d}\boldsymbol{N}$. Moreover, the affinity along an edge is inverted by the free energy shift.

We proceed by demonstrating that the above transformation, up to a non-fluctuating quantity, indeed leaves the generating function invariant. For better readability, we will split the action functional (A.88) into two parts and investigate how they transform under the time reversal in Eq. (A.89). First, the invariance of the biased Hamiltonian under this time-reversal transformation can be seen as follows,

$$\begin{aligned} &\tilde{H}_{\tilde{\boldsymbol{\gamma}}}[\boldsymbol{N}(t'), \boldsymbol{\pi}(t')] = \\ &= \sum_{\nu=1}^{L} \sum_{i,j=1}^{q} \Big[\exp\Big\{ \pi_j(t') - \pi_i(t') + N\beta^{(1)}(\partial_{n_i} - \partial_{n_j}) A^{(1)}(\boldsymbol{n}(t')) + \big(\gamma_F^{(\nu)} - 1\big)\beta^{(\nu)} f_{ij}^{(\nu)} + \\ &+ \big(\gamma_E^{(\nu)} - 1\big) N[\beta^{(1)} - \beta^{(\nu)}](\partial_{n_i} - \partial_{n_j}) E^{(\nu)}(\boldsymbol{n}(t')) \Big\} - 1 \Big] W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'}, \boldsymbol{N}(t')) \\ &= \sum_{\nu=1}^{L} \sum_{i,j=1}^{q} \Big[\exp\Big\{ \pi_j(t') - \pi_i(t') - N\beta^{(1)}(\partial_{n_j} - \partial_{n_i}) A^{(1)}(\boldsymbol{n}(t')) + \big(1 - \gamma_F^{(\nu)}\big)\beta^{(\nu)} f_{ji}^{(\nu)} + \end{aligned}$$

$$+\left(1-\gamma_E^{(\nu)}\right)N[\beta^{(1)}-\beta^{(\nu)}](\partial_{n_j}-\partial_{n_i})E^{(\nu)}(\boldsymbol{n}(t'))\Big\}-1\Big]W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'},\boldsymbol{N}(t')) \tag{A.90}$$

$$=\sum_{\nu=1}^{L}\sum_{i,j=1}^{q}\Big[\exp\Big\{\pi_j(t')-\pi_i(t')-\beta^{(\nu)}\big[N(\partial_{n_j}-\partial_{n_i})A^{(\nu)}(\boldsymbol{n}(t'))-f_{ji}^{(\nu)}\big]-\gamma_F^{(\nu)}\beta^{(\nu)}f_{ji}^{(\nu)}- \tag{A.91}$$

$$-\gamma_E^{(\nu)}N[\beta^{(1)}-\beta^{(\nu)}](\partial_{n_j}-\partial_{n_i})E^{(\nu)}(\boldsymbol{n}(t'))\Big\}-1\Big]W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'},\boldsymbol{N}(t'))$$

$$=\sum_{\nu=1}^{L}\sum_{i,j=1}^{q}\Big[\exp\big[\pi_j(t')-\pi_i(t')\big]\,\frac{W_{ji}^{(\nu)}(\boldsymbol{\lambda}_{t'},\boldsymbol{N}(t'))}{W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'},\boldsymbol{N}(t'))}\cdot$$

$$\cdot\exp\Big\{-\gamma_F^{(\nu)}\beta^{(\nu)}f_{ji}^{(\nu)}-\gamma_E^{(\nu)}N[\beta^{(1)}-\beta^{(\nu)}](\partial_{n_j}-\partial_{n_i})E^{(\nu)}(\boldsymbol{n}(t'))\Big\}-1\Big]W_{ij}^{(\nu)}(\boldsymbol{\lambda}_{t'},\boldsymbol{N}(t'))$$

$$=\sum_{\nu=1}^{L}\sum_{i,j=1}^{q}\Big[\mathrm{e}^{\pi_j(t')-\pi_i(t')-\gamma_F^{(\nu)}\beta^{(\nu)}f_{ji}^{(\nu)}-\gamma_E^{(\nu)}N[\beta^{(1)}-\beta^{(\nu)}](\partial_{n_j}-\partial_{n_i})E^{(\nu)}(\boldsymbol{n}(t'))}-1\Big]W_{ji}^{(\nu)}(\boldsymbol{\lambda}_{t'},\boldsymbol{N}(t'))$$

$$=H_{\boldsymbol{\gamma}}[\boldsymbol{N}(t'),\boldsymbol{\pi}(t')]. \tag{A.92}$$

Furthermore, we find for the sum of the kinetic and non-autonomous driving terms together with the initial condition under time-reversal,

$$\int_0^t \mathrm{d}t'\,\Big\{\big[N\,\beta^{(1)}\,\nabla_{\boldsymbol{n}}\,A^{(1)}(\boldsymbol{n}(t'))-\boldsymbol{\pi}(t')\big]\cdot\dot{\boldsymbol{N}}(t')+N(1-\gamma_{\Lambda})\,\beta^{(1)}\dot{\boldsymbol{\lambda}}_{t'}\cdot\big[\nabla_{\boldsymbol{\lambda}_{t'}}E_{\boldsymbol{n}}(\boldsymbol{\lambda}_{t'})\big]\Big\}+\ln P_{\boldsymbol{N}_t}^{eq}(\boldsymbol{\lambda}_t)\;=$$

$$=\int_0^t \mathrm{d}t'\Big\{N\beta^{(1)}\big(\mathrm{d}_{t'}A^{(1)}(\boldsymbol{n}(t'))-\dot{\boldsymbol{\lambda}}_{t'}\cdot\big[\nabla_{\boldsymbol{\lambda}_{t'}}A^{(1)}(\boldsymbol{n}(t'))\big]-\boldsymbol{\pi}(t')\cdot\dot{\boldsymbol{N}}(t')+N(1-\gamma_{\Lambda})\,\beta^{(1)}\dot{\boldsymbol{\lambda}}_{t'}\cdot\big[\nabla_{\boldsymbol{\lambda}_{t'}}E_{\boldsymbol{n}}(\boldsymbol{\lambda}_{t'})\big]\Big\}+\ln P_{\boldsymbol{N}_t}^{eq}(\boldsymbol{\lambda}_t)$$

$$=-\int_0^t \mathrm{d}t'\{\boldsymbol{\pi}(t')\cdot\dot{\boldsymbol{N}}(t')+N\gamma_{\Lambda}\beta^{(1)}\dot{\boldsymbol{\lambda}}_{t'}\cdot\big[\nabla_{\boldsymbol{\lambda}_{t'}}E_{\boldsymbol{n}}(\boldsymbol{\lambda}_{t'})\big]\Big\}+N\beta^{(1)}\big[A^{(1)}(\boldsymbol{n}(t))-A^{(1)}(\boldsymbol{n}(0))\big]+\ln P_{\boldsymbol{N}_t}^{eq}(\boldsymbol{\lambda}_t)$$

$$=-\int_0^t \mathrm{d}t'\{\boldsymbol{\pi}(t')\cdot\dot{\boldsymbol{N}}(t')+N\gamma_{\Lambda}\,\beta^{(1)}\dot{\boldsymbol{\lambda}}_{t'}\cdot\big[\nabla_{\boldsymbol{\lambda}_{t'}}E_{\boldsymbol{n}}(\boldsymbol{\lambda}_{t'})\big]\Big\}+\ln\frac{P_{\boldsymbol{N}_0}^{eq}(\boldsymbol{\lambda}_0)}{P_{\boldsymbol{N}_t}^{eq}(\boldsymbol{\lambda}_t)}+\beta^{(1)}\big[A^{eq}(\boldsymbol{\lambda}_t)-A^{eq}(\boldsymbol{\lambda}_0)\big]+\ln P_{\boldsymbol{N}_t}^{eq}(\boldsymbol{\lambda}_t)$$

$$=-\int_0^t \mathrm{d}t'\,\{\boldsymbol{\pi}(t')\cdot\dot{\boldsymbol{N}}(t')+N\gamma_{\Lambda}\,\beta^{(1)}\dot{\boldsymbol{\lambda}}_{t'}\cdot\big[\nabla_{\boldsymbol{\lambda}_{t'}}E_{\boldsymbol{n}}(\boldsymbol{\lambda}_{t'})\big]\Big\}+\ln P_{\boldsymbol{N}_0}^{eq}(\boldsymbol{\lambda}_0)+\beta^{(1)}\big[A^{eq}(\boldsymbol{\lambda}_t)-A^{eq}(\boldsymbol{\lambda}_0)\big]. \tag{A.93}$$

Collecting results, we thus find that the size-intensive action functional is invariant under the time reversal (A.89) up to a non-fluctuating term corresponding to the change in the size-intensive part of the equilibrium free energy, *i.e.*

$$\mathcal{L}_{\boldsymbol{\gamma}}[\boldsymbol{n},\boldsymbol{\pi}]=\tilde{\mathcal{L}}_{\tilde{\boldsymbol{\gamma}}}[\boldsymbol{n},\boldsymbol{\pi}]-\beta^{(1)}\Delta\mathcal{A}_1^{eq}(\boldsymbol{\lambda}). \tag{A.94}$$

In the macroscopic limit, the scaled cumulant generating function is equal to the extremal action functional, cf. Eq. (4.86). Moreover, the action functional contains the initial condition

of the trajectories so that its extremization does not give rise to additional boundary terms,

$$\delta \mathcal{L}_{\boldsymbol{\gamma}}[\boldsymbol{n}, \boldsymbol{\pi}] = \delta \tilde{\mathcal{L}}_{\tilde{\boldsymbol{\gamma}}}[\boldsymbol{n}, \boldsymbol{\pi}]. \tag{A.95}$$

Hence the invariance of the action functional is preserved in the macroscopic limit that in turn stipulates the following symmetry for the scaled-cumulant generating function,

$$\mathcal{G}(\boldsymbol{\gamma}, t) = \tilde{\mathcal{G}}(\tilde{\boldsymbol{\gamma}}, t) - \beta^{(1)} \Delta \mathcal{A}_1^{eq}(\boldsymbol{\lambda}), \tag{A.96}$$

which is exactly Eq. (4.87).

A.8 Generalization to Many-Body Systems

Setup

The procedure detailed in Sec. 4.3 for two interacting and underdamped particles can be straightforwardly generalized to N underdamped Brownian particles. We consider two subsystems $\boldsymbol{E}$ and $\boldsymbol{M}$ that consist of E and M particles, respectively, such that $E + M = N$. The particles in the subsystems $\boldsymbol{E}$ and $\boldsymbol{M}$ are interacting with each other via the potential $u_{\boldsymbol{E}}^{int}(\boldsymbol{x}_{\boldsymbol{E}}, t)$ and $u_{\boldsymbol{M}}^{int}(\boldsymbol{x}_{\boldsymbol{M}}, t)$, respectively, while the subsystems are coupled to each other according to the cumulated interaction potential $u_{\boldsymbol{EM}}^{int}(\{\boldsymbol{x}_i\})$. We label the particles by the index i, with the mass m_i by $i = 1, 2, \ldots, N$ and the phase-space coordinate $\boldsymbol{\Gamma}_i = (\boldsymbol{x}_i, \boldsymbol{v}_i)^\top$. Again, each particle is subjected to velocity-independent nonconservative forces $\boldsymbol{f}_i(\boldsymbol{x}_i, t)$ and coupled to a heat reservoir at inverse temperature β_i, as depicted in Fig. A.1.

