Henrik Wilming

A Quantum of Thermodynamics

From ground state cooling to spontaneous symmetry breaking
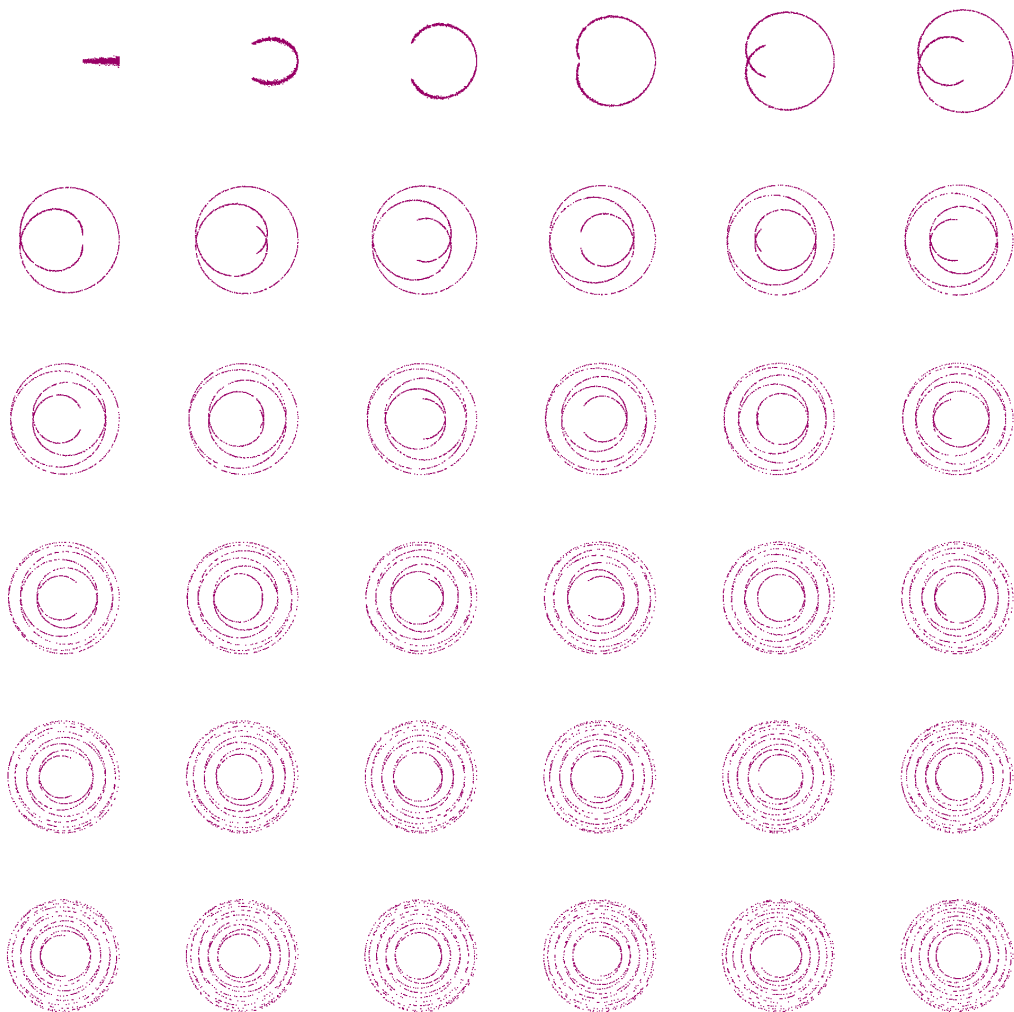
Erstgutachter: Prof. Dr. Jens Eisert

Zweitgutachter: Prof. Felix von Oppen, PhD

Datum der Disputation: 19.7.2018

Publications of the author of this thesis

Peer-reviewed publications:

[1]   H. Wilming, R. Gallego, and J. Eisert. "Second law of thermodynamics under control restrictions". *Phys. Rev. E* 93.4 (2016). DOI: 10.1103/physreve.93.042126.

[2]   R. Gallego, J. Eisert, and H. Wilming. "Thermodynamic work from operational principles". *New J. Phys.* 18.10 (2016), p. 103017. DOI: 10.1088/1367-2630/18/10/103017.

[3]   M. Perarnau-Llobet, A. Riera, R. Gallego, H. Wilming, and J. Eisert. "Work and entropy production in generalised Gibbs ensembles". *New J. Phys.* 18.12 (2016), p. 123035. DOI: 10.1088/1367-2630/aa4fa6.

[4]   H. Wilming, M. J. Kastoryano, A. H. Werner, and J. Eisert. "Emergence of spontaneous symmetry breaking in dissipative lattice systems". *J. Math. Phys.* 58.3 (2017), p. 033302. DOI: 10.1063/1.4978328.

[5]   F. Pastawski, J. Eisert, and H. Wilming. "Towards Holography via Quantum Source-Channel Codes". *Phys. Rev. Lett.* 119.2 (2017), p. 020501. DOI: 10.1103/PhysRevLett.119.020501.

[6]   J Lekscha, H Wilming, J Eisert, and R Gallego. "Quantum thermodynamics with local control". *Phys. Rev. E* 97.2 (2018), p. 022142. DOI: 10.1103/PhysRevE.97.022142. arXiv: 10.1103/PhysRevD.97.086015.

[7]   H. Wilming, R. Gallego, and J. Eisert. "Axiomatic Characterization of the Quantum Relative Entropy and Free Energy". *Entropy* 19.6 (2017), p. 241. DOI: 10.3390/e19060241.

[8]   H. Wilming and R. Gallego. "Third Law of Thermodynamics as a Single Inequality". *Phys. Rev. X* 7 (4 2017), p. 041033. DOI: 10.1103/PhysRevX.7.041033.

[9]   M. Perarnau-Llobet, H. Wilming, A. Riera, R. Gallego, and J. Eisert. "Strong Coupling Corrections in Quantum Thermodynamics". *Phys. Rev. Lett.* 120.12 (2018). DOI: 10.1103/physrevlett.120.120602. arXiv: 1704.05864.

[11] P. Boes, H. Wilming, J. Eisert, and R. Gallego. "Statistical ensembles without typicality". *Nature Comm.* 9.1 (2018), p. 1022. DOI: 10.1038/s41467-018-03230-y. arXiv: 1707.08218.

[12] R. Gallego, H. Wilming, J. Eisert, and C. Gogolin. "What it takes to avoid equilibration". *Phys. Rev. A* 98.2 (2018). DOI: 10.1103/physreva.98.022135. arXiv: 1711.09832.

Pre-prints:

[10] H. Wilming, M. Goihl, C. Krumnow, and J. Eisert. "Towards local equilibration in closed interacting quantum many-body systems" (2017). arXiv: 1704.06291 [quant-ph].

Contents

1 *Introduction and Overview*

THERMODYNAMICS is a tremendously useful theory. It can be used to explain the functioning of steam engines, refrigerators, power plants and cars. But at the same time it also explains the very basic mechanism behind climate-change (together with some basic properties of the interaction of certain molecules with light), chemical reactions, or why it helps to put ice-cubes in water if we want to have a cool drink. But it does not stop there: In fact, today it is also believed that even properties of black holes can and indeed have to be understood thermodynamically [13–15] to get a consistent physical description of our universe. Indeed, this is by now one of the guiding principles for obtaining a viable theory of quantum gravity and was essential for the currently much-discussed "Holographic" approach to quantum gravity [16–20].

The success and universality of thermodynamics is surprising given the fact that it was invented in the 19th century – long before quantum mechanics and statistical physics were developed. This was possible since the concepts and principles of thermodynamics can in fact be developed without any reference to the underlying microscopic physical mechanisms that give rise to them: It is essentially a formal framework based on very general axioms [21–24]. Moreover, these axioms or principles, often called the *laws of thermodynamics*, can be stated in an operational way. This means that they state that certain operations or processes on physical objects are impossible to perform (creating energy out of nothing, lifting arbitrary weights only using a single heat bath, cooling a system to zero temperature with finite resources). All predictions of phenomenological thermodynamics can be derived from these basic principles together with experimentally measurable quantities describing matter at the level of thermodynamic quantities (the equations of state).

Of course, today we know that it is quantum mechanics and not thermodynamics that describes the material world around us on the fundamental level. Once we accept that quantum mechanics underlies our physical reality at the microscopic level, we must then also accept that the thermodynamic principles have to follow from quantum mechanics, even though they can be formulated without referring to it. From a theoretical point of view it is then most desirable to understand exactly *how* thermodynamics emerges from quantum mechanics. In particular, the following questions have to be answered:

1. What exactly is (thermal) equilibrium and how does it emerge?

2. Can we understand the second law of thermodynamics from a microscopic point of view? In particular, how do the concepts "work" and "heat" emerge?

3. Can we understand the third law from a microscopic point of view?

4. Do thermodynamic principles apply to individual quantum systems?

Unsurprisingly, these questions have been studied and debated intensively over the 20th and 21st century, providing many interesting insights about physics, particularly the interconnections between thermodynamics, statistical physics and quantum mechanics [1]. Nevertheless, it seems fair to say that the above questions have not been answered conclusively in general, in the sense that a unique line of reasoning could be found that explains the emergence of all thermodynamic principles. However, in principle there is also no logical ground to assume that such a unique explanation would exist. Rather, many different explanations can be given, which should be considered complimentary as long as they do

---

[1] It is hopeless to reference all the relevant literature in this vast field; a small sample of seminal contributions can be found in Refs. [13, 16, 25–37] together with the reviews [38–42] and references therein.

not contradict each other. This is, to some extend, analogous to the emergence of classical mechanics from quantum mechanics or the use of low-energy effective field theories [43]: In general, there are many different microscopic quantum systems which give rise to the same behaviour in a suitable limit. As long as one is only interested in this limit, all the different microscopic descriptions are equivalent.