The total potential of the many-body system reads

$$u(\{\boldsymbol{x}_i\}, t) = \sum_{i=1}^{N} u_i(\boldsymbol{x}_i, t) + u_{\boldsymbol{E}}^{int}(\boldsymbol{x}_{\boldsymbol{E}}, t) + u_{\boldsymbol{M}}^{int}(\boldsymbol{x}_{\boldsymbol{M}}, t) + \underbrace{\sum_{j=1}^{E} \sum_{k=N-M}^{N} u_{jk}^{int}(\boldsymbol{x}_j, \boldsymbol{x}_k, t)}_{\equiv u_{\boldsymbol{EM}}^{int}(\{\boldsymbol{x}_i\}, t)}. \tag{A.97}$$

We stress that the interaction potentials $u_{\boldsymbol{E}}^{int}(\boldsymbol{x}_{\boldsymbol{E}}, t)$ and $u_{\boldsymbol{M}}^{int}(\boldsymbol{x}_{\boldsymbol{M}}, t)$ need to be distinguished from the sum of the single-particle bare potentials in the respective subsystem which we denote by $u_{\boldsymbol{E}}(\boldsymbol{x}_{\boldsymbol{E}}, t) = \sum_{i=1}^{E} u_i(\boldsymbol{x}_i, t)$ and $u_{\boldsymbol{M}}(\boldsymbol{x}_{\boldsymbol{M}}, t) = \sum_{i=N-M}^{N} u_i(\boldsymbol{x}_i, t)$, respectively. The Hamiltonian is given by

$$e = \underbrace{\sum_{i=1}^{E} e_i(\boldsymbol{\Gamma}_i, t) + u_{\boldsymbol{E}}^{int}(\boldsymbol{x}_{\boldsymbol{E}}, t)}_{e_{\boldsymbol{E}}} + \underbrace{\sum_{i=N-M}^{N} e_i(\boldsymbol{\Gamma}_i, t) + u_{\boldsymbol{M}}^{int}(\boldsymbol{x}_{\boldsymbol{M}}, t)}_{e_{\boldsymbol{M}}} + u_{\boldsymbol{EM}}^{int}(\{\boldsymbol{x}_i\}, t), \tag{A.98}$$

with the bare single-particle Hamiltonians $e_i(\boldsymbol{\Gamma}_i, t) = m_i \boldsymbol{v}_i^2/2 + u_i(\boldsymbol{x}_i, t)$.

Stochastic Dynamics And Thermodynamics

The resulting stochastic dynamics of the many-body system is ruled by the following Fokker-Planck equation

$$\partial_t \rho = -\nabla \cdot \boldsymbol{J} = -\nabla \cdot \left(\boldsymbol{L}^{det} + \boldsymbol{L}^{diss} \right) \rho, \tag{A.99}$$

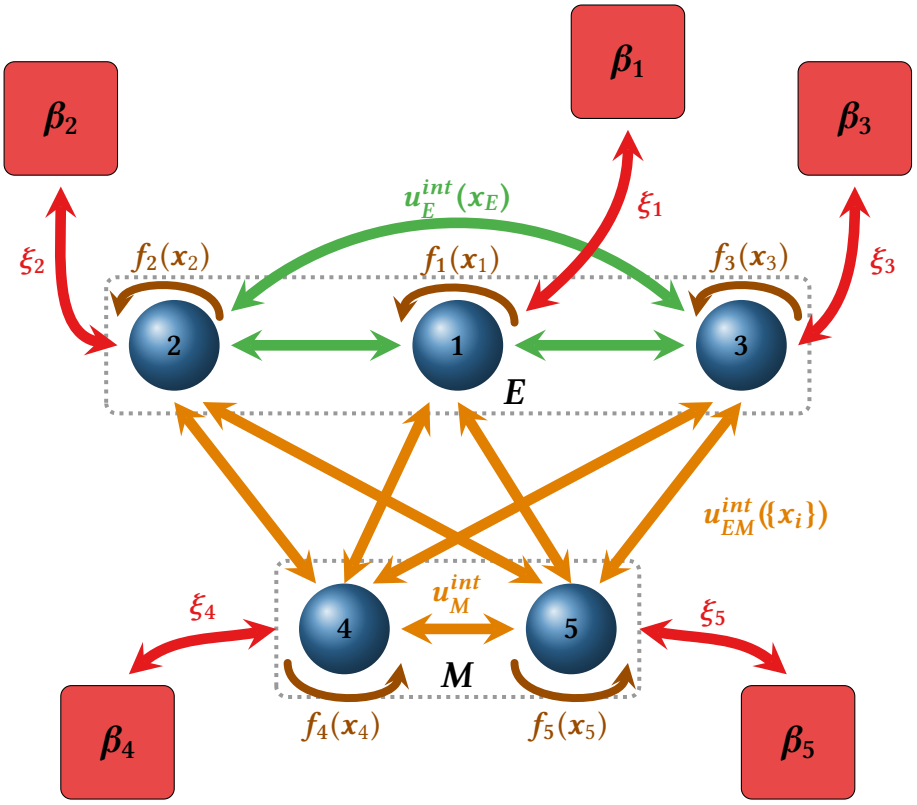

Figure A.1: *Schematic depiction of a system of five interacting underdamped Brownian particles. The $E = 3$ particles in subsystem $\boldsymbol{E}$ are interacting via $u_{\boldsymbol{E}}^{int}(x_E)$, while the $M = 2$ particles belonging to the subsystem $\boldsymbol{M}$ are interacting via $u_{\boldsymbol{M}}^{int}(x_M)$. The two subsystems are coupled by the interaction potential $u_{\boldsymbol{EM}}^{int}(\{\boldsymbol{x}_i\})$. Moreover, each particle is coupled to a heat reservoir at inverse temperature β_i and is subjected to a nonconservative force $\boldsymbol{f}_i(\boldsymbol{x}_i)$.*

with $\nabla = (\partial_{\boldsymbol{x}_1}, \partial_{\boldsymbol{v}_1}, \ldots, \partial_{\boldsymbol{x}_N}, \partial_{\boldsymbol{v}_N})^\top$ and

$$\boldsymbol{L}^{det} = \begin{pmatrix} \boldsymbol{v}_1 \\ \frac{1}{m_1}\left[-\partial_{\boldsymbol{x}_1} u(\{\boldsymbol{x}_i\}, t) + \boldsymbol{f}_1(\boldsymbol{x}_1, t)\right] \\ \vdots \\ \boldsymbol{v}_N \\ \frac{1}{m_N}\left[-\partial_{\boldsymbol{x}_N} u(\{\boldsymbol{x}_i\}, t) + \boldsymbol{f}_N(\boldsymbol{x}_N, t)\right] \end{pmatrix} \tag{A.100}$$

$$\boldsymbol{L}^{diss} = \begin{pmatrix} 0 \\ \frac{-\xi_1}{m_1^2}(m_1\boldsymbol{v}_1 + \beta_1^{-1}\partial_{\boldsymbol{v}_1}\ln\rho) \\ \vdots \\ 0 \\ \frac{-\xi_N}{m_N^2}(m_N\boldsymbol{v}_N + \beta_N^{-1}\partial_{\boldsymbol{v}_N}\ln\rho) \end{pmatrix}. \tag{A.101}$$

The average energy of the system is defined as

$$E = \int \mathrm{d}\boldsymbol{\Gamma}\; e\,\rho, \tag{A.102}$$

and gives rise to the first law of thermodynamics

$$d_t E = \dot{Q} + \dot{W}, \tag{A.103}$$

with the heat current

$$\dot{Q} = \int \mathrm{d}\boldsymbol{\Gamma}\; e\,\partial_t\rho - \sum_{i=1}^{N}\int \mathrm{d}\boldsymbol{\Gamma}\;\rho\;\boldsymbol{v}_i\cdot\boldsymbol{f}_i. \tag{A.104}$$

and the work current

$$\dot{W} = \int \mathrm{d}\boldsymbol{\Gamma}\;\rho\,\partial_t e + \sum_{i=1}^{N}\int \mathrm{d}\boldsymbol{\Gamma}\;\rho\;\boldsymbol{v}_i\cdot\boldsymbol{f}_i, \tag{A.105}$$

Like in the two-particle case (4.175), the heat current splits into additive contributions that each are functionals of the single-particle probabilities ρ_i only,

$$\dot{Q} = \sum_{i=1}^{N}\dot{q}^{(i)},\;\; \dot{q}^{(i)} = -\xi_i\int \mathrm{d}\boldsymbol{\Gamma}\rho\left(\boldsymbol{v}_i + \frac{1}{\beta_i m_i}\partial_{\boldsymbol{v}_i}\ln\rho\right)\boldsymbol{v}_i. \tag{A.106}$$

We define the N-particle Shannon entropy as the nonequilibrium system entropy

$$S = -\int \mathrm{d}\boldsymbol{\Gamma}\;\rho\;\ln\rho, \tag{A.107}$$

which gives rise to the entropy balance

$$\mathrm{d}_t S = \sum_{i=1}^{N}\beta_i\,\dot{q}^{(i)} + \dot{\Sigma}, \tag{A.108}$$

with the non-negative entropy production rate,

$$\dot{\Sigma} = \sum_{i=1}^{N}\dot{\sigma}^{(i)},\;\; \dot{\sigma}^{(i)} = \beta_i\,\xi_i\int \mathrm{d}\boldsymbol{\Gamma}\;\rho\left(\boldsymbol{v}_i + \frac{1}{\beta_i m_i}\partial_{\boldsymbol{v}_i}\ln\rho\right)^2 \geq 0. \tag{A.109}$$

Coarse-Graining

Marginalization

We proceed by formulating an effective description for a reduced dynamics in the subsystem $\boldsymbol{E}$. To Integration of the N-particle distribution ρ over $\boldsymbol{\Gamma_M}$ yields the marginal probability, $\rho_{\boldsymbol{E}} \equiv \int \mathrm{d}\boldsymbol{\Gamma_M}\, \rho$, that satisfies the following effective Fokker-Planck equation

$$\partial_t \rho_{\boldsymbol{E}} = -\nabla_{\boldsymbol{E}} \cdot \left(\boldsymbol{L}_{\boldsymbol{E}}^{det} + \boldsymbol{L}_{\boldsymbol{E}}^{diss} \right) \rho_{\boldsymbol{E}}, \tag{A.110}$$

with $\nabla_{\boldsymbol{E}} = (\partial_{\boldsymbol{x}_1}, \partial_{\boldsymbol{v}_1}, \ldots, \partial_{\boldsymbol{x}_E}, \partial_{\boldsymbol{v}_E})^\top$ and

$$\boldsymbol{L}_{\boldsymbol{E}}^{det} = \begin{pmatrix} \boldsymbol{v}_1 \\ \frac{1}{m_1}\left[-\partial_{\boldsymbol{x}_1}[u_1 + u_{\boldsymbol{E}}^{int}](\boldsymbol{x}_{\boldsymbol{E}}, t) + \boldsymbol{f}_1(\boldsymbol{x}_1, t) + \boldsymbol{f}^{(1)}(\boldsymbol{\Gamma}_{\boldsymbol{E}}, t)\right] \\ \vdots \\ \boldsymbol{v}_E \\ \frac{1}{m_E}\left[-\partial_{\boldsymbol{x}_E}[u_E + u_{\boldsymbol{E}}^{int}](\boldsymbol{x}_{\boldsymbol{E}}, t) + \boldsymbol{f}_E(\boldsymbol{x}_E, t) + \boldsymbol{f}^{(E)}(\boldsymbol{\Gamma}_{\boldsymbol{E}}, t)\right] \end{pmatrix} \tag{A.111}$$

$$\boldsymbol{L}_{\boldsymbol{E}}^{diss} = \begin{pmatrix} 0 \\ \frac{-\xi_1}{m_1^2}(m_1 \boldsymbol{v}_1 + \beta_1^{-1} \partial_{\boldsymbol{v}_1} \ln \rho) \\ \vdots \\ 0 \\ \frac{-\xi_E}{m_E^2}(m_E\, \boldsymbol{v}_E + \beta_E^{-1} \partial_{\boldsymbol{v}_E} \ln \rho) \end{pmatrix}. \tag{A.112}$$

The interaction between the effective $\boldsymbol{E}$ and the coarse-grained subsystem $\boldsymbol{M}$ are encoded in the nonconservative forces

$$\boldsymbol{f}^{(i)}(\boldsymbol{\Gamma}_{\boldsymbol{E}}, t) = -\int \mathrm{d}\boldsymbol{\Gamma_M}\, \rho_{\boldsymbol{M}|\boldsymbol{E}}\, \partial_{\boldsymbol{x}_i} u_{\boldsymbol{EM}}^{int}, \tag{A.113}$$

for $i \in \boldsymbol{E}$.