Different approaches to deriving statements akin to the thermodynamic laws from more microscopic principles will in general yield different results, applicable to different settings. For example, when studying the thermodynamics of the processes inside biological cells, the (classical) statistical mechanical approach to studying the kinetics and dynamics of biomolecules is very useful. In this formulation, the second law is encoded in so-called "fluctuation relations" [37, 44, 45] and this branch of research has been enormously successful over the last years, in particular in also explaining non-equilibrium aspects of thermodynamics. At the same time it operates essentially in a semi-classical description, which does not need to make reference to the underlying quantum mechanical description of matter.

On the other hand, if we study what can be done in experiments on individual, small *quantum* systems, which became doable due to astonishing experimental progress over the last years, this approach might not be the most useful one. Over the last years, however, a very fruitful approach to (quantum) thermodynamics has emerged, which shows that, at least in principle, thermodynamic laws also exist for individual quantum systems which can be brought in contact with thermal systems (in the sense of quantum statistical mechanics), which can be large but do not have to be large. Indeed, this new approach, known under the name of "(quantum) resource theory of thermodynamics" (see Refs. [46–49] and chapter 2), directly operates on the level of usual unitary quantum mechanics and makes only three fundamental assumptions: i) the validity of energy-conservation on the level of probability distributions, ii) the existence and availability of thermal heat baths in the sense of (quantum) statistical mechanics, i.e., Gibbs-states, and iii) that we can assume that heat baths are uncorrelated to other systems of interest. This approach has recently provided strong bounds on work-extraction in so-called single-shot settings [47, 49–53], has elucidated the role of coherences in quantum thermodynamics [54–58], provided new quantum fluctuation relations [59, 60], and provided a quantitative formulation of the unattainability principle [61].

IN THE RESEARCH PRESENTED IN THIS THESIS[2] we employ (variants of) this new approach to thermodynamics to provide new perspectives on questions 1.–4. To start, in chapter 2 I first review the resource theoretic approach to the thermodynamics of individual quantum systems. Then I explain how basic (quantum) information theoretic tools can be used to derive a set of second laws of thermodynamics without any additional assumptions. I also illustrate this approach with several applications, re-deriving well-known results in statistical mechanics.

In the following chapter 3, I show how to use the machinery developed in chapter 2 to derive an operational version of the third law of thermodynamics. More precisely, I quantify the non-equilibrium resources that are both necessary and sufficient to cool a system close to absolute zero temperature and quantify how these resources diverge as the final temperature approaches absolute zero. This provides a quantification of the third law of thermodynamics from basic principles, however from a perspective that is complementary to Ref. [61]. To arrive at this result, I introduce a novel state function on non-equilibrium states, called *vacancy*. The vacancy is formally closely related to the non-equilibrium free energy, but its physical interpretation is very different. The main result of this chapter essentially shows that the vacancy controls thermodynamic behaviour at very small temperatures.

If thermodynamic principles apply to small quantum systems, then it is interesting to study whether phenomenological concepts such as "work" and "heat" can also be transferred in a meaningful way to this setting. This is interesting since energy is in general a fluctuating quantity for small systems: first due to thermal fluctuations, but ultimately due to quantum fluctuations, since we cannot expect that a small quantum system resides in an energy-eigenstate. Imagine, for example, a small quantum system undergoing a state transition form an energy-eigenstate to a super-position of two energy-eigenstates in the

[2] The content of this thesis is based on the publications [1–4, 6–12].

presence of a heat bath. This is a basic elementary operation in a quantum computer [62]. Can we meaningfully associate a certain amount of "work" or "heat" to such a transition?

This broad problem is the topic of chapter 4, where we discuss how the performance of quantum machines can be quantified in the presence of thermal baths from an abstract point of view. The results in this chapter are derived by introducing an axiomatic, operational framework and putting special emphasis on the connection between a consistent definition of "thermodynamic work", the second law of thermodynamics and correlations between quantum systems. As a side-result we obtain a new axiomatic characterization of the quantum relative entropy and non-equilibrium free energy related to the build-up of correlations in thermodynamic processes. This characterization is detailed in the following chapter 5.

The results in chapters 3–5 assume the availability of heat baths well described by Gibbs states, i.e., states of the form

$$\omega_\beta(H) := \frac{\mathrm{e}^{-\beta H}}{Z_\beta(H)}, \quad Z_\beta(H) := \mathrm{Tr}(\mathrm{e}^{-\beta H}). \tag{1.1}$$

This description for states in thermal equilibrium can be justified from many different perspectives. The most common ones use some version of a typicality argument, i.e., the statement that most microstates, according to some reasonable measure, give the same measurement statistics for physically relevant observables. Alternatively, one can make use of an ergodic hypothesis together with an argument about the equivalence of ensembles [38]. Quite distinct from these approaches is that of Jaynes' Maximum Entropy Principle. It states that one should always assign that probability distribution (density matrix) to a (quantum) system which is compatible with all the information that one has about its state and otherwise maximizes the Shannon entropy (von Neumann entropy in the quantum case) [31, 32]. Apart from the reasoning based on ergodicity, which, however, seems difficult to translate to the quantum setting since there is no clear notion of ergodicity and non-integrability [63], on the conceptual side these approaches suffer from introducing either some probability measure or an information theoretic entropy measure in an ad-hoc way.

In chapter 6, I show a new way of deriving the canonical ensemble, which does not rely on any probability measure and entropy measure. Instead, it follows from an operational formulation of thermodynamics under partial information, analogous to the resource theoretical approach to thermodynamics. This approach works by assuming that one only knows the expectation value of certain observables about any system and asks what kind of state-transformations are possible to implement on such a system. It thus does not operate on quantum states, but on *macrostates*. Nevertheless, it does not assume that the system on which one acts is *macroscopic* and is thus also applicable to single quantum systems under a setting of partial information.

Although the functional form that quantum states in thermal equilibrium take is derived and not assumed in chapter 6, the results in chapters 2 to 6 assume the existence of thermal equilibrium in some form from the beginning. How such thermal equilibrium is attained dynamically is completely left open, however. This is discussed in chapters 7 and 8. First, I discuss general results about equilibration of complex many body systems. Then I provide intuition on how such equilibration behaviour can be understood more "mechanically" and why we can expect that equilibration in fact happens quickly, even though general bounds on equilibration times diverge with the system size. Following up, in chapter 8, I present arguments that quantitatively show that it is difficult to prepare a quantum system in a state that does *not* equilibrate. These arguments provide a novel link between the resource theoretic approach to thermodynamics and the problem of equilibration in complex quantum systems. We also discuss in detail the role of correlations and their connection with the existence of a "second law of equilibration". This closes the first part of the thesis that deals with fundamental questions in thermodynamics from an abstract point of view.

From an operational point of view, the success of phenomenological thermodynamics is partly based on the fact that macroscopic systems only need to be controlled on a macroscopic level to be able to saturate the thermodynamic bounds arbitrarily well. For example, the movement of the piston in a heat engine does not need to depend on the microscopic motion of each of the molecules in the gas to achieve high thermodynamic efficiency. It is thus interesting whether this feature persists on the level of thermodynamics of quantum

systems.

In the second part of the thesis, in chapters 9 – 11, I discuss the thermodynamic bounds for thermal machines whose working system is made up of individual quantum systems. In particular, I focus on how such thermodynamic bounds are modified when experimental restrictions are introduced. Examples of such restrictions are limited control over the coupling to the heat bath, limitations in terms of available field strengths or the inability to control the fundamental interactions between particles. To discuss these questions, a more hands-on model of thermodynamics is required, which is introduced first in chapter 9. This chapter also discusses the above mentioned control restrictions. I also discuss in detail the example of a situation in which an experimenter can only bring a many-body system in contact with either a hot or a cold bath and control external magnetic fields, but not the inter-particle interactions. I derive general bounds of the efficiency of such a model as a function of the interaction strengths in the system and apply them to the example of an Ising model. I find that in the limit of strong interactions, thermodynamic properties depend strongly on whether the coupling between the individual spins is ferro-magnetic or anti-ferro-magnetic and provide explanations for this behaviour.

The discussion in chapters 7 and 8 provided an understanding of the equilibration behaviour of many-body systems, which we now use to study thermal machines from the point of view of closed quantum systems. First, in chapter 10, I develop a general framework to discuss thermodynamics of closed, equilibrating many-body systems from a general point of view. This framework uses an effective description of equilibrium states in terms of so-called *Generalized Gibbs ensembles* and we discuss in great detail the emergence of reversible transformations, entropy production and the validity of the minimum work principle.

The results in chapter 9 are derived under the assumption that the working system in a thermal machine is coupled only weakly to thermal baths. For small systems, however, this assumption is in general not justified and it is of great interest to study how thermodynamic bounds behave under strong coupling to heat baths. In the next chapter 11, we apply the framework developed in chapter 10 to discuss in detail the thermodynamics of thermal machines that are strongly-coupled to heat baths. First, we derive general and tight corrections for the fundamental work-extraction bounds and the efficiency of a thermal machine. These bounds hold for arbitrary interaction strengths and can in principle be evaluated for any model. However, in general the bounds are complicated functions of the involved Hamiltonians. Therefore, we also derive the fully general, explicit correction to leading order in the coupling strength. This correction takes a surprisingly simple form and can nicely be interpreted in terms of a "dressed interaction". All these results show that strong coupling is detrimental for thermodynamic efficiency. However, the general form of the corrections allows us to argue that a finite interaction strength is optimal in terms of the power of thermal machines, as is expected, and discuss general bounds on the power in leading orders of the coupling strength. Finally, the results are illustrated in an explicit model of quantum Brownian motion [64, 65], showing very good agreement with the theoretical predictions. This closes the discussion of thermal machines.

IN THE LAST CHAPTER of this thesis, chapter 12, I come back to a classic problem in quantum statistical mechanics, namely the emergence of spontaneous symmetry breaking in lattice models. Usually, spontaneous symmetry breaking in quantum statistical mechanics is discussed mostly in terms of kinematics by showing that thermal equilibrium states of certain systems in the thermodynamic limit are not unique and that the different equilibrium states break a symmetry of the Hamiltonian. This can, for example, be done by adding an infinitesimal small symmetry-breaking field which is removed after the thermodynamic limit has been taken. In such an approach, the actual dynamics that thermalizes the system is not modelled explicitly.

Here, I discuss the emergence of spontaneous symmetry breaking from a dynamical perspective of Markovian open system dynamics. The main result can informally be stated in the following way. Suppose that local, purely dissipative dynamics acts on a lattice system and that this dynamics is in detailed balance with some state which has finite fluctuations in the *density* of an order parameter. Furthermore suppose that this order parameter vanishes in this state, i.e., the state is symmetric. Then the main result from this chapter shows

that there *necessarily* also exist steady states in which the order parameter takes a finite value. Hence, these steady-states break the symmetry associated to the order parameter. As a simple example of the implications of this result, the existence of spontaneous symmetry breaking in the Glauber-dynamics of a two-dimensional Ising model [66] follows as a corollary by noting that the Gibbs-state of the two-dimensional Ising model has finite fluctuations in the magnetization density on any finite lattice with periodic boundary conditions below the critical temperature. However, the result also generalizes to continuous symmetries and in this case implies the existence of *dissipative Goldstone-modes*. I also derive explicit bounds on the equilibration time-scale of the symmetry-breaking states in finite systems.

Many of the results in this thesis are (or can be) proven mathematically fully rigorously. However the main ideas can usually be understood without referring to the full technical proofs. To prioritize accessibility I therefore took the liberty of omitting technical arguments or proofs in the main-text whenever I felt that they deviate too much from the important points at hand. The technical proofs missing in the main-text can be found in chapter 14 and are referenced in the main-text. Unless stated explicitly otherwise, I choose units such that $\hbar = 1$ and $k_{\mathrm{B}} = 1$ throughout the thesis.

2 *Thermodynamics as a resource theory*

TO DERIVE FUNDAMENTAL BOUNDS in quantum thermodynamics we need to find a model which allows us to formulate thermodynamic problems in a very general way, while at the same time allowing us to obtain concrete results. In this chapter I will introduce the model that we will work with in the following chapters.

After formalizing the model I will then exemplify the usefulness of the model using operational tasks such as Landauer's principle, extraction of energy from non-equilibrium systems or inducing population inversion. Importantly, the model will not postulate the existence of quantities such as entropy or free energy – these will be derived notions, using ideas from information theory, and will play major roles in the chapters 3 and 4.

To motivate the model, imagine an extremely capable experimenter called Alice living and experimenting in an environment at inverse temperature β. She has full quantum control over arbitrary quantum systems, i.e., she can implement any unitary evolution. The only constraint is energy-conservation, so that the energy is always exactly conserved on the systems that she is acting upon.

If Alice wants to implement a state-transition between two quantum states on some system and the final state does not have the same energy as the initial state, she therefore has to involve a second system and take energy from it. Since Alice is living in an environment of inverse temperature β, a particularly convenient way to obtain this energy is by simply using systems in thermal equilibrium, which are essentially provided for free to her. In case Alice can implement the state-transition on her system by only using systems from the thermal environment we will therefore say that the transition is a "free transition" – free in the sense that she has not used up any valuable "resources": Since the environment is by assumption macroscopic, while Alice acts on small systems, she could repeat the transition on another small system. Using words from classical thermodynamics, we could say that she only used "heat" from her environment to induce the state transition.

At this stage one might ask: If Alice has full quantum control over her systems and the macroscopic thermal environment, that is any single particle in the air around her, in fact all the particles making up her lab, couldn't she prepare any state that she wants to prepare? Maybe surprisingly, the answer is no, if she does not also has a large amount of systems which are not in thermal equilibrium that she can use up (bring to equilibrium) in the process.

In fact we will see that Alice is highly constrained in which state transitions she can implement. An interesting question is then to ask: How many and which kind of "resources", that is systems in certain states out of equilibrium, does she need to be able to implement processes? Which resources can be inter-converted to which other states? How does the concept of "work" come into play? And can we derive only from the above assumptions

restrictions akin to the second and third law of thermodynamics?

All these questions will be answered, at least partially, in the rest of this and the following chapters. The model that I have described in words above has been introduced at least two times in the literature [46, 47] and is now commonly called the framework of "thermal operations". As we will see, it is particularly useful to obtain fundamental limitations, while other, more simplified and less abstract models, which will we will use in chapters 9–11, are sometimes better suited to study more specific problems with additional physical constraints.

2.1   Formalizing the model

To obtain concrete answers from the model of thermal operations, we first have to formalize it into a mathematical model. To do this we have to specify what we mean by the fact that Alice has access to systems in thermal equilibrium at a given inverse temperature β and what we mean by energy-conservation.

In the model of thermal operations we formalize energy-conservation in the strict sense that the exact probability distribution of energies of the system of interest is preserved. Mathematically this is guaranteed if Alice acts on systems S and B, with Hamiltonians H_S and H_B, and the *overall* time evolution is described by a unitary operator U that commutes with the total Hamiltonian:

$$[U, H_S \otimes \mathbf{1} + \mathbf{1} \otimes H_B] = 0. \tag{2.1}$$

It is important that U describes the entire physical process that Alice implements. It does *not* mean, that this condition has to be fulfilled at every instant of time. Alice can let the two systems interact very strongly using some time-dependent interaction V_t for some time, as long as the total evolution

$$U = \mathcal{T} e^{-i \int (H_S + H_B + V_t) dt} \tag{2.2}$$

conserves the energy in the above sense. Here, $\mathcal{T}$ denotes the time-ordering operator.

The second assumption that we need to formalize is the access to systems in the thermal environment. We formalize this by assuming that Alice can include *arbitrary* systems described by some Hamiltonian H_B and the corresponding Gibbs-state

$$\omega_\beta(H) := \frac{e^{-\beta H}}{Z_\beta}, \quad Z_\beta = \mathrm{Tr}(e^{-\beta H}). \tag{2.3}$$

There are several ways to argue why we allow Alice to choose arbitrary Hamiltonians H_B. First, if we want to derive fundamental bounds, we can never exclude that Alice indeed simply has access to the one particular Hamiltonian in the environment that she needs since we do not want to make very specific assumptions about her environment. Second, we assume that Alice has access to a macroscopic thermal environment and has full quantum control. From the fact that she can implement arbitrary energy-conserving unitaries, it follows then that the only important property of H_B is its energy spectrum (with degeneracies). But because Alice has full quantum control she can in principle choose to only make her system of interest interact with some hand-picked energy-eigenstates of her macroscopic environment. Since one can find essentially any combination of energy-eigenstates with any wished-for degeneracies in a macroscopic environment, Alice can then effectively make her system interact with arbitrary Hamiltonians. This works due to the special property of Gibbs-states that the normalized probability distributions on a subset of energy-eigenstates is the Gibbs-state of the Hamiltonian made up by only taking those energy-eigenstates. Formally, if P is the projector onto a subset of energy-eigenstates, then we have:

$$\frac{\omega_\beta(H)\big|_P}{\mathrm{Tr}(P\omega_\beta(H))} = \omega_\beta(H|_P),$$

where $|_P$ means that the preceding operator is restricted to the subspace P.

Finally, we also assume that the thermal environment is uncorrelated with the system she wishes to act on. This can be justified by arguing that any correlations of the thermal environment with a specific, small system should be diluted quickly over the whole environment, so that any finite subsystem of the bath is essentially uncorrelated with the given small system. While this assumption seems quite innocent at first, it is in fact crucial to obtain a second law of thermodynamics [67] and similar to a modern form of Boltzmann's "Stosszahlenansatz" [25, 38, 39].

Combining the unitary operations with the access to uncorrelated thermal systems, we then see that the operations, or quantum channels, that Alice can implement for free are all the operations of the form

$$\rho \mapsto \mathcal{G}(\rho) = \mathrm{Tr}_B \left(U \rho \otimes \omega_\beta(H_B) U^\dagger \right). \tag{2.4}$$

In the following I will often call a transition $\rho \to \sigma = \mathcal{G}(\rho)$ of the above form a *free transition* and denote it by $\rho \rightsquigarrow \sigma$. It is an important property of the free transitions that they are *Gibbs-preserving*, meaning that when acting on a state in thermal equilibrium with the environment, it cannot be changed:

$$G(\omega_\beta(H)) = \omega_\beta(H). \tag{2.5}$$

This follows immediately from (2.4) and energy-conservation:

$$U \omega_\beta(H_S) \otimes \omega_\beta(H_B) U^\dagger = U \omega_\beta(H_S + H_B) U^\dagger = \omega_\beta(U(H_S + H_B) U^\dagger) \tag{2.6}$$

$$= \omega_\beta(H_S + H_B). \tag{2.7}$$

It is of course not true if the system starts initially in a Gibbs-state at a different temperature than the environment.

In fact the free transitions as stated so far are arguably still too restrictive: Imagine Alice would, in performing the state-transition, use a third auxiliary system, which is mapped back exactly to its initial condition at the end. Clearly, in this case she also has not used up any resource, and nobody could tell afterwards that she had used this auxiliary system as a tool. This tool then plays the role that "cyclic machines" play in phenomenological thermodynamics and we should clearly allow the use of such tools in the definition of free transitions. In the literature of quantum thermodynamics, these tools are usually called "catalysts" [49, 68], since they may allow for free transitions that are otherwise not possible, and the corresponding transitions are called *catalytic* free transitions. Formally, I will write

$$\rho \underset{c}{\rightsquigarrow} \sigma \quad \text{if} \quad \rho \otimes \chi \rightsquigarrow \sigma \otimes \chi, \tag{2.8}$$

for some auxiliary system described by the density matrix and Hamiltonian (χ, H_C). Here, we assume that the catalyst remains uncorrelated with the system of interest and does not change at all during the process. In chapters 4, 5 and 8, we discuss the consequences of allowing for a built-up of correlations between the catalyst and system of interest in different situations. Similarly, once can also consider what happens if one only requires the catalyst to be returned approximately [49, 68]. This will be discussed in some detail in chapter 3 in the context of ground state cooling.