Following the reasoning in the context of Eq. (4.166), we define an effective entropy balance for the reduced dynamics of the subsystem $\boldsymbol{E}$ as

$$\mathrm{d}_t S = \dot{Q}^{(\boldsymbol{E})} + \dot{\Sigma}^{(\boldsymbol{E})} + \sum_{i=1}^{E} \frac{1}{m_i} \int \mathrm{d}\boldsymbol{\Gamma}_i\, \rho_i\, \partial_{\boldsymbol{v}_i} \cdot \boldsymbol{f}^{(i)}\,, \tag{A.114}$$

where the probability associated with the ith particle is denoted as ρ_i. Here, the effective heat

$$\dot{Q}^{(\boldsymbol{E})} = \sum_{i=1}^{E} \dot{q}^{(i)} + \beta_E^{-1}\, \mathrm{d}_t S_{\boldsymbol{M}|\boldsymbol{E}}, \tag{A.115}$$

is supplemented by the conditional Shannon entropy

$$S_{M|E} = S - S_E = -\int \mathrm{d}\Gamma_E\, \rho_E \int \mathrm{d}\Gamma_M\; \rho_{M|E}\, \ln \rho_{M|E}, \tag{A.116}$$

with the Shannon entropy of subsystem $\boldsymbol{E}$

$$S_E = -\int \mathrm{d}\Gamma_E\, \rho_E\, \ln \rho_E. \tag{A.117}$$

The last three equations imply for the effective entropy production rate

$$\dot{\Sigma}^{(E)} = \sum_{i=1}^{E} \beta_i \xi_i \int \mathrm{d}\Gamma_E \rho_E \left(\boldsymbol{v}_i + \frac{1}{\beta_i m_i} \partial_{\boldsymbol{v}_i} \ln \rho_E \right)^2 \geq 0. \tag{A.118}$$

The difference between the real (A.106) and effective heat current (A.115) reads

$$\dot{Q} - \dot{Q}^{(E)} \equiv \dot{Q} - \sum_{i=1}^{E} \dot{q}^{(i)} - \mathrm{d}_t S_{M|E} \equiv \dot{q}^{(M)} - \mathrm{d}_t S_{M|E}, \tag{A.119}$$

where the additive components of the heat current can, as in the two-particle case (4.176), be rewritten as follows

$$\dot{q}^{(i)} = \int \mathrm{d}\Gamma_i\, e_i\, \dot{\rho}_i - \int \mathrm{d}\Gamma_i\; \rho_i\; \boldsymbol{v}_i \cdot \left(\boldsymbol{f}_i + \boldsymbol{f}^{(i)} \right). \tag{A.120}$$

Here, the nonconservative forces $\boldsymbol{f}^{(i)}$ in the subsystem $\boldsymbol{M}$ read

$$\boldsymbol{f}^{(i)}(\Gamma_M, t) = -\int \mathrm{d}\Gamma_E\; \rho_{E|M}\; \partial_{\boldsymbol{x}_i} u^{int}_{EM}, \tag{A.121}$$

for $i \in \boldsymbol{M}$.

Next, we have for the difference between the entropy production rates

$$\dot{\Sigma} - \dot{\Sigma}^{(E)} = \int \mathrm{d}\Gamma_E\; \rho_E\, \dot{\mathbb{\Sigma}}_E = \int \mathrm{d}\Gamma_E\; \rho_E \left(\dot{\mathbb{\Sigma}}'_E + \dot{\mathbb{\Sigma}}''_E \right), \tag{A.122}$$

swith the non-negative integral kernels

$$\dot{\mathbb{\Sigma}}'_E = \sum_{i=N-M}^{N} \beta_i\, \xi_i \int \mathrm{d}\Gamma_M\; \rho_{M|E} \left(\boldsymbol{v}_i + \frac{1}{\beta_i m_i} \partial_{\boldsymbol{v}_i} \ln \rho_{M|E} \right)^2 \tag{A.123}$$

$$\dot{\mathbb{\Sigma}}''_E = \sum_{i=1}^{E} \frac{\xi_i}{\beta_i m_i^2} \int \mathrm{d}\Gamma_M\; \rho_{M|E} \left(\partial_{\boldsymbol{v}_i} \ln \rho_{M|E} \right)^2, \tag{A.124}$$

and we note that the following relation holds true

$$\int \mathrm{d}\Gamma_E\; \rho_E\, \dot{\mathbb{\Sigma}}'_E = \sum_{i=N-M}^{N} \dot{\sigma}^{(i)} = \dot{\Sigma} - \dot{\sigma}^{(E)} \equiv \dot{\sigma}^{(M)}. \tag{A.125}$$

Bipartite System

The relative entropy between subsystem $\boldsymbol{E}$ and subsystem $\boldsymbol{M}$ reads

$$I = S_{\boldsymbol{M}} + S_{\boldsymbol{E}} - S = \int \mathrm{d}\boldsymbol{\Gamma}\, \rho \, \ln \frac{\rho}{\rho_{\boldsymbol{E}}\, \rho_{\boldsymbol{M}}} \geq 0, \tag{A.126}$$

and its time-derivative can be decomposed into two directional information flows

$$\dot{I}^{(\boldsymbol{M}\to\boldsymbol{E})} = \int \mathrm{d}\boldsymbol{\Gamma}_{\boldsymbol{E}}\, \rho_{\boldsymbol{E}} \left(\sum_{i=1}^{E} \frac{1}{m_i} \partial_{\boldsymbol{v}_i} \cdot \boldsymbol{f}^{(i)} - \dot{\mathbb{\Sigma}}''_{\boldsymbol{E}} \right) \tag{A.127}$$

$$\dot{I}^{(\boldsymbol{E}\to\boldsymbol{M})} = \int \mathrm{d}\boldsymbol{\Gamma}_{\boldsymbol{M}}\, \rho_{\boldsymbol{M}} \left(\sum_{i=N-M}^{N} \frac{1}{m_i} \partial_{\boldsymbol{v}_i} \cdot \boldsymbol{f}^{(i)} - \dot{\mathbb{\Sigma}}''_{\boldsymbol{M}} \right), \tag{A.128}$$

where we introduced the integral kernel specifying the difference between the entropy production rates

$$\dot{\Sigma} - \dot{\Sigma}^{(\boldsymbol{M})} = \int \mathrm{d}\boldsymbol{\Gamma}_{\boldsymbol{M}}\, \rho_{\boldsymbol{M}}\, \dot{\mathbb{\Sigma}}_{\boldsymbol{M}} = \int \mathrm{d}\boldsymbol{\Gamma}_{\boldsymbol{M}}\, \rho_{\boldsymbol{M}} (\dot{\mathbb{\Sigma}}'_{\boldsymbol{M}} + \dot{\mathbb{\Sigma}}''_{\boldsymbol{M}}), \tag{A.129}$$

with

$$\dot{\mathbb{\Sigma}}'_{\boldsymbol{M}} = \sum_{i=1}^{E} \beta_i\, \xi_i \int \mathrm{d}\boldsymbol{\Gamma}_{\boldsymbol{E}}\, \rho_{\boldsymbol{E}|\boldsymbol{M}} \left(\boldsymbol{v}_i + \frac{1}{\beta_i m_i} \partial_{\boldsymbol{v}_i} \ln \rho_{\boldsymbol{E}|\boldsymbol{M}} \right)^2 \tag{A.130}$$

$$\dot{\mathbb{\Sigma}}''_{\boldsymbol{M}} = \sum_{i=N-M}^{N} \frac{\xi_i}{\beta_i m_i^2} \int \mathrm{d}\boldsymbol{\Gamma}_{\boldsymbol{E}}\, \rho_{\boldsymbol{E}|\boldsymbol{M}} \left(\partial_{\boldsymbol{v}_i} \ln \rho_{\boldsymbol{E}|\boldsymbol{M}} \right)^2. \tag{A.131}$$

We note the following relation

$$\int \mathrm{d}\boldsymbol{\Gamma}_{\boldsymbol{M}}\, \rho_{\boldsymbol{M}}\, \dot{\mathbb{\Sigma}}'_{\boldsymbol{M}} = \dot{\sigma}^{(\boldsymbol{E})}. \tag{A.132}$$

The first term in the directional information flow, (A.127) and (A.128), represents the force contribution $\dot{I}_F$ while the second term is its entropic part $\dot{I}_S$. From the identities

$$\dot{I}_S^{(\boldsymbol{E}\to\boldsymbol{M})} = \dot{\Sigma}^{(\boldsymbol{M})} - \dot{\sigma}^{(\boldsymbol{M})} \tag{A.133}$$

$$\dot{I}_S^{(\boldsymbol{M}\to\boldsymbol{E})} = \dot{\Sigma}^{(\boldsymbol{E})} - \dot{\sigma}^{(\boldsymbol{E})}, \tag{A.134}$$

follows, using Eq. (A.114), the effective entropy balance equations for the two parts of the bipartite system, $\boldsymbol{E}$ and $\boldsymbol{M}$,

$$\mathrm{d}_t S_{\boldsymbol{E}} = \sum_{i=1}^{E} \beta_i\, \dot{q}^{(i)} + \dot{\sigma}^{(\boldsymbol{E})} + \dot{I}^{(\boldsymbol{M}\to\boldsymbol{E})} \tag{A.135}$$

$$\mathrm{d}_t S_{\boldsymbol{M}} = \sum_{i=N-M}^{N} \beta_i\, \dot{q}^{(i)} + \dot{\sigma}^{(\boldsymbol{M})} + \dot{I}^{(\boldsymbol{E}\to\boldsymbol{M})}. \tag{A.136}$$

Hamiltonian of Mean Force

In the following, we assume that the system is isothermal, $\beta_i = \beta \; \forall i$, that the nonconservative forces inside the subsystem $\boldsymbol{M}$ are zero, $\boldsymbol{f}^{(i)} = 0$, $i \in \boldsymbol{M}$ and that the Hamiltonian of the subsystem $\boldsymbol{M}$ is time-independent, $\partial_t e_{\boldsymbol{M}} = \partial_t u_{\boldsymbol{M}} = 0$. Following the same reasoning as in Sec. 4.3.2.2, we impose the following initial condition

$$\rho(0) = \rho_{\boldsymbol{E}}(0)\, \rho^{eq}_{\boldsymbol{M}|\boldsymbol{E}}(0) = \rho_{\boldsymbol{E}}(0)\, \mathrm{e}^{-\beta\left(e - A^{eq}_{\boldsymbol{M}|\boldsymbol{E}}\right)}, \tag{A.137}$$

where $\rho_{\boldsymbol{E}}(0)$ is an arbitrary initial state and $A^{eq}_{\boldsymbol{M}|\boldsymbol{E}}$ the free-energy landscape of the subsystem $\boldsymbol{E}$ for a locally equilibrated subsystem $\boldsymbol{M}$,

$$A^{eq}_{\boldsymbol{M}|E}(\boldsymbol{\Gamma}_{\boldsymbol{E}}) = e_{\boldsymbol{E}} - \beta^{-1} \ln \int \mathrm{d}\boldsymbol{\Gamma}_{\boldsymbol{M}}\, \mathrm{e}^{-\beta\left(e_{\boldsymbol{M}} + u^{int}_{\boldsymbol{E}\boldsymbol{M}} - A^{eq}_{\boldsymbol{M}}\right)} + A^{eq}_{\boldsymbol{M}}, \tag{A.138}$$

whereas $A^{eq}_{\boldsymbol{M}}$ is the equilibrium free energy of $\boldsymbol{M}$. Next, the Hamiltonian of mean force is defined as

$$H^{hmf} \equiv e_{\boldsymbol{E}} - \beta^{-1} \ln \langle \mathrm{e}^{-\beta u^{int}_{\boldsymbol{E}\boldsymbol{M}}} \rangle^{eq}_{\boldsymbol{M}} = A^{eq}_{\boldsymbol{M}|\boldsymbol{E}} - A^{eq}_{\boldsymbol{M}}, \tag{A.139}$$

which in turn allows us to rewrite Eq. (A.137) as follows

$$\rho(0) = \rho_{\boldsymbol{E}}(0)\, \mathrm{e}^{-\beta\left(e - H^{hmf} - A^{eq}_{\boldsymbol{M}}\right)}, \tag{A.140}$$

Next, we employ the definitions

$$S^{hmf}(t) \equiv S_{\boldsymbol{E}}(t) + \beta^2 \langle \partial_\beta H^{hmf} \rangle(t), \tag{A.141}$$

and

$$E^{hmf}(t) \equiv \langle \partial_\beta \left(\beta H^{hmf}\right) \rangle(t). \tag{A.142}$$

Using the definition of work that coincides with the one of the full system in Eq. (A.105),

$$W^{hmf}(t) \equiv \int_0^t \mathrm{d}t \left[\langle \dot{e} \rangle(t) + \sum_{i=1}^{E} \int \mathrm{d}\boldsymbol{\Gamma}_{\boldsymbol{E}}\, \rho_{\boldsymbol{E}}\, \boldsymbol{v}_i \cdot \boldsymbol{f}_i \right], \tag{A.143}$$

imposes the following definition for the heat

$$Q^{hmf}(t) \equiv -W(t) + \langle \partial_\beta \left(\beta H^{hmf}\right) \rangle(t) - \langle \partial_\beta \left(\beta H^{hmf}\right) \rangle(0). \tag{A.144}$$

With the nonequilibrium free energy

$$A^{hmf}(t) \equiv E^{hmf}(t) - \frac{S^{hmf}(t)}{\beta} = \langle H^{hmf} \rangle(t) - \frac{S_{\boldsymbol{E}}(t)}{\beta}, \tag{A.145}$$

we can rewrite the entropy balance

$$\Delta S^{hmf}(t) = \beta Q^{hmf}(t) + \Sigma^{hmf}(t) \geq 0. \tag{A.146}$$

in the form of a second law of thermodynamics as follows

$$\Sigma^{hmf}(t) = \beta\left[W(t) - \Delta A^{hmf}(t)\right] \geq 0. \tag{A.147}$$

To proof that inequality, we first note that along the lines of the derivation of Eq. (4.217) we can obtain

$$\Sigma^{hmf}(t) - \Sigma(t) = D[\rho(t) \,||\, \rho^{eq}_{\boldsymbol{M}|\boldsymbol{E}}\, \rho_{\boldsymbol{E}}(t)] \geq 0, \tag{A.148}$$

which, because of Eq. (A.109), proves the inequality in Eq. (A.147).