After formulating in a mathematically precise way the model of catalytic thermal operations, let us begin to study what Alice can and cannot achieve using such operations. Before exploring the general limitations that arise, let us consider a simple example.

2.2   *Example: Anomalous heat-flow*

In this section we will discuss a simple example [1] that illustrates some of the features of the model of thermal operations. The example works as follows. We imagine that Alice has a two-level system with energy gap Δ that is initially isolated from the environment and at a thermal state much colder than the environment. For simplicity we can imagine it to be in the ground state $|0\rangle$. For some reason, however, Alice needs a very hot two-level system.

[1] This example was first developed in [1] and will be used again in Chapter 9.

We can then ask: How hot can Alice make the two-level system? By letting it interact in a weak and arbitrary way for a sufficiently long time with the environment, she could clearly achieve inverse temperature β. A simpler way to achieve this, in the set-up of thermal operations, is to simply swap the system with a two-level system (with the same energy-gap) from the environment. Similarly, she could achieve any temperature in the range 0 to $1/\beta$: One way to do this is by letting the system interact with the environment for a time that only brings the system to an intermediate temperature. More formally, she can do this by using the fact that thermal operations are *convex*. If $\mathcal{G}_1$ and $\mathcal{G}_2$ are two quantum channels representing thermal operations, then $p\mathcal{G}_1 + (1-p)\mathcal{G}_2$ is also a quantum channel representing a thermal operation. This follows from the fact that one can use a further thermal system as a source of randomness and condition a thermal operation on events of this source of randomness using a controlled unitary operation. Therefore, if one can reach a state ρ' from some initial state ρ, then it is also possible to reach the state $p\rho + (1-p)\rho'$. In the case of a two-level system, all states of the form $p\,|0\rangle\langle 0| + (1-p)\omega_\beta(H)$ are thermal states with temperatures between 0 and $1/\beta$.

On the other hand, one might think that she cannot go all the way to $T = +\infty$ or even to negative temperatures, that is, achieve a population inversion. I will now show that, depending on the energy-gap Δ and the environment temperature β, the latter is in fact possible and then explain why this is not problematic from a thermodynamic perspective.

If we disregard possible coherences, i.e., off-diagonal elements of the density matrix in the energy eigenbasis, any state of the two-level system can be parametrized by a single quantity, namely the probability p to find the system in the excited state $|1\rangle$ upon measuring the energy. The state-space of such systems is thus given by the pairs (p,Δ) (see Figure 2.1).

Figure 2.1: The state-space of a two-level system and illustration of anomalous heat transfer. The red line shows the thermal excitation probability at the environment temperature. Anomalous heat flow (AHF) allows a system to reach a state with higher excitation probability than the thermal one by interacting in an energy- and entropy-preserving way with a heat bath.
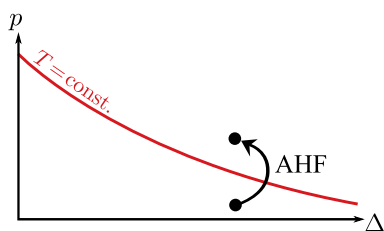
In this description, Alice's goal is to maximize the final excitation probability p' of her working system using a thermal operation. In Fig. 2.2, I describe an algorithm to achieve maximum final excitation probability. This algorithm uses n thermal two-level systems with the same energy gap Δ as the working system. When implemented, the algorithm has the following effect: If applied to an initial state with $p < p_\beta$, it maximizes p'. If applied to a state $p > p_\beta$ it minimizes it. As $p \to p_\beta$, we of course get $p' \to p_\beta$.

What is the largest excitation probability (and hence temperature) that can be achieved with this algorithm? A detailed analysis (see section 14.1.1) shows that the final probability of Alice's system to be excited is given by

$$p' = p_n^*(\beta) - p\left(\frac{p_n^*(\beta)}{p_\beta} - 1\right), \tag{2.9}$$

with the function $p_n^*(\beta)$ given by

$$p_n^*(\beta) = \sum_{m=1}^{n}\binom{2n}{m-1}\frac{e^{-\beta\Delta m}}{Z_\beta} + \sum_{m=n+1}^{2n}\binom{2n}{m}\frac{e^{-\beta\Delta m}}{Z_\beta}. \tag{2.10}$$

Let us analyse this expression. First, we see that if the initial inverse temperature of the system is equal to β, i.e., $p = p_\beta$, the final excitation probability will be unchanged. This has to be the case because a thermal operation cannot bring a system out of equilibrium (see (2.5)). Let us now suppose the system is initially in the ground state. In this case we obtain $p' = p_n^*(\beta)$. It is then easy to check that for $n = 2, \Delta = 1$ and $\beta = 0$ we already obtain $p_n^*(\beta) = p_{\beta=0} + 1/8 = 5/8 > 1/2$. More interestingly, we can consider the limit
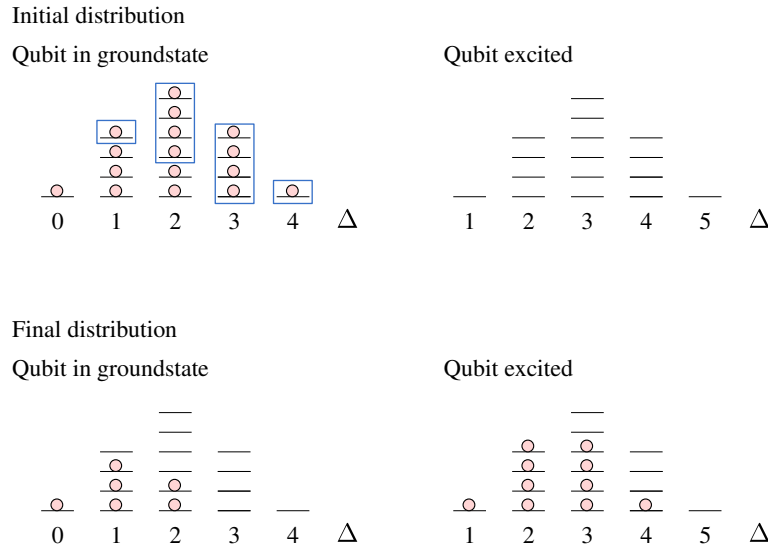
Figure 2.2: Illustration of the algorithm for anomalous heat flow if the system-qubit is initially in the ground state. The heat bath consists of $2n$ two-level systems with the same gap Δ (here $n = 2$). The top row shows the initial configuration, with the left figure illustrating the energy levels E_k^0 of the total system conditioned on the system-qubit being in the ground state and the right figure the energy levels E_k^1 conditioned on the system-qubit being in the excited state. We have $E_k^0 = k\Delta$ and $E_k^1 = (k+1)\Delta$ with $k = 0, \ldots, 2n$. The energy-levels $E_k^{0/1}$ have degeneracy $\binom{2n}{k}$. The red balls denote occupied energy-levels of the full system, the numbers below give the total energy E_k^0 and E_k^1 of the corresponding states, respectively, in units of the gap of the qubits Δ. Initially, only states in which the qubit is in the ground state are occupied. Using an energy-preserving unitary, as many balls as possible are moved to the subspace in which the system-qubit is excited, resulting in the lower row. In the given example, the balls circumscribed by rectangles are mapped to states where the system-qubit is excited, while the remaining balls are untouched. The choice depicted maximizes the final excitation probability for the given size of the bath (note that each of the red balls has a different initial probability, depending on its energy, due to the thermal distribution on the bath). As $n \to \infty$ the final excitation probability approaches $\mathrm{e}^{-\beta\Delta}$.

$n \to \infty$, in which we obtain

$$\lim_{n\to\infty} p_n^*(\beta) = \mathrm{e}^{-\beta\Delta} \tag{2.11}$$

and therefore $p' = \mathrm{e}^{-\beta\Delta}(1 - p)$. This result shows that i) the final temperature of the system becomes larger as the initial temperature becomes colder and ii) that we can obtain a population inversion, and therefore a negative temperature, as long as

$$\Delta < k_\mathrm{B}T \log(2). \tag{2.12}$$

We can also easily prove that the final excitation probability $p' = \mathrm{e}^{-\beta\Delta}(1 - p)$ is optimal in the sense that no thermal operation can achieve $p' > \mathrm{e}^{-\beta\Delta}(1 - p)$. As explained in the previous section, any thermal operation is Gibbs-preserving, i.e. maps thermal states to thermal states (at the environment temperature). On the energy-distribution it therefore acts as a stochastic matrix[2] which has the Gibbs-distribution as a fixed-point. In the case of a two-level system any such matrix can be written as

$$\mathcal{G}_\beta(r) = r\mathbf{1} + (1 - r)\begin{pmatrix} 0 & \mathrm{e}^{-\beta\Delta} \\ 1 & 1 - \mathrm{e}^{-\beta\Delta} \end{pmatrix}, \quad 0 \le r \le 1. \tag{2.13}$$

[2] A stochastic matrix is any matrix that maps probability distributions to probability distributions. Equivalently, it is any matrix with positive entries whose columns sum to one.

Such a map is always a mixture of the identity map and $\mathcal{G}_\beta(0)$. Since the excitation probability is a linear function of the state, its final value is extremized by the extremal operation $\mathcal{G}_\beta(0)$, which yields exactly $p' = \mathrm{e}^{-\beta\Delta}(1 - p)$.