Limiting Cases

Fast-Dynamics Limit

We assume time-scale separation between the stochastic dynamics of subsystems $\boldsymbol{E}$ and $\boldsymbol{M}$, with $\boldsymbol{M}$ evolving much faster than $\boldsymbol{E}$, and furthermore take $\boldsymbol{f}_i = 0$, $i \in \boldsymbol{M}$ and $\beta_i = \beta\, \forall i$. As a result, the subsystem $\boldsymbol{M}$ instantaneously equilibrates with respect to the coordinates of $\boldsymbol{E}$ and the conditional probability is given by

$$\rho^{eq}_{\boldsymbol{M}|\boldsymbol{E}}(\boldsymbol{x}_{\boldsymbol{E}}, \boldsymbol{\Gamma}_{\boldsymbol{M}}) = \mathrm{e}^{-\beta\left(e - A^{eq}_{\boldsymbol{M}|\boldsymbol{E}}\right)}. \tag{A.149}$$

Consequently, the effective forces $\boldsymbol{f}^{(i)}$ in subsystem $\boldsymbol{E}$ ($i \in \boldsymbol{E}$) become conservative and the total force acting on subsystem $\boldsymbol{E}$ is generated by the free-energy landscape of the latter,

$$\left[\boldsymbol{f}^{(i)} - u_{\boldsymbol{E}} - u^{int}_{\boldsymbol{E}}\right]\Big|_{tss} = -\partial_{\boldsymbol{x}_i} A^{eq}_{\boldsymbol{M}|\boldsymbol{E}}. \tag{A.150}$$

Marginalization Inserting Eq. (A.149) into Eq. (A.116), we obtain

$$\beta^{-1}\, \mathrm{d}_t S_{\boldsymbol{M}|\boldsymbol{E}}\big|_{tss} = \int \mathrm{d}\boldsymbol{\Gamma}\, [\partial_t \rho_{\boldsymbol{E}}(t)]\, \rho^{eq}_{\boldsymbol{M}|\boldsymbol{E}} \left(e - A^{eq}_{\boldsymbol{M}|\boldsymbol{E}}\right) = \int \mathrm{d}\boldsymbol{\Gamma}\, [\partial_t \rho(t)]\, e - \int \mathrm{d}\boldsymbol{\Gamma}_{\boldsymbol{E}}\, [\partial_t \rho_{\boldsymbol{E}}(t)]\, A^{eq}_{\boldsymbol{M}|\boldsymbol{E}}. \tag{A.151}$$

From Eqs. (A.120) and (A.138) follows the relation

$$\dot{q}^{(\boldsymbol{E})}\Big|_{tss} = \int \mathrm{d}\boldsymbol{\Gamma}_{\boldsymbol{E}}\, [\partial_t \rho_{\boldsymbol{E}}(t)]\, A^{eq}_{\boldsymbol{M}|\boldsymbol{E}}, \tag{A.152}$$

and with Eq. (A.115) we get

$$\dot{Q}^{(\boldsymbol{E})}\Big|_{tss} = \dot{q}^{(\boldsymbol{E})}\Big|_{tss} + \beta^{-1}\mathrm{d}_t S_{\boldsymbol{M}|\boldsymbol{E}}\big|_{tss} = \dot{Q}\big|_{tss}\,, \tag{A.153}$$

and therefore prove that within time-scale separation, the equivalence of the effective and real first law,

$$\mathrm{d}_t E|_{tss} = \dot{Q}^{(\boldsymbol{E})}\Big|_{tss} + \dot{W}\big|_{tss} = \dot{Q}\big|_{tss} + \dot{W}\big|_{tss}\,. \tag{A.154}$$

as well as that the effective and real second law agree,

$$\dot{\Sigma}^{(\boldsymbol{E})}\Big|_{tss} = \mathrm{d}_t S|_{tss} - \beta\dot{Q}^{(\boldsymbol{E})}\Big|_{tss} = \mathrm{d}_t S|_{tss} - \beta\,\dot{Q}\big|_{tss} = \dot{\Sigma}\big|_{tss} \geq 0. \tag{A.155}$$

Physically, the coarse-grained system $\boldsymbol{M}$ corresponds to an ideal heat reservoir the subsystem $\boldsymbol{E}$ is coupled with.

Bipartite System Moreover, substituting (A.149) into Eq. (A.127), shows that the directional information flow from the fast $\boldsymbol{M}$ to the slow subsystem $\boldsymbol{E}$ is identically zero,

$$\dot{I}_F^{(\boldsymbol{M}\to\boldsymbol{E})}\Big|_{tss} = \dot{I}_S^{(\boldsymbol{M}\to\boldsymbol{E})}\Big|_{tss} = \dot{I}^{(\boldsymbol{M}\to\boldsymbol{E})}\Big|_{tss} = 0, \tag{A.156}$$

Hence both the additive and effective entropy production rate (A.118) agree with the real one (A.109),

$$\dot{\sigma}^{(\boldsymbol{E})}\Big|_{tss} = \dot{\Sigma}^{(\boldsymbol{E})}\Big|_{tss} = \dot{\Sigma}\big|_{tss}\,, \tag{A.157}$$

which implies that $\dot{\sigma}^{(\boldsymbol{M})}\big|_{tss} = 0$.

Yet, the effective entropy balance of the slow particle (A.135) deviates from the real entropy balance (A.108) by the conditional Shannon entropy (A.151),

$$\mathrm{d}_t S_{\boldsymbol{E}}|_{tss} = \beta\dot{q}^{(\boldsymbol{E})}\Big|_{tss} + \dot{\sigma}^{(\boldsymbol{E})}\Big|_{tss} = \mathrm{d}_t S|_{tss} - \mathrm{d}_t S_{\boldsymbol{M}|\boldsymbol{E}}\big|_{tss}\,, \tag{A.158}$$

such that, using Eqs. (A.158) and (A.153), the effective entropy balance of the second particle reads

$$\mathrm{d}_t S_{\boldsymbol{M}}|_{tss} - \mathrm{d}_t\, S_{\boldsymbol{M}|\boldsymbol{E}}\big|_{tss} = \mathrm{d}_t I^{(\boldsymbol{E}\to\boldsymbol{M})}\Big|_{tss}\,. \tag{A.159}$$

Thus, the totally asymmetric information flow $\mathrm{d}_t I|_{tss} = \mathrm{d}_t I^{(\boldsymbol{E}\to\boldsymbol{M})}\big|_{tss}$ from the slowly to the quickly evolving subsystem is finite. This is physically plausible since the subsystems are still correlated. The information flow corresponds to time-varying correlations between the two subsets due to the changes of their out-of-equilibrium probability distributions. This means that at a global equilibrium state, $\rho^{eq} = \rho^{eq}_{\boldsymbol{M}|\boldsymbol{E}}\,\rho^{eq}_{\boldsymbol{E}}$, the information flow is zero.

Hamiltonian of Mean Force We note the standard equilibrium identities

$$A^{eq}_{\boldsymbol{M}|\boldsymbol{E}} = E^{eq}_{\boldsymbol{M}|\boldsymbol{E}} - \beta^{-1} S^{eq}_{\boldsymbol{M}|\boldsymbol{E}}, \tag{A.160}$$

$$E^{eq}_{\boldsymbol{M}|\boldsymbol{E}} = \partial_\beta(\beta A^{eq}_{\boldsymbol{M}|\boldsymbol{E}}) = \int \mathrm{d}\boldsymbol{\Gamma}_{\boldsymbol{M}}\, \rho^{eq}_{\boldsymbol{M}|\boldsymbol{E}}\, e \tag{A.161}$$

$$S^{eq}_{\boldsymbol{M}|\boldsymbol{E}} = \beta^2 \partial_\beta A^{eq}_{\boldsymbol{M}|\boldsymbol{E}} = -\int \mathrm{d}\boldsymbol{\Gamma}_{\boldsymbol{M}}\, \rho^{eq}_{\boldsymbol{M}|\boldsymbol{E}}\, \ln \rho^{eq}_{\boldsymbol{M}|\boldsymbol{E}}, \tag{A.162}$$

which, using Eq. (A.139), can be rewritten as

$$E^{eq}_{\boldsymbol{M}|\boldsymbol{E}} = \partial_\beta \left[\beta \left(H^{hmf} + A^{eq}_{\boldsymbol{M}}\right)\right] \tag{A.163}$$

$$S^{eq}_{\boldsymbol{M}|\boldsymbol{E}} = \beta^2\, \partial_\beta \left(H^{hmf} + A^{eq}_{\boldsymbol{M}}\right). \tag{A.164}$$

In the limit of time-scale separation and local equilibrium of the subsystem $\boldsymbol{M}$, Eqs. (A.163) and (A.164) are valid at any time t. A comparison with Eqs. (A.142) and (A.141), respectively, shows that

$$E^{hmf}(t)\big|_{tss} = E(t)|_{tss} - \partial_\beta(\beta\, A^{eq}_{\boldsymbol{M}}) \tag{A.165}$$

$$S^{hmf}(t)\big|_{tss} = S(t)|_{tss} - \beta^2\, \partial_\beta\, A^{eq}_{\boldsymbol{M}}. \tag{A.166}$$

Since we choose a time-independent Hamiltonian $e_{\boldsymbol{M}}$ and the definitions of the Hamiltonian and mean force (A.143) and the full work agree (A.105), it follows that the definitions of heat *current* coincide

$$\dot{Q}^{hmf}(t)\big|_{tss} = \mathrm{d}_t E^{hmf}(t)\big|_{tss} - \dot{W}(t)\big|_{tss} = \dot{Q}(t)\big|_{tss}\,. \tag{A.167}$$

According to Eqs. (A.165) and (4.237) the entropy production *rates* are also identical,

$$\dot{\Sigma}^{hmf}(t)\big|_{tss} = \mathrm{d}_t S^{hmf}(t)\big|_{tss} - \beta \dot{Q}^{hmf}(t)\big|_{tss} = \dot{\Sigma}(t)\big|_{tss}\,, \tag{A.168}$$

hence the Hamiltonian of mean-force formalism captures the full differential thermodynamics in the limit of time-scale separation. Furthermore, with Eq. (A.155) we have proven that

$$\dot{\Sigma}(t)\big|_{tss} = \dot{\Sigma}^{(1)}(t)\Big|_{tss} = \dot{\sigma}^{(1)}(t)\Big|_{tss} = \dot{\Sigma}^{hmf}(t)\big|_{tss}\,. \tag{A.169}$$

Large-Mass Limit

We now study the limit of diverging masses $m_i \to \infty$ for $i \in \boldsymbol{M}$ and assume, to avoid trivialities, the following scaling: $\mathbb{O}(\partial_{\boldsymbol{x}_i} u_i / m_i) = \mathbf{1}$ while $\partial_{\boldsymbol{x}_i} u^{int}/m_i \to 0$ as $m_i \to \infty$, for $i \in \boldsymbol{M}$. Consequently, the distributions $\rho_{\boldsymbol{E}}$ and $\rho_{\boldsymbol{M}}$ become statistically independent and the conditional distribution reads

$$\rho^{det}_{\boldsymbol{M}|\boldsymbol{E}}(\boldsymbol{\Gamma}_{\boldsymbol{M}}, t) = \sum_{i=N-M}^{N} \delta\left(\boldsymbol{x}_i - \boldsymbol{x}_t^{(i)}\right) \delta\left(\boldsymbol{v}_i - \boldsymbol{v}_t^{(i)}\right), \tag{A.170}$$

where $\boldsymbol{x}_t^{(i)}$ and $\boldsymbol{v}_t^{(i)}$ are the solutions of the deterministic equations of motion (4.142) in the subsystem $\boldsymbol{M}$. As a result, the effective forces (A.113) in subsystem $\boldsymbol{E}$ ($i \in \boldsymbol{E}$) are conservative,

$$\boldsymbol{f}^{(i)}(\boldsymbol{x}_{\boldsymbol{E}}, t)\Big|_{det} = -\partial_{\boldsymbol{x}_i} u^{int}_{\boldsymbol{EM}}(\boldsymbol{x}_{\boldsymbol{E}}, \boldsymbol{x}_{\boldsymbol{M}}, t)\Big|_{\boldsymbol{x}_{\boldsymbol{M}}=\boldsymbol{x}_t^{(M)}} . \tag{A.171}$$