Let me finally argue heuristically why it is in fact not too surprising that it is possible to bring the system to a higher temperature than the heat bath if it was at a colder temperature before. We may imagine that Alice's two-level system is a regular thermodynamic system. If it is initially very cold, this means that its *non-equilibrium free energy* with respect to the environment (exergy) $\Delta F = E - S/\beta - (E_\beta - S_\beta/\beta)$ is large. Here, E and S denote the initial energy and entropy, respectively, and E_β and S_β the corresponding quantities at

thermal equilibrium with the environment. We can therefore in principle use a thermodynamic machine to perform an amount of work $W = \Delta F$ while bringing the system into thermal equilibrium. We can then use this amount of work to run a heat-pump and heat up the system above the environment's temperature. It is clear, that by this procedure the free energy of the system cannot increase, since otherwise we could repeat the procedure and could extract work cyclically from a heat bath.

To make connection to the example, let us define the *non-equilibrium free energy* as

$$\Delta F_\beta(\rho, H) := F_\beta(\rho, H) - F_\beta(\omega_\beta(H), H), \tag{2.14}$$

with $F_\beta(\rho, H) = \mathrm{Tr}(\rho H) - S(\rho)/\beta$ and the von Neumann entropy $S(\rho) = -\mathrm{Tr}(\rho \log \rho)$. If, for simplicity, we assume that Alice's system starts in the ground state and ends in a state with $p' = e^{-\beta\Delta}$ we can then easily see that ΔF_β decreases: Since the thermal parts of the free energies cancel, we have (denoting the final state as ρ' and writing $H_S = \Delta \left|1\right\rangle\left\langle 1\right|$)

$$
\begin{aligned}
\Delta F_\beta(\left|0\right\rangle\left\langle 0\right|, H_S) - \Delta F_\beta(\rho', H_S) &= F_\beta(\left|0\right\rangle\left\langle 0\right|, H_S) - F_\beta(\rho', H_S) \\
&= -p' \log(p')/\beta - (1 - p') \log(1 - p')/\beta - p'\Delta \\
&= -(1 - e^{-\beta\Delta}) \log(1 - e^{-\beta\Delta})/\beta > 0.
\end{aligned}
$$

In the next section we will see that this is a general feature of (catalytic) free transitions: The non-equilibrium free energy, defined as ΔF_β, can never increase under a (catalytic) free transition.

To summarize, we have seen in this simple example how thermal operations can be used to obtain non-trivial effects, such as population inversions, without having to invest additional "resources". The decisive property of the initial state for this to be possible is that it is far from being in thermal equilibrium with the environment – a property which is in general neither measured by the energy nor the entropy alone. In the next section, we will return to the general setting and discuss general laws for free transitions akin to the second law of thermodynamics.

2.3   Monotones, the data-processing inequality and the second laws

IN THE ABOVE EXAMPLE, we have seen that even though it was possible to prepare a population inversion from a ground state using free transitions, it was not possible to increase the non-equilibrium free energy ΔF_β. I will now show that this is a general feature of free transitions and show that their are in fact infinitely many functions that can only decrease under (catalytic) free transitions.

Such functions are called *monotones* of the free transitions, since if $\rho \underset{c}{\rightsquigarrow} \sigma$, we have $f(\rho) \geq f(\sigma)$ for every monotone f. Hence, they are monotonic with respect to the partial order $\underset{c}{\rightsquigarrow}$ induced on states by catalytic free transitions. The importance of monotones comes from the fact that every monotone gives *necessary* conditions for whether a transition $\rho \rightarrow \sigma$ can be realised as a free transition, i.e., without use of additional non-equilibrium resources. They can therefore be interpreted as second laws of thermodynamics [49].

But how can we find such monotones? In fact, using concepts of information theory, it is not difficult. To explain these concepts, let us first consider the case of classical information, represented by probability distributions over some set of events X. Suppose we have two such distributions p and q, inferred from some measurement statistics, and want to quantify how different the two distributions are. By applying a (possibly probabilistic) function to X we can obtain new distributions p' and q' on a, possibly different, space of events Y. Such a transformation is called a *post-processing* or *coarse-graining* of the data and clearly it cannot increase the distinguishability between p and q. On the level of the probability distributions, the post-processing is represented by a stochastic matrix T such that $p' = Tp$ and $q' = Tq$. Therefore, an important property of any measure of distinguishability D between probability distributions p and q is that it fulfills the *data-*

processing inequality:

$$D(p\|q) \geq D(Tp\|Tq)$$

for every stochastic map T. Similarly, in the case of quantum information, the distinguishability between two quantum states ρ and σ can only decrease under any quantum channel $\mathcal{T}$. We therefore ask that any sensible measure of distinguishability of quantum states fulfills the corresponding quantum version of the data-processing inequality:

$$\mathcal{D}(\rho\|\sigma) \geq \mathcal{D}(\mathcal{T}(\rho)\|\mathcal{T}(\sigma)). \tag{2.15}$$

We can use this fact to obtain monotones in thermodynamics (or other resource theories) in a quite general fashion. Intuitively, for a system to be out of equilibrium it means that its quantum state is far from being the Gibbs-state – where the distance is measured with respect to some distinguishability measure between quantum states. Making use of the data-processing inequality, we can now show that this property can only decrease under free transitions:

$$\mathcal{D}(\rho\|\omega_\beta(H)) \geq \mathcal{D}(\mathcal{G}(\rho)\|\mathcal{G}(\omega_\beta(H))) = \mathcal{D}(\mathcal{G}(\rho)\|\omega_\beta(H)), \tag{2.16}$$

where we have used the fact that free transitions always map Gibbs-states to Gibbs-states. If we furthermore assume that the measure $\mathcal{D}$ is additive over tensor-products we can also show the same for catalytic free transitions:

$$\begin{aligned}
\mathcal{D}(\rho\|\omega_\beta(H)) + \mathcal{D}(\chi\|\omega_\beta(H_C)) &= \mathcal{D}(\rho \otimes \chi\|\omega_\beta(H) \otimes \omega_\beta(H_C)) \\
&\geq \mathcal{D}(\mathcal{G}(\rho) \otimes \chi\|\omega_\beta(H) \otimes \omega_\beta(H_C)) \\
&= \mathcal{D}(\mathcal{G}(\rho)\|\omega_\beta(H)) + \mathcal{D}(\chi\|\omega_\beta(H_C)). \tag{2.17}
\end{aligned}$$

Cancelling the terms involving the catalyst[3] χ then again shows $\mathcal{D}(\mathcal{G}(\rho)\|\omega_\beta(H)) \leq \mathcal{D}(\rho\|\omega_\beta(H))$. Catalytic free transitions can therefore only bring a system closer to equilibrium if this distance is measured by a proper distinguishability measure. As we have seen in the previous example, the change of temperature is in fact in general not a good indicator for whether a system has gotten closer to equilibrium.

[3] We are assuming that the distinguishability is always finite or the catalyst can be chosen in such a way.

Any distinguishability measure that fulfills the data-processing inequality automatically defines a monotone of free transitions by setting $f(\rho, H) = \mathcal{D}(\rho\|\omega_\beta(H))$. Important examples of such measures are given by *(quantum) Rényi-divergences*. I will now give the basic definitions that are important for the rest of the thesis, but omit many details. For a detailed treatment, see, for example. Ref. [69] and the references therein.

The case that will be most important for us in the following are the (Petz) Rényi-divergences defined as

$$D_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log \operatorname{Tr}\left(\rho^\alpha \sigma^{1-\alpha}\right). \tag{2.18}$$

For $0 \leq \alpha \leq 2$, they fulfill the data-processing inequality for arbitrary quantum states when they are well defined. If ρ and σ commute and hence share a common eigenbasis, they fulfill the data-processing inequality for any channel $\mathcal{G}$ such that $G(\rho)$ and $G(\sigma)$ also commute. In this case, this holds for any α.

For general quantum states, we can also define the *minimal* quantum Rényi-divergences

$$\tilde{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log \operatorname{Tr}\left(\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right). \tag{2.19}$$

Both D_α and $\tilde{D}_\alpha$ diverge if the support of ρ is not contained in the support of σ and $\alpha > 1$ and we define them to be equal to $+\infty$ in this case. Similarly, D_α diverges for $\alpha < 0$ if the support of σ is not contained in the support of ρ. While these definitions might at first seem a little bewildering, they have the nice property that

$$\lim_{\alpha \to 1} D_\alpha(\rho\|\sigma) = \lim_{\alpha \to 1} \tilde{D}_\alpha(\rho\|\sigma) = D(\rho\|\sigma), \tag{2.20}$$

with the *quantum relative entropy* D defined as

$$D(\rho\|\sigma) = \begin{cases} \text{Tr}(\rho\log(\rho)) - \text{Tr}(\rho\log(\sigma)) & \text{if} \quad \rho \ll \sigma \\ +\infty & \text{otherwise.} \end{cases} \quad (2.21)$$

Here, $\rho \ll \sigma$ means that the support of the operator ρ is contained in that of σ.

In fact, the quantum relative entropy is nothing but the non-equilibrium free energy in disguise:

$$D(\rho\|\omega_\beta(H)) = \beta\Delta F_\beta(\rho, H), \quad (2.22)$$

as can be checked by a simple direct calculation. All the quantum Rényi-divergences are additive over tensor-products and hence also provide monotones of catalytic free transitions. We thus arrive at the important insight that the non-equilibrium free energy ΔF_β, and its cousins defined using (quantum) Rényi-divergences for different values of α, are monotones of catalytic free transitions. They can never increase in a free transitions. We have seen a particular example of this in the case of anomalous heat flow, which also showed that both the excitation probability in a qubit and the temperature are in general not monotones of free transitions. Of course, the monotonicity of the free energy does not imply that we can never increase the free energy, but if we want to do so, we have to take it from other "resource" systems.