Marginalization. Obviously, $S_{\boldsymbol{M}|\boldsymbol{E}}\big|_{det} = 0$, that is the naive definition for the system entropy in $\boldsymbol{E}$, (A.117), and the real system entropy (A.107) agree,

$$\mathrm{d}_t S_{\boldsymbol{E}}|_{det} = \mathrm{d}_t S|_{det} , \tag{A.172}$$

such that we find for the difference between the real (A.106) and the effective heat current (A.115),

$$\dot{Q}\big|_{det} - \dot{q}^{(\boldsymbol{E})}\Big|_{det} = \sum_{i=N-M}^{N} \left[\dot{q}^{(i)} - \xi_i \left(\boldsymbol{v}_t^{(i)}\right)^2\right]\Big|_{det} , \tag{A.173}$$

and for the difference between the real (A.109) and the effective entropy production rate (A.118),

$$\dot{\Sigma}\big|_{det} - \dot{\Sigma}^{(\boldsymbol{E})}\Big|_{det} = \sum_{i=N-M}^{N} \beta_i \left(\boldsymbol{v}_t^{(i)}\right)^2 . \tag{A.174}$$

This means that the effective thermodynamics defined for the subsystem $\boldsymbol{E}$ agrees - up to the cumulated macroscopic friction associated with the particles in $\boldsymbol{M}$ - with the full one. If the subsystem $\boldsymbol{M}$ is furthermore Hamiltonian, $\xi_i = 0 \; \forall i \in \boldsymbol{M}$, the work sources are non-dissipative and the effective description coincides with the full one. Physically, the particles of $\boldsymbol{M}$ represent M independent work sources that modulate the energy landscape of the subsystem $\boldsymbol{E}$ according to a protocol $\boldsymbol{x}_t^{(\boldsymbol{M})}, \boldsymbol{v}_t^{(\boldsymbol{M})}$.

Owing to the statistical independence of $\rho_{\boldsymbol{E}}$ and $\rho_{\boldsymbol{M}}$, the mutual information and thus the information flow is identically zero,

$$I|_{det} = \dot{I}^{\boldsymbol{M}\to\boldsymbol{E}}\big|_{det} = \dot{I}^{\boldsymbol{E}\to\boldsymbol{M}}\big|_{det} = 0. \tag{A.175}$$

Therefore, in the large-mass limit the effective entropy balance of the subsystem $\boldsymbol{E}$ coincides with the full one,

$$\mathrm{d}_t S_{\boldsymbol{E}}|_{det} = \sum_{i=1}^{E} \beta_i \, \dot{q}^{(i)}\Big|_{det} + \dot{\sigma}^{(\boldsymbol{E})}\Big|_{det} = \mathrm{d}_t S|_{tss} , \tag{A.176}$$

while the corresponding effective entropy balance equation for the subsystem $\boldsymbol{M}$ becomes macroscopic and thus a triviality,

$$\dot{\sigma}^{(\boldsymbol{M})}\Big|_{det} = -\sum_{i=N-M}^{N} \beta_i \, \dot{q}^{(i)}\Big|_{det} = \sum_{i=N-M}^{N} \beta_i \left(\boldsymbol{v}_t^{(i)}\right)^2 . \tag{A.177}$$

Hamiltonian of Mean Force. As already pointed out in the two-particle case, the Hamiltonian of mean force formalism and the deterministic limit are incompatible since the assumption of a conditional Gibbs state (A.137) is inconsistent with the independent subsystem distributions (A.170).

To sum up, the generalization of the two-particle discussion to multiple particles is formally straightforward and leads to identical results.

Bibliography

[1] T. Herpich, J. Thingna, and M. Esposito. "Collective Power: Minimal Model for Thermodynamics of Nonequilibrium Phase Transitions". In: *Phys. Rev. X* **8**, 031056 (2018). DOI: 10.1103/PhysRevX.8.031056.

[2] T. Herpich and M. Esposito. "Universality in driven Potts models". In: *Phys. Rev. E* **99**, 022135 (2019). DOI: 10.1103/PhysRevE.99.022135.

[3] S. Sheng, T. Herpich, G. Diana, and M. Esposito. "Thermodynamics of Majority-Logic Decoding in Information Erasure". In: *Entropy* **21**, 284 (2019). DOI: 10.3390/e21030284.

[4] T. Herpich, K. Shayanfard, and M. Esposito. "Effective thermodynamics of two interacting underdamped Brownian particles". In: *Phys. Rev. E* **101**, 022116 (2020). DOI: 10.1103/PhysRevE.101.022116.

[5] T. Herpich, T. Cossetto, G. Falasco, and M. Esposito. *Stochastic thermodynamics of all-to-all interacting many-body systems*. 2020. arXiv: 2001.09744.

[6] S. Carnot. *Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance*. Bachelier, 1824.

[7] S. Carnot. *From Watt to Clausius: The Rise of Thermodynamics in the Early Industrial Age*. Iowa State University Press, 1971.

[8] H. B. Callen. *Thermodynamics and an Introduction to Thermostatistics*. 2nd ed. John Wiley & Sons, 1985.

[9] L. Onsager. "Reciprocal Relations in Irreversible Processes. I." In: *Phys. Rev.* **37**, 405–426 (1931). DOI: 10.1103/PhysRev.37.405.

[10] L. Onsager. "Reciprocal Relations in Irreversible Processes. II." In: *Phys. Rev.* **38**, 2265–2279 (1931). DOI: 10.1103/PhysRev.38.2265.

[11] S. De Groot and P. Mazur. *Non-Equilibrium Thermodynamics*. Dover Publication, New York, 1984.

[12] D. Kondepudi and I. Prigogine. *Modern Thermodynamics: From Heat Engines to Dissipative Structures*. 2nd ed. John Wiley & Sons, 2014. DOI: 10.1002/9781118698723.

[13] U. Seifert. "Stochastic thermodynamics, fluctuation theorems and molecular machines". In: *Reports on Progress in Physics* **75**, 126001 (2012). DOI: 10.1088/0034-4885/75/12/126001.

[14] C. Van den Broeck and M. Esposito. "Ensemble and trajectory thermodynamics: A brief introduction". In: *Physica A: Statistical Mechanics and its Applications* **418**, 6–16 (2015). DOI: 10.1016/j.physa.2014.04.035.

[15] K. Sekimoto. *Stochastic Energetics*. 1st ed. Springer-Verlag, 2010.

[16] X.-J. Zhang, H. Qian, and M. Qian. "Stochastic theory of nonequilibrium steady states and its applications. Part I". In: *Physics Reports* **510**, 1–86 (2012). DOI: 10.1016/j.physrep.2011.09.002.

[17] H. Ge, M. Qian, and H. Qian. "Stochastic theory of nonequilibrium steady states. Part II: Applications in chemical biophysics". In: *Physics Reports* **510**, 87–118 (2012). DOI: 10.1016/j.physrep.2011.09.001.

[18] C. Jarzynski. "Equalities and Inequalities: Irreversibility and the Second Law of Thermodynamics at the Nanoscale". In: *Annual Review of Condensed Matter Physics* **2**, 329–351 (2011). DOI: 10.1146/annurev-conmatphys-062910-140506.

[19] D. J. Evans and D. J. Searles. "Equilibrium microstates which generate second law violating steady states". In: *Phys. Rev. E* **50**, 1645–1648 (1994). DOI: 10.1103/PhysRevE.50.1645.

[20] U. Seifert. "Entropy Production along a Stochastic Trajectory and an Integral Fluctuation Theorem". In: *Phys. Rev. Lett.* **95**, 040602 (2005). DOI: 10.1103/PhysRevLett.95.040602.

[21] J. Parrondo, J. Horowitz, and T. Sagawa. "Thermodynamics of information". In: *Nature Physics* **11**, 131–139 (2015). DOI: 10.1038/nphys3230.

[22] R. Rao and M. Esposito. "Nonequilibrium Thermodynamics of Chemical Reaction Networks: Wisdom from Stochastic Thermodynamics". In: *Phys. Rev. X* **6**, 041064 (2016). DOI: 10.1103/PhysRevX.6.041064.

[23] C. Nardini, É. Fodor, E. Tjhung, F. van Wijland, J. Tailleur, and M. E. Cates. "Entropy Production in Field Theories without Time-Reversal Symmetry: Quantifying the Non-Equilibrium Character of Active Matter". In: *Phys. Rev. X* **7**, 021007 (2017). DOI: 10.1103/PhysRevX.7.021007.

[24] L. Dabelow, S. Bo, and R. Eichhorn. "Irreversibility in Active Matter Systems: Fluctuation Theorem and Mutual Information". In: *Phys. Rev. X* **9**, 021009 (2019). DOI: 10.1103/PhysRevX.9.021009.

[25] J. P. Pekola, O.-P. Saira, V. F. Maisi, A. Kemppinen, M. Möttönen, Y. A. Pashkin, and D. V. Averin. "Single-electron current sources: Toward a refined definition of the ampere". In: *Rev. Mod. Phys.* **85**, 1421–1472 (2013). DOI: 10.1103/RevModPhys.85.1421.

[26] S. Ciliberto. "Experiments in Stochastic Thermodynamics: Short History and Perspectives". In: *Phys. Rev. X* **7**, 021051 (2017). DOI: 10.1103/PhysRevX.7.021051.

[27] M. Esposito, K. Lindenberg, and C. Van den Broeck. "Universality of Efficiency at Maximum Power". In: *Phys. Rev. Lett.* **102**, 130602 (2009). DOI: 10.1103/PhysRevLett.102.130602.

[28] G. Verley, T. Willaert, C. V. d. Broeck, and M. Esposito. "The unlikely Carnot efficiency". In: *Nature Communications* **5** (2014).

[29] M. Polettini and M. Esposito. "Effective Thermodynamics for a Marginal Observer". In: *Phys. Rev. Lett.* **119**, 240601 (2017). DOI: 10.1103/PhysRevLett.119.240601.

[30] N. Shiraishi, K. Saito, and H. Tasaki. "Universal Trade-Off Relation between Power and Efficiency for Heat Engines". In: *Phys. Rev. Lett.* **117**, 190601 (2016). DOI: 10.1103/PhysRevLett.117.190601.

[31] P. Pietzonka and U. Seifert. "Universal Trade-Off between Power, Efficiency, and Constancy in Steady-State Heat Engines". In: *Phys. Rev. Lett.* **120**, 190602 (2018). DOI: 10.1103/PhysRevLett.120.190602.

[32] J. P. Pekola. "Towards quantum thermodynamics in electronic circuits". In: *Nature Physics* **11**, 118–123 (2015). DOI: 10.1038/nphys3169.

[33] M. Esposito, K. Lindenberg, and C. V. den Broeck. "Thermoelectric efficiency at maximum power in a quantum dot". In: *EPL (Europhysics Letters)* **85**, 60010 (2009). DOI: 10.1209/0295-5075/85/60010.

[34] G. B. Cuetara and M. Esposito. "Double quantum dot coupled to a quantum point contact: a stochastic thermodynamics approach". In: *New Journal of Physics* **17**, 095005 (2015). DOI: 10.1088/1367-2630/17/9/095005.

[35] B. Rutten, M. Esposito, and B. Cleuren. "Reaching optimal efficiencies using nanosized photoelectric devices". In: *Phys. Rev. B* **80**, 235122 (2009). DOI: 10.1103/PhysRevB.80.235122.

[36] P. Gaspard and E. Gerritsma. "The stochastic chemomechanics of the F1-ATPase molecular motor". In: *Journal of Theoretical Biology* **247**, 672–686 (2007). DOI: 10.1016/j.jtbi.2007.03.034.

[37] A. W. C. Lau, D. Lacoste, and K. Mallick. "Nonequilibrium Fluctuations and Mechanochemical Couplings of a Molecular Motor". In: *Phys. Rev. Lett.* **99**, 158102 (2007). DOI: 10.1103/PhysRevLett.99.158102.

[38] U. Seifert. "Stochastic thermodynamics of single enzymes and molecular motors". In: *The European Physical Journal E* **34**, 26 (2011). DOI: 10.1140/epje/i2011-11026-7.