In many applications, we can make the simplifying assumption that all states in question commute with the Hamiltonian, i.e., are time-invariant. In this case, drastic simplifications occur. First, the minimal quantum Rényi-divergences simplify to the classical Rényi-divergences:

$$D_\alpha(\rho\|\sigma) = \tilde{D}_\alpha(\rho\|\sigma), \quad \text{if} \quad [\rho, \sigma] = 0. \quad (2.23)$$

Second, we have the following important result, which shows that the Rényi-divergences not only imply necessary conditions for a catalytic free transition, but also provide *sufficient conditions*:

Theorem 2.1 (Necessary and sufficient conditions for catalytic free transitions [49])**.** *Let ρ and ρ' commute with the Hamiltonian H. Then $\rho \underset{c}{\rightsquigarrow} \rho'$ if and only if*

$$D_\alpha(\rho\|\omega_\beta(H)) \geq D_\alpha(\rho'\|\omega_\beta(H)) \quad (2.24)$$

for all $\alpha \in \mathbb{R}$.

This theorem allows to check whether a transition between quasi-classical states can be achieved as a catalytic free transition using only the Rényi-divergences. In chapter 3, we will see how these conditions can be used to study the general problem of ground state cooling.

The conditions given by theorem 2.1 can further be simplified if we slightly modify the concept of catalytic transitions. Suppose we do not require that a catalyst has to be returned exactly, but that the error can be made arbitrarily small. In this case, we can include in the catalyst a pure state and return it in any full-rank state that approximates the pure state to arbitrary accuracy. Then for negative α the initial Rényi-divergence D_α of the whole compound diverges while it is finite in the final state. In this way we can get rid of the conditions for negative α.

Let us now illustrate thermal operations in a few examples. First, we discuss how thermal operations and the non-equilibrium free energy can be used to easily re-derive a famous result: Landauer's bound on the production of heat when erasing information.

2.4   *Illustration: Landauer Erasure and side-information*

INFORMATION IS PHYSICAL in the sense that it has to be represented by states of physical systems. Landauer argued that this implies that the erasure of information then in

general leads to irreversible dissipation of heat by an amount $k_{\mathrm{B}}T\log(2)$ [33, 35], where T is the environment temperature. I will now explain how this bound follows from the previous considerations about monotones. This is of course not a new result, but serves as a simple illustration of the framework.

The setting is the following: We imagine a physical two-level systems (with Hamiltonian H_S), whose levels encode a single bit of information. To erase the information encoded in the system means to reset its state to $|0\rangle$ (or any other fixed, pure reference state), independent of the initial state. We want to achieve this using a (catalytic) free transition. On the one hand, using such a thermal operation, we can always thermalize the system to the environment's temperature for free, for example by swapping first the system with an identical, but thermalized two-level system from the environment. On the other hand, the bit erasure also has to work if the system is initially already thermalized. We will therefore consider without loss of generality the situation in which the initial state of the system is a thermal state.

The first observation that we make is that since the system is initially in equilibrium with the bath, a thermal operation without any additional system cannot be used to erase the bit, because thermal operations cannot bring a system out of equilibrium. We thus have to have access to an additional system, which we call "resource", with a quantum state ρ_R and Hamiltonian H_R. I will assume that this resource is initially uncorrelated with the bit. Later, I will also discuss the case where side-information about the initial state of the system is available. Let the final state of the resource be ρ_R'. From the monotonicity and additivity of the free energy we then get:

$$
\begin{aligned}
\Delta F_\beta(\rho_R \otimes \omega_\beta(H_S), H_R + H_S\otimes) &= \Delta F_\beta(\rho_R, H_R) \\
&\geq \Delta F_\beta(\rho_R' \otimes |0\rangle\langle 0|, H_R + H_S) \\
&= \Delta F_\beta(\rho_R', H_R) + \Delta F_\beta(|0\rangle\langle 0|, H_S),
\end{aligned} \tag{2.25}
$$

where we have used that a system in a pure state cannot be correlated to another system. Since $\Delta F_\beta(|0\rangle\langle 0|, H_S) = 0 - E_\beta + S_\beta/\beta$ we thus obtain:

$$
F_\beta(\rho_R, H_R) - F_\beta(\rho_R', H_R) \geq k_{\mathrm{B}}T S_\beta - E_\beta. \tag{2.26}
$$

Here, S_β and E_β denote the thermal entropy and energy of the bit to be erased. Let us consider some special cases:

- If the Hamiltonian of the bit does not discriminate between the two levels ($H_S \propto \mathbf{1}$), we obtain $F_\beta(\rho_R, H_R) - F_\beta(\rho_R', H_R) \geq k_{\mathrm{B}}T\log(2)$.

- In the case where the state of the resource system does not change its entropy, we may want to interpret the change of energy on it as "work" provided by the resource (see, however, chapter 4). In this case we get $W \geq k_{\mathrm{B}}T\log(2)$.

- In the case where the resource system does not change its energy, but only its entropy, we see that its entropy has to increase at least by the thermal entropy of the bit S_β. We can then say that an amount of *heat* $Q = T S_\beta$ has been generated in the resource.

Intuitively, it should be easier to erase the information, if we have some knowledge about the state of the bit before we want to erase it. For example, if the bit is known to be in state $|1\rangle$ and H_S is trivial, then we can simply flip the spin and no resource is necessary. I will now demonstrate in general how such correlations affect the erasure bound in a simple example.

Since information is physical, having prior knowledge means that there exists a physical system which is correlated with the bit. We will call this system a memory M. The memory and system together are initially in some correlated state described by the density matrix ρ_{SM}. To simplify the problem, let us assume $H_S \propto \mathbf{1}$ and $H_M \propto \mathbf{1}$ and assume that the initial state on S is maximally mixed, $\rho_S = \mathbf{1}/2$. The final state of the system definitely is $|0\rangle$ and is hence uncorrelated with the memory. I will now make the restriction that the memory ends up in the same internal state ρ_M since I want to highlight only the effect of the correlations. In principle, if ρ_M is not given by the maximally mixed state, we could

use the "non-equilibriumness" in the memory later on to perform further thermodynamic tasks. Summarizing, we thus find that the initial entropy of SM is given by $S(\rho_{SM})$ and the final entropy by $S(\rho_M)$. Arguing as above by regarding the compound SM as a single system, we then find

$$F_\beta(\rho_R, H_R) - F_\beta(\rho'_R, H_R) \geq k_\mathrm{B}T\left(S(\rho_{SM}) - S(\rho_M)\right)$$
$$= k_\mathrm{B}TH(S|M)_\rho, \tag{2.27}$$

where we have introduced the *conditional entropy* $H(S|M)_\rho := S(\rho_{SM}) - S(\rho_M)$, which is a measure for the information about S contained in M. Interestingly, in quantum mechanics, the conditional entropy can be negative if the initial state ρ_{SM} is entangled. In such a case, one can in principle *extract non-equilibriumness* by erasure instead of investing non-equilibriumness to erase information. For a more thorough discussion of this problem, see Ref. [70].

We have seen how the formalism of thermal operations allows us to derive in a simple way the fundamental Landauer bound including side-information. However, we have only derived *necessary* conditions for erasure. In particular, we did not make use of theorem 2.1, which tells us that we should check the infinite family of second laws given by the Rényi-divergences. We will come back to this problem in chapter 3, where we will find that there are much more stringent conditions then just Landauer's bound when we want to bring a system to the ground state, which can be interpreted as a quantitative version of the third law of thermodynamics.

2.5   Illustration: The infinite ladder battery and fluctuations of energy

In recent years, the concept of *fluctuation relations* [37, 44, 45, 71, 72] had big impact on the field of statistical mechanics. It originates from studying stochastic thermodynamics, which is concerned with the thermodynamic properties of individual particles which can come in contact with heat baths. In stochastic thermodynamics, the particle in question undergoes dynamical trajectories under influence of stochastic forces, leading to stochastic notions of basic properties in thermodynamics, such as entropy production (see section 2.6), heat dissipation or work being done by or on the particle. Fluctuation relations provide, under quite general assumptions, detailed and universal relations for the probability distributions of various quantities, such as the work being done by the particle along a trajectory.

In the context of quantum mechanics, a simple fluctuation relation is concerned with the energy fluctuations when a system that is initially in thermal equilibrium undergoes unitary dynamics. To understand the relation, consider the following operational protocol. Assume you have a system initially in a thermal state. First you perform an energy measurement, obtaining an outcome $E^{(i)}$ and projecting the state of the system to the energy eigenstate $|E^{(i)}\rangle$. Then you let the system undergo unitary dynamics U, which possibly changes the Hamiltonian over time and is *not* required to conserve the energy. Then you measure again the final energy to obtain some outcome $E^{(f)}$. Repeat the experiment many times and determine the distribution $P(\Delta E)$ of the random variable $\Delta E = E^{(i)} - E^{(f)}$. Then the distribution $P(\Delta E)$ fulfills the *Jarzysnki equation*

$$\mathbb{E}(e^{\beta \Delta E}) = \frac{Z_\beta(H^{(f)})}{Z_\beta(H^{(i)})}, \tag{2.28}$$

where $\mathbb{E}(\cdot)$ denotes the expectation value of a random variable. For a simple derivation of this relation, see Ref. [73]. Often, the energy-difference ΔE is interpreted as fluctuating, mechanical work being done by the system. Indeed, using Jensen's inequality, we can obtain the inequality $\mathbb{E}(\Delta E) \leq F_\beta(H^{(i)}) - F_\beta(H^{(f)})$, which can then be interpreted as an expression of the second law of thermodynamics relating work to free energy changes. On the other hand, (2.28) is an *equality* which contains information about higher cumulants of ΔE and hence provides more detailed information about possible fluctuations of work. Fluctuation relations can, however, also be de-rived under different assumptions in the

quantum regime and there is by now a vast body of literature dealing with fluctuation relations in the quantum setting (for a selection, see Refs. [71–79] and references therein).

It is then probably not surprising that similar relations can also be derived in the framework of thermal operations [59, 60]. I will now sketch how this can be done. Clearly, to derive such relations in the framework of thermal operations, it is necessary to introduce an auxiliary system W which can take up energy from the system of interest or give energy to it. Here, we will take the stance that this external energy is supposed to model intuitively what one considers to be mechanical work. Such a system should then have the property that it resembles the classical notion of "lifted weight", being able to take up and release arbitrary amounts of energy. Furthermore, the weight should not act as an entropy sink. Finally, at least the average work provided by the weight or done on the weight should not be dependent on the initial state of the weight.

Indeed, there is a simple, but highly idealized model which achieves all these demands [60, 61, 80]. It is given by a continuous one-dimensional degree of freedom whose energy is simply proportional to the position operator X: $H = mgX$. The additional constraint on the unitary U that implements the thermal operations and ensures that the weight does not act as an entropy sink is that the unitary commutes with the momentum operator P on the weight. This has the consequence that the operation on the system and heat bath is a mixture of unitaries [80]:

$$\Gamma_{SB}(\rho_S \otimes \omega_\beta(H_B)) := \mathrm{Tr}_W(U\rho_S \otimes \omega_\beta(H_B) \otimes \rho_W U^\dagger)$$
$$= \int U_{SB}(p)\rho_S \otimes \omega_\beta(H_B)U_{SB}^\dagger(p)P_W(p)\mathrm{d}p, \qquad (2.29)$$

where ρ_W is the initial density matrix on the weight, $P_W(p)$ is the initial momentum distribution on the weight and $U_{SB}(p)$ is a momentum-dependent family of unitaries on SB. Such a mixture of unitaries cannot decrease the von Neumann entropy of a system[4] and therefore, the weight does not act as an entropy sink. In particular, this assumption implies

[4] This can easily be seen by considering a mixture of unitaries as a Gibbs-preserving map for a trivial Hamiltonian.

$$\Gamma_{SB}(\mathbf{1}_{SB}) = \mathbf{1}_{SB}. \qquad (2.30)$$

Note that if the initial state on the weight is (approximated arbitrarily well by) a momentum-eigenstate, the dynamics on the system together with the heat-bath is simply (approximated arbitrarily well by) unitary dynamics.

As shown in Ref. [60], the classical Jarzynski relation (2.28) can then be recovered if one additionally assumes that the induced map on the system and *weight* Γ_{SW} maps energy-eigenstates to mixtures of energy-eigenstates. I will now present a formal derivation of this result. In this classical case, the action on SW for states diagonal in the energy-eigenbasis can be represented by conditional probabilities $P(E_{s'}, w|E_s)$:

$$\Gamma_{SW}(|E_s\rangle\langle E_s| \otimes |x\rangle\langle x|) = \sum_{s',w} P(E_{s'}, w|E_s) |E_{s'}\rangle\langle E_{s'}| \otimes |x+w\rangle\langle x+w|, \quad (2.31)$$

where $|x\rangle$ denote (generalized) energy-eigenstates of the weight and $\{E_s\}$ is the spectrum of H_S[5]. Due to the commutation constraint $[U, P] = 0$, the conditional probabilities $P(E_{s'}, w|E_s)$ are in fact independent of the initial state on the weight as long as the initial state on W is diagonal in the position basis (this excludes the possibility of a purely unitary dynamics on the system and bath and is analogous to the requirement of performing energy-measurements before and after a protocol in different frameworks). One can thus choose without loss of generality that this state is $|0\rangle$ and refer to the event that this state is changed to a state $|w\rangle$ as *extracting an amount of work w*. The quantity $P(E_{s'}, w|E_s)$ then denotes the conditional probability that the system goes from an energy eigenstate $|E_s\rangle$ to an energy eigenstate $|E_{s'}\rangle$ while an amount of work w is extracted.

[5] We assume that the Hamiltonian on S remains unchanged, see [60] for a generalization which allows for changing the Hamiltonian on H_S.

The Jarzynski relation follows by making use of the explicit definition of Γ_{SW}. First, we can use that U conserves the energy in form of the relation

$$U\mathrm{e}^{-\beta H_B} = \mathrm{e}^{-\beta(H_S+H_B+H_W)}U\mathrm{e}^{\beta(H_S+H_W)}. \qquad (2.32)$$

With this relation we can write

$$P(E_{s'}, w|E_s) = \langle E_{s'}, w| \mathrm{Tr}_B\left(U\omega_\beta(H_B) \otimes |s\rangle\langle s| \otimes |0\rangle\langle 0| U^\dagger\right) |E_{s'}, w\rangle$$
$$= \langle E_{s'}, w| \mathrm{e}^{-\beta(E_{s'}-E_s+w)}\mathrm{Tr}_B\left(\omega_\beta(H_B)U\mathbf{1}_B \otimes |s\rangle\langle s| \otimes |0\rangle\langle 0| U^\dagger\right) |E_{s'}, w\rangle.$$

Re-arranging and summing over s and w, we then obtain (using (2.30))

$$\sum_{s,w} P(E_{s'}, w|E_s) e^{-\beta(E_s - w)} = e^{-\beta E_{s'}} \langle E_{s'} | \operatorname{Tr}_B \left(\omega_\beta(H_B) \operatorname{Tr}_W \left(U \mathbf{1}_{SB} \otimes |0\rangle\langle 0| U^\dagger \right) \right) | E_{s'} \rangle$$

$$= e^{-\beta E_{s'}}. \tag{2.33}$$

Dividing both sides by the partition function of H_S and summing over s', this yields the Jarzynski relation for thermal initial states:

$$\mathbb{E}\left(e^{\beta w} \right) = \sum_{s,s',w} P(E_{s'}, w|E_s) \frac{e^{-\beta E_s}}{Z_\beta(H_S)} e^{\beta w} = 1. \tag{2.34}$$

In fact, in the setting of thermal operations, more general *quantum fluctuation relations* can be de-rived using similar methods as presented above. For details, see Ref. [59, 60].

2.6   Connection to thermodynamics in terms of open system dynamics

Before concluding this chapter, let me briefly comment on the connection between the monotones as defined in this section and more conventional models of quantum thermodynamics in terms of open system dynamics. Often dynamics of open systems in contact with heat baths are modelled by general Markovian dynamics, which need not be unitary. For example, in classical statistical mechanics one often describes the dynamics of an open system by a Markov chain or its continuous time analogues. Similarly, in the context of quantum mechanics, the dynamics of open systems is frequently modelled by a so-called dynamical semi-group (see, for example, Refs. [71, 81, 82]). Such a dynamical semi-group, which is simply the most general form of Markovian dynamics in quantum mechanics, for example arises if one weakly couples a small system to a reservoir. Formally, the dynamics then takes the form

$$\rho(t) = \mathcal{E}_t(\rho(0)), \quad \mathcal{E}_{t+s} = \mathcal{E}_t \circ \mathcal{E}_s, \quad \mathcal{E}_0 = 0. \tag{2.35}$$

Here, $\mathcal{E}_t$ is a quantum channel for every $t \geq 0$ and is generated by a *Lindbladian* or *Liouvillian* $\mathcal{L}^*$:

$$\mathcal{E}_t(\rho) = e^{t\mathcal{L}^*}(\rho). \tag{2.36}$$

The general form of $\mathcal{L}^*$ has been derived by Lindblad in Ref. [83], but is not important for the present discussion (see chapter 12 for the explicit form of a Liouvillian). This formulation naturally includes the description of classical systems by embedding them diagonally in any arbitrarily chosen orthonormal basis of the Hilbert space. The action of the channel $\mathcal{E}_t$ in this basis is then simply given by a stochastic matrix.

If the dynamics $\mathcal{E}_t$ is supposed to describe dynamics that occurs while the system is immersed in a thermal bath at temperature β, it is a very natural assumption that after sufficiently long time the system reaches a thermal state $\omega_\beta(H)$. Indeed, in the case mentioned above, where a finite system is weakly coupled to a heat bath, one can show under very general assumptions show that this is true [84]. This then leads to the conclusion that the channels $\mathcal{E}_t$ have the thermal state as a fixed-point. They are hence examples of the Gibbs-preserving maps that we discussed earlier, although here derived from a different point of view. In the field of open system dynamics, a seminal observation was that one can in this case define an *entropy production* function as [85–88]

$$\sigma(t) := -\frac{\mathrm{d}}{\mathrm{d}t} D(\rho(t) \| \omega_\beta(H)), \tag{2.37}$$

which is always non-negative due to the data-processing inequality: $\sigma(t) \geq 0$. It is natural to interpret this inequality as an expression of the second law of thermodynamics.

From the modern perspective of quantum thermodynamics as a resource theory, the definition of the entropy-production $\sigma(t)$ is completely natural, although maybe the name *non-equilibrium free energy decrease* would be more appropriate. Indeed, we could now

introduce an infinite family of entropy-production functions $\sigma_\alpha(t)$ using the (quantum) Rényi-divergences introduced in section 2.3. The question is then rather: Which property singles out the entropy-production function defined through the quantum relative entropy $\sigma(t)$? The answer will be given in chapter 5, where we will see that only the quantum relative entropy remains as a monotone of catalytic thermal transitions if we allow the system to become correlated with potential catalysts. This is a natural assumption in any open-system framework, since the bath and potential catalysts are not even included in the description. Note that a connection between entropy production and the fact that a system becomes correlated with its surroundings has been made before, however from a different point of view, see for example [89].

2.7   Conclusions and Outlook

IN THIS CHAPTER I have reviewed the framework of catalytic thermal operations as a way to model thermodynamics in a general way.

Even though this framework allows for arbitrary quantum mechanical control over arbitrary systems, we have seen that it nevertheless puts tight constraints on possible physical processes that involve thermal systems. It thus establishes a non-obvious fact: Namely that thermodynamic laws (or rather, restrictions) also play an important role for individual, small quantum systems. At the same time we have seen, in the example of anomalous heat flow, that it allows for processes that might seem unintuitive at first sight.

In this chapter, I only discussed the basics of this framework in the most idealized setting. Recently, progress has been made in deriving conditions that arise from finite heat baths [90] as well as deriving explicit ways to implement a given thermal transition using a small set of "elementary operations", somewhat similar to a universal gate set in the theory of computation [91]. It is also possible to include more conserved quantities than the energy, such as angular momentum, using Generalized Gibbs ensembles instead of canonical ensembles (see Refs. [53, 92, 93] and chapter 10).