[39] B. Altaner, A. Wachtel, and J. Vollmer. "Fluctuating currents in stochastic thermodynamics. II. Energy conversion and nonequilibrium response in kinesin models". In: *Phys. Rev. E* **92**, 042133 (2015). DOI: 10.1103/PhysRevE.92.042133.

[40] K. Proesmans, B. Cleuren, and C. Van den Broeck. "Power-Efficiency-Dissipation Relations in Linear Thermodynamics". In: *Phys. Rev. Lett.* **116**, 220601 (2016). DOI: 10.1103/PhysRevLett.116.220601.

[41] J. S. Lee and H. Park. "Carnot efficiency is reachable in an irreversible process". In: *Scientific Reports* **7**, 2045–2322 (2017). DOI: 10.1038/s41598-017-10664-9.

[42] M. Polettini and M. Esposito. "Carnot efficiency at divergent power output". In: *EPL (Europhysics Letters)* **118**, 40003 (2017). DOI: 10.1209/0295-5075/118/40003.

[43] C. Van den Broeck. "Thermodynamic Efficiency at Maximum Power". In: *Phys. Rev. Lett.* **95**, 190602 (2005). DOI: 10.1103/PhysRevLett.95.190602.

[44] M. Esposito, R. Kawai, K. Lindenberg, and C. Van den Broeck. "Efficiency at Maximum Power of Low-Dissipation Carnot Engines". In: *Phys. Rev. Lett.* **105**, 150603 (2010). DOI: 10.1103/PhysRevLett.105.150603.

[45] M. Polettini, G. Verley, and M. Esposito. "Efficiency Statistics at All Times: Carnot Limit at Finite Power". In: *Phys. Rev. Lett.* **114**, 050601 (2015). DOI: 10.1103/PhysRevLett.114.050601.

[46] H. Stanley. *Introduction to Phase Transitions and Critical Phenomena*. 1st ed. Oxford University Press, 1971.

[47] N. Goldenfeld. *Lectures on Phase Transitions and the Renormalization Group*. Perseus Books, 1992.

[48] J. Yeomans. *Statistical Mechanics of Phase Transitions*. Oxford University Press, 1992.

[49] L. Landau and E. Lifshitz. *Course of Theoretical Physics Volum 5: Statistical Physics Part 1*. Pergamon Press, 1994.

[50] L. P. Kadanoff. "More is the same: phase transitions and mean field theories". In: *Journal of Statistical Physics* **137**, 777 (2009). DOI: 10.1007/s10955-009-9814-1.

[51] L. S. Schulman. "Magnetisation probabilities and metastability in the Ising model". In: *Journal of Physics A: Mathematical and General* **13**, 237–250 (1980). DOI: 10.1088/0305-4470/13/1/025.

[52] B. Gaveau and L. S. Schulman. "Dynamical metastability". In: *Journal of Physics A: Mathematical and General* **20**, 2865–2873 (1987). DOI: 10.1088/0305-4470/20/10/031.

[53] B. Schmittmann and R. Zia. "Statistical mechanics of driven diffusive systems". In: *Statistical Mechanics of Driven Diffusive System*. Vol. 17. Phase Transitions and Critical Phenomena. Academic Press, 1995, pp. 3–214. DOI: https://doi.org/10.1016/S1062-7901(06)80014-5.

[54] J.-P. Bouchaud, L. Cugliandolo, J. Kurchan, and M. Mézard. "Spin Glasses and Random Fields". In: *Series on Directions in Condensed Matter Physics*. Vol. 12. 1997, pp. 161–223. DOI: 10.1142/9789812819437_0006.

[55] G. Ódor. *Universality in Nonequilibrium Lattice Systems*. World Scientific, 2008. DOI: 10.1142/6813.

[56] J. A. Acebrón, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler. "The Kuramoto model: A simple paradigm for synchronization phenomena". In: *Rev. Mod. Phys.* **77**, 137–185 (2005). DOI: 10.1103/RevModPhys.77.137.

[57] B. Cleuren and C. V. den Broeck. "Ising model for a Brownian donkey". In: *Europhysics Letters (EPL)* **54**, 1–6 (2001). DOI: 10.1209/epl/i2001-00274-6.

[58] J. P. Garrahan, R. L. Jack, V. Lecomte, E. Pitard, K. van Duijvendijk, and F. van Wijland. "First-order dynamical phase transition in models of glasses: an approach based on ensembles of histories". In: *Journal of Physics A: Mathematical and Theoretical* **42**, 075007 (2009). DOI: 10.1088/1751-8113/42/7/075007.

[59] S. Vaikuntanathan, T. R. Gingrich, and P. L. Geissler. "Dynamic phase transitions in simple driven kinetic networks". In: *Phys. Rev. E* **89**, 062108 (2014). DOI: 10.1103/PhysRevE.89.062108.

[60] M. Marcuzzi, E. Levi, S. Diehl, J. P. Garrahan, and I. Lesanovsky. "Universal Nonequilibrium Properties of Dissipative Rydberg Gases". In: *Phys. Rev. Lett.* **113**, 210401 (2014). DOI: 10.1103/PhysRevLett.113.210401.

[61] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim. "Macroscopic fluctuation theory". In: *Rev. Mod. Phys.* **87**, 593–636 (2015). DOI: 10.1103/RevModPhys.87.593.

[62] R. L. Jack and P. Sollich. "Large Deviations and Ensembles of Trajectories in Stochastic Models". In: *Progress of Theoretical Physics Supplement* **184**, 304–317 (2010). DOI: 10.1143/PTPS.184.304.

[63] H. Vroylandt, M. Esposito, and G. Verley. "Collective effects enhancing power and efficiency". In: *EPL (Europhysics Letters)* **120**, 30009 (2017). DOI: 10.1209/0295-5075/120/30009/meta.

[64] N. Golubeva and A. Imparato. "Maximum power operation of interacting molecular motors". In: *Phys. Rev. E* **88**, 012114 (2013). DOI: 10.1103/PhysRevE.88.012114.

[65] N. Golubeva and A. Imparato. "Efficiency at Maximum Power of Interacting Molecular Machines". In: *Phys. Rev. Lett.* **109**, 190602 (2012). DOI: 10.1103/PhysRevLett.109.190602.

[66] A. Imparato. "Stochastic thermodynamics in many-particle systems". In: *New Journal of Physics* **17**, 125004 (2015). DOI: 10.1088/1367-2630/17/12/125004.

[67] C. E. F. Noa, P. E. Harunari, M. J. de Oliveira, and C. E. Fiore. "Entropy production as a tool for characterizing nonequilibrium phase transitions". In: *Phys. Rev. E* **100**, 012104 (2019). DOI: 10.1103/PhysRevE.100.012104.

[68] S. Sasa. "Collective dynamics from stochastic thermodynamics". In: *New Journal of Physics* **17**, 045024 (2015). DOI: 10.1088/1367-2630/17/4/045024.

[69] T. Tomé and M. J. de Oliveira. "Entropy Production in Nonequilibrium Systems at Stationary States". In: *Phys. Rev. Lett.* **108**, 020601 (2012). DOI: 10.1103/PhysRevLett.108.020601.

[70] L. Crochik and T. Tomé. "Entropy production in the majority-vote model". In: *Phys. Rev. E* **72**, 057103 (2005). DOI: 10.1103/PhysRevE.72.057103.

[71] Y. Zhang and A. C. Barato. "Critical behavior of entropy production and learning rate: Ising model with an oscillating field". In: *Journal of Statistical Mechanics: Theory and Experiment* **2016**, 113207 (2016).

[72] E. Crosato, R. E. Spinney, R. Nigmatullin, J. T. Lizier, and M. Prokopenko. "Thermodynamics and computation during collective motion near criticality". In: *Phys. Rev. E* **97**, 012120 (2018). DOI: 10.1103/PhysRevE.97.012120.

[73] C. Leduc, K. Padberg-Gehle, V. Varga, D. Helbing, S. Diez, and J. Howard. "Molecular crowding creates traffic jams of kinesin motors on microtubules". In: *Proceedings of the National Academy of Sciences of the United States of America* **109**, 6100 (2012). DOI: 10.1073/pnas.1107281109.

[74] O. Campàs, Y. Kafri, K. B. Zeldovich, J. Casademunt, and J.-F. Joanny. "Collective Dynamics of Interacting Molecular Motors". In: *Phys. Rev. Lett.* **97**, 038101 (2006). DOI: 10.1103/PhysRevLett.97.038101.

[75] A. Pikovsky, M. Rosenblum, and J. Kurths. *Synchronization: a universal concept in nonlinear sciences*. Vol. 12. Cambridge University Press, 2003. DOI: 10.1017/CBO9780511755743.

[76] C. Huygens. In: *Letters to de Sluse, (letters; no. 1333 of 24 February 1665, no. 1335 of 26 February 1665, no. 1345 of 6 March 1665)* (1895).

[77] S. Strogatz. *Sync: The emerging science of spontaneous order*. 1st ed. Penguin UK, 2004.

[78] A. T. Winfree. "Biological rhythms and the behavior of populations of coupled oscillators". In: *Journal of Theoretical Biology* **16**, 15–42 (1967). DOI: 10.1016/0022-5193(67)90051-3.

[79] Y. Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Dover Books on Chemistry Series. Dover Publications, 2003.

[80] S. H. Strogatz. "From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators". In: *Physica D: Nonlinear Phenomena* **143**, 1–20 (2000). DOI: 10.1016/S0167-2789(00)00094-4.

[81] H. Sakaguchi, S. Shinomoto, and Y. Kuramoto. "Local and Grobal Self-Entrainments in Oscillator Lattices". In: *Progress of Theoretical Physics* **77**, 1005–1010 (1987). DOI: 10.1143/PTP.77.1005.

[82] H. Daido. "Lower Critical Dimension for Populations of Oscillators with Randomly Distributed Frequencies: A Renormalization-Group Analysis". In: *Phys. Rev. Lett.* **61**, 231–234 (1988). DOI: 10.1103/PhysRevLett.61.231.

[83] D. Walgraef, G. Dewel, and P. Borckmans. "Chemical waves in a two-dimensional oscillating system". In: *The Journal of Chemical Physics* **78**, 3043–3051 (1983). DOI: 10.1063/1.445266.

[84] Y. Izumida, H. Kori, and U. Seifert. "Energetics of synchronization in coupled oscillators rotating on circular trajectories". In: *Phys. Rev. E* **94**, 052221 (2016). DOI: 10.1103/PhysRevE.94.052221.

[85] K. Wood, C. Van den Broeck, R. Kawai, and K. Lindenberg. "Universality of Synchrony: Critical Behavior in a Discrete Model of Stochastic Phase-Coupled Oscillators". In: *Phys. Rev. Lett.* **96**, 145701 (2006). DOI: 10.1103/PhysRevLett.96.145701.

[86] K. Wood, C. Van den Broeck, R. Kawai, and K. Lindenberg. "Critical behavior and synchronization of discrete stochastic phase-coupled oscillators". In: *Phys. Rev. E* **74**, 031113 (2006). DOI: 10.1103/PhysRevE.74.031113.

[87] K. Wood, C. Van den Broeck, R. Kawai, and K. Lindenberg. "Continuous and discontinuous phase transitions and partial synchronization in stochastic three-state oscillators". In: *Phys. Rev. E* **76**, 041132 (2007). DOI: 10.1103/PhysRevE.76.041132.

[88] N. Van Kampen. *Stochastic Processes in Physics and Chemistry*. 3rd ed. North Holland, 2007. DOI: 10.1016/B978-0-444-52965-7.X5000-4.

[89] B. Gaveau and L. S. Schulman. "Theory of nonequilibrium first-order phase transitions for stochastic dynamics". In: *Journal of Mathematical Physics* **39**, 1517–1533 (1998). DOI: 10.1063/1.532394.

[90] G. Biroli and J. Kurchan. "Metastable states in glassy systems". In: *Phys. Rev. E* **64**, 016101 (2001). DOI: 10.1103/PhysRevE.64.016101.

[91] K. Macieszczak, M. Guta, I. Lesanovsky, and J. P. Garrahan. "Towards a Theory of Metastability in Open Quantum Dynamics". In: *Phys. Rev. Lett.* **116**, 240404 (2016). DOI: 10.1103/PhysRevLett.116.240404.

[92] D. C. Rose, K. Macieszczak, I. Lesanovsky, and J. P. Garrahan. "Metastability in an open quantum Ising model". In: *Phys. Rev. E* **94**, 052132 (2016). DOI: 10.1103/PhysRevE.94.052132.

[93] D. H. Wolpert. "The stochastic thermodynamics of computation". In: *Journal of Physics A: Mathematical and Theoretical* **52**, 193001 (2019). DOI: 10.1088/1751-8121/ab0850.

[94] R. Landauer. "Irreversibility and Heat Generation in the Computing Process". In: *IBM J. Res. Dev.* **5**, 183–191 (1961). DOI: 10.1147/rd.53.0183.

[95] M. Esposito, R. Kawai, K. Lindenberg, and C. Van den Broeck. "Finite-time thermodynamics for a single-level quantum dot". In: *EPL (Europhysics Letters)* **89**, 20003 (2010). DOI: 10.1209/0295-5075/89/20003.

[96] G. Diana, G. B. Bagci, and M. Esposito. "Finite-time erasing of information stored in fermionic bits". In: *Phys. Rev. E* **87**, 012111 (2013). DOI: 10.1103/PhysRevE.87.012111.

[97] P. R. Zulkowski and M. R. DeWeese. "Optimal finite-time erasure of a classical bit". In: *Phys. Rev. E* **89**, 052140 (2014). DOI: 10.1103/PhysRevE.89.052140.

[98] M. Gavrilov, R. Chétrite, and J. Bechhoefer. "Direct measurement of weakly nonequilibrium system entropy is consistent with Gibbs–Shannon form". In: *Proceedings of the National Academy of Sciences* **114**, 11097–11102 (2017). DOI: 10.1073/pnas.1708689114.

[99] B. Altaner and J. Vollmer. "Fluctuation-Preserving Coarse Graining for Biochemical Systems". In: *Phys. Rev. Lett.* **108**, 228101 (2012). DOI: 10.1103/PhysRevLett.108.228101.

[100] A. Celani, S. Bo, R. Eichhorn, and E. Aurell. "Anomalous Thermodynamics at the Microscale". In: *Phys. Rev. Lett.* **109**, 260603 (2012). DOI: 10.1103/PhysRevLett.109.260603.

[101] G. Bulnes Cuetara, M. Esposito, and P. Gaspard. "Fluctuation theorems for capacitively coupled electronic currents". In: *Phys. Rev. B* **84**, 165114 (2011). DOI: 10.1103/PhysRevB.84.165114.

[102] A. Wachtel, R. Rao, and M. Esposito. "Thermodynamically consistent coarse graining of biocatalysts beyond Michaelis–Menten". In: *New Journal of Physics* **20**, 042002 (2018). DOI: 10.1088/1367-2630/aab5c9.

[103] J. Mehl, B. Lander, C. Bechinger, V. Blickle, and U. Seifert. "Role of Hidden Slow Degrees of Freedom in the Fluctuation Theorem". In: *Phys. Rev. Lett.* **108**, 220601 (2012). DOI: 10.1103/PhysRevLett.108.220601.

[104] K. Sekimoto. "Microscopic heat from the energetics of stochastic phenomena". In: *Phys. Rev. E* **76**, 060103 (2007). DOI: 10.1103/PhysRevE.76.060103.

[105] A. Gomez-Marin, J. M. R. Parrondo, and C. Van den Broeck. "Lower bounds on dissipation upon coarse graining". In: *Phys. Rev. E* **78**, 011107 (2008). DOI: 10.1103/PhysRevE.78.011107.

[106] T. Leonard, B. Lander, U. Seifert, and T. Speck. "Stochastic thermodynamics of fluctuating density fields: Non-equilibrium free energy differences under coarse-graining". In: *The Journal of Chemical Physics* **139**, 204109 (2013). DOI: 10.1063/1.4833136.

[107] G. Bisker, M. Polettini, T. R. Gingrich, and J. M. Horowitz. "Hierarchical bounds on entropy production inferred from partial information". In: *Journal of Statistical Mechanics: Theory and Experiment* **2017**, 093210 (2017). DOI: 10.1088/1742-5468/aa8c0d.

[108] F. Knoch and T. Speck. "Cycle representatives for the coarse-graining of systems driven into a non-equilibrium steady state". In: *New Journal of Physics* **17**, 115004 (2015). DOI: 10.1088/1367-2630/17/11/115004.

[109] I. Martinez, G. Bisker, J. Horowitz, and J. Parrondo. "Inferring broken detailed balance in the absence of observable currents". In: *Nature Communications* **10**, 3542 (2019). DOI: 10.1038/s41467-019-11051-w.

[110] M. Kahlen and J. Ehrich. "Hidden slow degrees of freedom and fluctuation theorems: an analytically solvable model". In: *Journal of Statistical Mechanics: Theory and Experiment* **2018**, 063204 (2018). DOI: 10.1088/1742-5468/aac2fd.

[111] M. Uhl, P. Pietzonka, and U. Seifert. "Fluctuations of apparent entropy production in networks with hidden slow degrees of freedom". In: *Journal of Statistical Mechanics: Theory and Experiment* **2018**, 023203 (2018). DOI: 10.1088/1742-5468/aaa78b.

[112] M. Esposito. "Stochastic thermodynamics under coarse graining". In: *Phys. Rev. E* **85**, 041125 (2012). DOI: 10.1103/PhysRevE.85.041125.

[113] S. Bo and A. Celani. "Entropy Production in Stochastic Systems with Fast and Slow Time-Scales". In: *Journal of Statistical Physics* **154**, 1325–1351 (2014). DOI: 10.1007/s10955-014-0922-1.

[114] J. M. Horowitz and M. Esposito. "Thermodynamics with Continuous Information Flow". In: *Phys. Rev. X* **4**, 031015 (2014). DOI: 10.1103/PhysRevX.4.031015.

[115] J. M. Horowitz and H. Sandberg. "Second-law-like inequalities with information and their interpretations". In: *New Journal of Physics* **16**, 125007 (2014). DOI: 10.1088/1367-2630/16/12/125007.

[116] J. M. Horowitz. "Multipartite information flow for multiple Maxwell demons". In: *Journal of Statistical Mechanics: Theory and Experiment* **2015**, P03006 (2015). DOI: 10.1088/1742-5468/2015/03/p03006.

[117] D. Hartich, A. C. Barato, and U. Seifert. "Stochastic thermodynamics of bipartite systems: transfer entropy inequalities and a Maxwell's demon interpretation". In: *Journal of Statistical Mechanics: Theory and Experiment* **2014**, P02016 (2014). DOI: 10.1088/1742-5468/2014/02/p02016.

[118] U. Seifert. "First and Second Law of Thermodynamics at Strong Coupling". In: *Phys. Rev. Lett.* **116**, 020601 (2016). DOI: 10.1103/PhysRevLett.116.020601.

[119] C. Jarzynski. "Stochastic and Macroscopic Thermodynamics of Strongly Coupled Systems". In: *Phys. Rev. X* **7**, 011008 (2017). DOI: 10.1103/PhysRevX.7.011008.

[120] P. Strasberg and M. Esposito. "Stochastic thermodynamics in the strong coupling regime: An unambiguous approach based on coarse graining". In: *Phys. Rev. E* **95**, 062101 (2017). DOI: 10.1103/PhysRevE.95.062101.

[121] C. W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. 3rd ed. Springer-Verlag, 2004.

[122] H. Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Berlin, Germany: Springer, 1989. DOI: 10.1007/978-3-642-61544-3.

[123] H.-P. Breuer and F. Petruccione. *The Theory of Open Quantum Systems*. New York, USA: Oxford University Press, 2006. DOI: 10.1093/acprof:oso/9780199213900.001.0001.

[124] K. Lindenberg and B. J. West. *The Nonequilibrium Statistical Mechanics of Open and Closed Systems*. John Wiley & Sons, 1990.

[125] F. L. Curzon and B. Ahlborn. "Efficiency of a Carnot engine at maximum power output". In: *American Journal of Physics* **43**, 22–24 (1975). DOI: 10.1119/1.10023.

[126] K. Sekimoto. "Langevin Equation and Thermodynamics". In: *Progress of Theoretical Physics Supplement* **130**, 17–27 (1998). DOI: 10.1143/PTPS.130.17.

[127] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. 2nd ed. Wiley-Interscience, 2006.

[128] M. Esposito, U. Harbola, and S. Mukamel. "Entropy fluctuation theorems in driven open systems: Application to electron counting statistics". In: *Phys. Rev. E* **76**, 031132 (2007). DOI: 10.1103/PhysRevE.76.031132.

[129] C. Jarzynski. "Nonequilibrium Equality for Free Energy Differences". In: *Phys. Rev. Lett.* **78**, 2690–2693 (1997). DOI: 10.1103/PhysRevLett.78.2690.

[130] C. Jarzynski. "Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach". In: *Phys. Rev. E* **56**, 5018–5035 (1997). DOI: 10.1103/PhysRevE.56.5018.

[131] M. Esposito and C. Van den Broeck. "Three Detailed Fluctuation Theorems". In: *Phys. Rev. Lett.* **104**, 090601 (2010). DOI: 10.1103/PhysRevLett.104.090601.

[132] G. Bulnes Cuetara, M. Esposito, and A. Imparato. "Exact fluctuation theorem without ensemble quantities". In: *Phys. Rev. E* **89**, 052119 (2014). DOI: 10.1103/PhysRevE.89.052119.

[133] R. Rao and M. Esposito. "Conservation laws shape dissipation". In: *New Journal of Physics* **20**, 023007 (2018). DOI: 10.1088/1367-2630/aaa15f.

[134] M. Esposito, U. Harbola, and S. Mukamel. "Nonequilibrium fluctuations, fluctuation theorems, and counting statistics in quantum systems". In: *Rev. Mod. Phys.* **81**, 1665–1702 (2009). DOI: 10.1103/RevModPhys.81.1665.

[135] G. E. Crooks. "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences". In: *Phys. Rev. E* **60**, 2721–2726 (1999). DOI: 10.1103/PhysRevE.60.2721.

[136] G. E. Crooks. "Path-ensemble averages in systems driven far from equilibrium". In: *Phys. Rev. E* **61**, 2361–2366 (2000). DOI: 10.1103/PhysRevE.61.2361.

[137] O. Campàs, Y. Kafri, K. B. Zeldovich, J. Casademunt, and J.-F. Joanny. "Collective Dynamics of Interacting Molecular Motors". In: *Phys. Rev. Lett.* **97**, 038101 (2006). DOI: 10.1103/PhysRevLett.97.038101.

[138] T. L. Hill. "Theoretical study of the effect of enzyme-enzyme interactions on steady-state enzyme kinetics". In: *PNAS* **74**, 3632–3636 (1977).

[139] T. L. Hill and Y. D. Chen. "Three-state, steady-state Ising systems: Monte Carlo and Bragg-Williams treatments". In: *PNAS* **78**, 4–8 (1981).

[140] F. M. Raymo, S. Giordani, A. J. P. White, and D. J. Williams. "Digital Processing with a Three-State Molecular Switch". In: *J. Org. Chem.* **68**, 4158–4169 (2003). DOI: 10.1021/jo0340455.

[141] C. Sim ão, M. Mas-Torrent, J. Casado-Montenegro, F. Ot ón, J. Veciana, and C. Rovira. "A Three-State Surface-Confined Molecular Switch with Multiple Channel Outputs". In: *J. Am. Chem. Soc.* **133**, 13256–13259 (2011). DOI: 10.1021/ja204898u.

[142] H. J. Richter. "Recent advances in the recording physics of thin-film media". In: *Journal of Physics D: Applied Physics* **32**, R147 (1999). DOI: 10.1088/0022-3727/32/21/201.

[143] A. Dholakia. *Introduction to Convolutional Codes with Applications. Chapter: Majority-Logic Decoding*. Boston, MA, USA: Springer, 1991, 127–136. DOI: 10.1007/978-1-4615-2712-1_7.

[144] J. P. Keener. "Infinite period bifurcation and global bifurcation branches". In: *SIAM Journal on Applied Mathematics* **41**, 127–144 (1981). DOI: 10.1137/0141010.

[145] W. Arnoldi. "The principle of minimized iterations in the solution of the matrix eigenvalue problem". In: *Quart. Appl. Math.* **9**, 17–29 (1951). DOI: 10.1090/qam/42792.

[146] D. T. Gillespie. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". In: *Journal of Computational Physics* **22**, 403–434 (1976). DOI: 10.1016/0021-9991(76)90041-3.

[147] D. T. Gillespie. "Exact stochastic simulation of coupled chemical reactions". In: *The Journal of Physical Chemistry* **81**, 2340–2361 (1977). DOI: 10.1021/j100540a008.

[148] S. Varrette, P. Bouvry, H. Cartiaux, and F. Georgatos. "Management of an Academic HPC Cluster: The UL Experience". In: *Proc. of the 2014 Intl. Conf. on High Performance Computing & Simulation (HPCS 2014)*. Bologna, Italy: IEEE, 2014, pp. 959–967.

[149] T. Herpich, J. Thingna, and M. Esposito. *Time Evolution of Joint Probability Distribution*. 2018. URL: https://figshare.com/articles/Time_Evolution_of_Joint_Probability_Distribution/5822097 (visited on 06/26/2019).

[150] J. Schnakenberg. "Network theory of microscopic and macroscopic behavior of master equation systems". In: *Rev. Mod. Phys.* **48**, 571–585 (1976). DOI: 10.1103/RevModPhys.48.571.

[151] R. D. Vale and R. A. Milligan. "The Way Things Move: Looking Under the Hood of Molecular Motor Proteins". In: *Science* **288**, 88–95 (2000). DOI: 10.1126/science.288.5463.88.

[152] J. Ashkin and E. Teller. "Statistics of Two-Dimensional Lattices with Four Components". In: *Phys. Rev.* **64**, 178–184 (1943). DOI: 10.1103/PhysRev.64.178.

[153] E. Ising. "Beitrag zur Theorie des Ferromagnetismus". In: *Zeitschrift fur Physik* **31**, 253–258 (1925). DOI: 10.1007/BF02980577.

[154] L. Onsager. "Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition". In: *Physical Review* **65**, 117–149 (1944). DOI: 10.1103/PhysRev.65.117.

[155] A. Böttcher, B. Silbermann, and A. Y. Karlovich. *Analysis of Toeplitz Operators (Springer Monographs in Mathematics)*. Berlin, Heidelberg: Springer-Verlag, 2006. DOI: 10.1007/3-540-32436-4.

[156] P. Davis. *Circulant Matrices*. New York: Wiley, 1970.

[157] M. Esposito and C. Van den Broeck. "Second law and Landauer principle far from equilibrium". In: *EPL (Europhysics Letters)* **95**, 40004 (2011). DOI: 10.1209/0295-5075/95/40004.

[158] T. Sagawa. "Thermodynamic and logical reversibilities revisited". In: *Journal of Statistical Mechanics: Theory and Experiment* **2014**, P03025 (2014). DOI: 10.1088/1742-5468/2014/03/P03025.

[159] M. Gavrilov and J. Bechhoefer. "Erasure without Work in an Asymmetric Double-Well Potential". In: *Phys. Rev. Lett.* **117**, 200601 (2016). DOI: 10.1103/PhysRevLett.117.200601.

[160] Y. Jun, M. Gavrilov, and J. Bechhoefer. "High-Precision Test of Landauer's Principle in a Feedback Trap". In: *Phys. Rev. Lett.* **113**, 190601 (2014). DOI: 10.1103/PhysRevLett.113.190601.

[161] E. Aurell, K. Gawedzki, C. Mejia-Monasterio, R. Mohayaee, and P. Muratore-Ginanneschi. "Foundations of Statistical Mechanics: A Deductive Treatment". In: *J. Stat. Phys.* **147**, 487–505 (2012). DOI: 10.1007/s10955-012-0478-x.

[162] T. Schmiedl and U. Seifert. "Efficiency at maximum power: An analytically solvable model for stochastic heat engines". In: *EPL (Europhysics Letters)* **81**, 20003 (2008). DOI: 10.1209/0295-5075/81/20003.

[163] H. Touchette. "The large deviation approach to statistical mechanics". In: *Physics Reports* **478**, 1–69 (2009). DOI: https://doi.org/10.1016/j.physrep.2009.05.002.

[164] P. C. Martin, E. D. Siggia, and H. A. Rose. "Statistical Dynamics of Classical Systems". In: *Phys. Rev. A* **8**, 423–437 (1973). DOI: 10.1103/PhysRevA.8.423.

[165] M. F. Weber and E. Frey. "Master equations and the theory of stochastic path integrals". In: *Reports on Progress in Physics* **80**, 046601 (2017). DOI: 10.1088/1361-6633/aa5ae2.

[166] H. S. Leff and A. F. Rex. *Maxwell's Demon: Entropy, classical and quantum information, computing*. Princeton University Press, 1990.

[167] C. Jarzynski. "Comparison of far-from-equilibrium work relations". In: *C. R. Physique* **8**, 495–506 (2007). DOI: https://doi.org/10.1016/j.crhy.2007.04.010.

[168] K. H. Kim and H. Qian. "Entropy Production of Brownian Macromolecules with Inertia". In: *Phys. Rev. Lett.* **93**, 120602 (2004). DOI: 10.1103/PhysRevLett.93.120602.

[169] K. H. Kim and H. Qian. "Fluctuation theorems for a molecular refrigerator". In: *Phys. Rev. E* **75**, 022102 (2007). DOI: 10.1103/PhysRevE.75.022102.

[170] P. Romanczuk, M. Bär, W. Ebeling, B. Lindner, and L. Schimansky-Geier. "Active Brownian particles". In: *Eur. Phys. J. Special Topics* **202**, 1–162 (2012). DOI: 10.1140/epjst/e2012-01529-y.

[171] T. Schreiber. "Measuring Information Transfer". In: *Phys. Rev. Lett.* **85**, 461–464 (2000). DOI: 10.1103/PhysRevLett.85.461.

[172] A. E. Allahverdyan, D. Janzing, and G. Mahler. "Thermodynamic efficiency of information and heat flow". In: *Journal of Statistical Mechanics: Theory and Experiment* **2009**, P09011 (2009). DOI: 10.1088/1742-5468/2009/09/p09011.

[173] X. S. Liang and R. Kleeman. "Information Transfer between Dynamical System Components". In: *Phys. Rev. Lett.* **95**, 244101 (2005). DOI: 10.1103/PhysRevLett.95.244101.

[174] A. J. Majda and J. Harlim. "Information flow between subspaces of complex dynamical systems". In: *Proceedings of the National Academy of Sciences* **104**, 9558–9563 (2007). DOI: 10.1073/pnas.0703499104.

[175] A. C. Barato, D. Hartich, and U. Seifert. "Information-theoretic versus thermodynamic entropy production in autonomous sensory networks". In: *Phys. Rev. E* **87**, 042104 (2013). DOI: 10.1103/PhysRevE.87.042104.

[176] S. Goldt and U. Seifert. "Stochastic Thermodynamics of Learning". In: *Phys. Rev. Lett.* **118**, 010601 (2017). DOI: 10.1103/PhysRevLett.118.010601.

[177] G. Diana and M. Esposito. "Mutual entropy production in bipartite systems". In: *Journal of Statistical Mechanics: Theory and Experiment* **2014**, P04010 (2014). DOI: 10.1088/1742-5468/2014/04/p04010.

[178] K. Ptaszyński and M. Esposito. "Thermodynamics of Quantum Information Flows". In: *Phys. Rev. Lett.* **122**, 150603 (2019). DOI: 10.1103/PhysRevLett.122.150603.

[179] J. G. Kirkwood. "Statistical Mechanics of Fluid Mixtures". In: *The Journal of Chemical Physics* **3**, 300–313 (1935). DOI: 10.1063/1.1749657.

[180] H. Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Berlin, Germany: Springer, 1989. DOI: 10.1007/978-3-642-61544-3.

[181] M. Suñé and A. Imparato. "Out-of-Equilibrium Clock Model at the Verge of Criticality". In: *Phys. Rev. Lett.* **123**, 070601 (2019). DOI: 10.1103/PhysRevLett.123.070601.

[182] M. Polettini and M. Esposito. "Effective fluctuation and response theory". In: *Journal of Statistical Physics* **176**, 94–168 (2019). DOI: 10.1007/s10955-019-02291-7.

[183] Y. A. Kuznetsov. *Elements of applied bifurcation theory*. 2nd ed. Springer-Verlag, 1998. DOI: 10.1007/978-1-4757-3978-7.

[184] D. T. Gillespie. "Stochastic Simulation of Chemical Kinetics". In: *Annual Review of Physical Chemistry* **58**, 35–55 (2007). DOI: 10.1146/annurev.physchem.58.032806.104637.

[185] T. Cossetto and G. Falasco. *Private Notes*.

[186] C. Aron, G. Biroli, and L. F. Cugliandolo. "Symmetries of generating functionals of Langevin processes with colored multiplicative noise". In: *Journal of Statistical Mechanics: Theory and Experiment* **2010**, P11018 (2010). DOI: 10.1088/1742-5468/2010/11/p11018.

Author's Contribution

The authorship of the five included references is not entirely due to myself. The first article [1] on the three-state model is the result of a very intense collaboration with Juzar Thingna. Starting from a well-defined setup and an *ad-hoc* mean-field equation proposed by Prof. Esposito, I realized that one can in fact start from a microscopic representation and arrive at the mean-field equation in the macroscopic limit. The numerical advice of Juzar Thingna was crucially important as he made me aware of how nontrivial the numerical problem is and kept me motivated during the tedious and at times frustrating process of coding. He furthermore introduced me to high-performance computing using the HPC facilities of the University of Luxembourg [148]. The calculations and the implementation of the numerics were mainly done by myself and I wrote the paper in close exchange with Prof. Esposito.

The underlying idea and implementation of the second article [2] was due to me. I realized that, using simple thermodynamic arguments, the low- and high-temperature limit phases of the three-state model are universal for all clock models with infinite-range interactions and numerically discovered the existence of two classes which exhibit universal (thermo)dynamical phenomenology. The discussions with Prof. Esposito inspired me to also derive the critical Hopf-bifurcation temperature for all clock models. The thermodynamic analysis was entirely conducted by myself. I wrote the paper myself with instructions from Prof. Esposito.

The manuscript which culminated into article [3] was already well-established regarding its scientific results by Shiqi Sheng. While I added minor results on the physics, my main contribution was to rewrite the entire draft in order to create a coherent narrative and to significantly improve the presentation of the results.

Similarly, the draft for the article [4] was already in an advanced stage due to the work of Kamran Shayanfard. Yet, the formulation of the coarse-graining contained inconsistencies, which I, under guidance by Prof. Esposito, could resolve. Besides, Prof. Esposito inspired me to also make contact with the theory of the Hamiltonian of mean force. Furthermore, I overall significantly improved the presentation of the results that were already established and shaped the draft into a paper.

Inspired by the works [1] and [2], I realized that it is possible to consistently formulate the stochastic thermodynamics across microscopic, mesoscopic scales as well as on the mean-field level for all-to-all interacting identical systems. The different thermodynamic descriptions for the averages at these levels had already been established for the three-state model and could be straighforwardly generalized to a much more generic setup. I

furthermore formulated the fluctuating thermodynamics at the microscopic and mesoscopic scale and identified the conditions under which these representations coincide. Regarding the fluctuating thermodynamics at macroscopic scales, Prof. Esposito realized the connection with the work of Gianmaria Falasco and Tommaso Cossetto on path integral representations of the nonequilibrium thermodynamics for large many-body systems. Together we succeeded in consistently formulating the fluctuating thermodynamics at macroscopic scales using path integral techniques. I wrote the entire paper [5] under guidance of Prof. Esposito.

Acknowledgments

In the following lines, I want to express my deepest gratitude to everyone that in some way contributed to this thesis: First of all, I am deeply indebted to my supervisor Prof. Esposito for giving me the possibility to work on this project in his internationally renowned research group. Working in such an ambitious environment definitely allowed me to progress as a researcher and, more importantly, as a human being. Although being very busy at times, Prof. Esposito still managed to find some time to discuss pertinent research questions. I also wish to thank Prof. Tkatchenko and Prof. Schmidt for their efforts as members of the thesis committee and their critical feedback during the obligatory thesis progress meetings as well as Prof. Imparato and Prof. Garrahan for their time spent as external referees for this thesis. Next, I am thankful to Juzar Thingna for his advice on numerical aspects and our innumerable discussions which resulted into a very fruitful collaboration. In that regard, I also would like to thank everyone else that worked together with me on joint projects and therefore contributed to this dissertation. Moreover, I am also grateful to Grégory Bulnes Cuetara for becoming a friend and sharing some memorable moments during our trips and mutual time in Luxembourg. There are many other persons, in fact too many as that they could be named individually, inside and outside the University of Luxembourg that enriched my time as a doctoral student in Luxembourg and whom I wish to thank. I would also like to acknowledge that this thesis was funded by an AFR PhD Grant awarded in 2016 by the National Research Fund Luxembourg (grant number 11271777). Finally, I am thankful for the love and support of my family, friends and Zahra, especially during the occasionally stressful times as a doctoral student.