The framework of catalytic thermal operations has the structure of what is today known in general as a "resource theory" [94–100]: It consists of a set of "free operations" (in our case: energy preserving unitaries or, more generally, Gibbs-preserving maps) and "free states" (systems in thermal equilibrium with the environment). The free operations always map free states to free states. Hence if one wants to prepare a state that is not a free state out of a free state, one has to have access to a non-free state and apply a free operation to the whole compound. Thus states that are not free (in our case any state that is not a Gibbs state at the environment temperature) can be used to implement non-free operations and are therefore considered to be "resources".

In the context of quantum physics, and especially quantum information theory, the concept of a resource theory is best known from the theory of entanglement, where the free states correspond to product states and the free operations to local operations with classical information [62]. It was noticed fairly early in the field of quantum information theory that this formulation of entanglement theory resembles the theory of thermodynamics [101, 102], which of course fits naturally in the language of a resource theory (see, in particular, the approach of Lieb and Yngvason to thermodynamics as an abstract axiomatic framework [23, 24, 103]). The explicit formulation of thermodynamics as a resource theory in the formal framework of quantum mechanics nevertheless only took place fairly recently and lead to a rich body of work on the formal resource theoretical aspects of quantum thermodynamics (see, for example, [46, 48, 53, 103–105]) Indeed, very recently results from thermodynamics have now found a way back into the theory of entanglement, providing "fluctuation relations of entanglement" [106].

By now, it is common place in quantum information theory to formulate problems in terms of a resource theory. As examples, let me mention here the formulation as resource theories of such diverse fields as coherence, reference frames and (a-)symmetry [54, 57, 58, 107–115], steering [116] or stabilizer quantum computation [117]. Recently, also (quantum) cryptography has been cast into a formulation which closely resembles a resource theory [118].

An important concept in almost any resource theory is that of a monotone, namely any function on the set of all possible states that can only decrease under free transitions and thereby provides necessary conditions for state transitions. Any monotone also provides a measure for the "resourcefulness" of a state. We have seen in our examples how such monotones can be used to derive fundamental bounds in a simple way. Unsurprisingly, they will continue to play important roles in the rest of this thesis.

The generality of the resource theoretical framework of quantum thermodynamics is at the same time a drawback: While it allows to derive fundamental bounds in a simple way, it does not tell us exactly how these bounds can be realized in concrete physical terms. Furthermore it is not so easy to include possible additional physical restrictions. That is why we will have to adapt and simplify the model in chapter 9, where I will discuss the role of additional experimental restrictions on thermodynamic bounds for thermal machines.

Before, however, we will now use the more general framework to discuss fundamental problems in thermodynamics from the point of view of resource theories. First, in chapter 3 we derive in a rigorous way quantitative bounds on the sufficient and necessary amount of resources needed to bring a system close to its ground state. In other words, we will see how the third law of thermodynamics emerges. Then, in chapter 4 we discuss whether and how one can generalize the concept of thermodynamic work to quantify in a meaningful way how useful a transition between two states of a system is. This will be followed up by a new axiomatic derivation of the relative entropy and non-equilibrium free energy in chapter 5. In chapter 6, we will see that a resource theoretical approach to thermodynamics can in fact be used to *derive* maximum entropy ensembles like the Gibbs state. Finally, in chapter 8 we discuss the problem of equilibration in closed many-body systems using tools inspired by the resource theory of thermal operations.

3 Ground state cooling and the unattainability principle

QUANTUM INFORMATION PROCESSING requires highly pure quantum states as a resource to perform computational tasks [62]. In this chapter I will discuss the fundamental limitations to prepare very pure quantum states from a thermodynamic point of view. To make the problem more concrete, I will focus on bringing the system to an energy-eigenstate. In the framework of catalytic thermal operations, any energy-eigenstate may be converted to the ground state without requiring additional resource since for any $\alpha \geq 0$ we have

$$D_\alpha(|E_i\rangle\langle E_i| \, \| \, \omega_\beta(H)) = \beta E_i + log(Z) \geq D_\alpha(|E_0\rangle\langle E_0| \, \| \, \omega_\beta(H)), \qquad (3.1)$$

where E_i denote the energy-eigenlevels of H and E_0 the ground state energy.

Thus, the ground state is the energy-eigenstate that is easiest to prepare from a thermodynamic perspective. It is therefore sensible to first study the problem of bringing a system to its ground state, or, in other words, the problem of ground state cooling [1].

But then we face a problem from the point of view of thermodynamics: The third law of thermodynamics, or more precisely, the unattainability principle, states that cooling a system exactly to its ground state requires infinite resource, being it in the form time, space, work or other resources [61, 119–122]. Similar no-go theorems have been derived for the task of bit-erasure (see section 2.4): They show that no unitary process on a system plus a finite dimensional heat bath can bring the system from a mixed state to a pure state [123–125]. This ultimately follows since the rank of a matrix cannot be decreased by a unitary operation, since unitary operations preserve the spectrum of any operator. Thus we can conclude that if we do not already have access to a pure quantum state, we cannot prepare a pure quantum state exactly.

From a more practical perspective the only sensible question is then to ask how many resources are necessary to approximate a pure state arbitrarily well. Indeed, recently a sizable number of studies deal with different protocols to cool a small quantum system by unitarily acting on heat bath and a certain number of systems out of equilibrium to be "used up" (known under the name of algorithmic or dynamical cooling) [126–130] or studying particular models of refrigerating small quantum systems [131–137], including ones that claim to challenge the unattainability principle in terms of required time (see Ref. [138–145] for such claims and counter-claims).

In this chapter, I will address the problem ground state cooling from a resource-theoretic perspective as introduced in chapter 2. There have been previous studies on cooling using

[1] This chapter is based on joint-work with Rodrigo Gallego, published in [8].

(parts of) the resource-theoretic framework, but focussing on very specific resources such as time, work-fluctuations or Hilbert-space dimension [61, 121, 122]. Here, instead of asking how long it takes to cool a system to a certain temperature, e.g., as measured by the cycles of a periodically working machine, or how much work is needed to do so, the discussion will be phrased in more abstract terms.

The task of cooling that I will be considering can be phrased as finding a cooling protocol that, given a resource (ρ_R, H_R) and a target system with Hamiltonian H_S initially in thermal equilibrium with the environment at inverse temperature β, brings the system to a thermal state ρ_S at a temperature T_S as low as possible. I will assume that the density matrix of the resource has full rank. Indeed if this was not so, for example if the resource was pure, the problem would trivialize. In the simplest case, one could simply swap the resource and the target and bring the system exactly to a pure state [2].

The cooling protocol consists of finding a catalyst (σ_C, H_C) and a corresponding catalytic thermal operation on the resource and the target which has the desired effect on the target system. It is therefore a unitary operation on the compound of heat bath, resource, target and catalyst and can therefore be seen as an instance of heat-bath algorithmic cooling. In the notation established in chapter 2 we ask when the following transition is possible:

$$\rho_R \otimes \omega_\beta(H_S) \underset{c}{\rightsquigarrow} \rho'_{RS}, \quad \mathrm{Tr}_R(\rho'_{RS}) = \omega_{\beta_S}(H_S), \tag{3.2}$$

with $\beta_S = 1/T_S$ as large as possible.

From theorem 2.1 and the discussion below it, we know that, in the case of states that are diagonal in the energy-eigenbasis, to check whether such a transition is possible, in principle we have to check the infinite set of inequalities

$$D_\alpha(\rho_R \otimes \omega_\beta(H_S) \| \omega_\beta(H_R) \otimes \omega_\beta(H_S)) \geq D_\alpha(\rho'_{RS} \| \omega_\beta(H_R) \otimes \omega_\beta(H_S)), \quad \alpha \geq 0.$$

If we trace out the resource in the final state we then get the set of inequalities

$$D_\alpha(\rho_R \| \omega_\beta(H_R)) \geq D_\alpha(\rho'_S \| \omega_\beta(H_S)), \quad \alpha \geq 0. \tag{3.3}$$

Here, we allowed for a catalyst that changes by an arbitrarily small amount to get rid of the conditions for negative α as discussed below theorem 2.1. We will later also discuss the case where the catalyst cannot change at all and the case where the catalyst is returned with a finite error. The functions D_α depend in a complicated way on the density matrices and Hamiltonians. Therefore it is in general difficult to decide for which final temperatures T_S the cooling task can be achieved. In this chapter, I will introduce a new *monotone* of thermal operations, called the *vacancy* $\mathcal{V}_\beta$, and show that it alone determines whether a cooling protocol can be found if T_S is very small. In rough terms we will see that the necessary and sufficient condition expressed in the infinite set of inequalities (3.3) essentially collapses to a single necessary and sufficient condition for cooling:

$$\mathcal{V}_\beta(\rho_R, H_R) \gtrapprox \mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S). \tag{3.4}$$

Moreover, as the target temperature approaches zero, $T_S \to 0$, the vacancy diverges, $\mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S) \to \infty$. Hence, an infinite amount of resources, as measured by the vacancy, are required to cool exactly to zero temperature. The fact that the vacancy is an important quantity for ground state cooling has been found previously in the very particular case where the target system is a qubit and the resources are *i.i.d.*, i.e., many copies of the same system [46]. Using a completely different approach, we will see that the vacancy is the crucial quantity also in the general setting.

As a very simple example, we can consider as resource a collection of N identical harmonic oscillators (with Hamiltonian $H_{\mathrm{osc.}}$), each initially in a thermal state at a temperature T_R hotter than the environment. We want to use the temperature difference to the environment to cool one of them to a very low temperature. Then our results show that this is possible in the above sense if and only if

$$N \geq \frac{\mathcal{V}_\beta(\omega_{\beta_S}(H_{\mathrm{osc}}), H_{\mathrm{osc}})}{\mathcal{V}_\beta(\omega_{\beta_R}(H_{\mathrm{osc}}), H_{\mathrm{osc}})}. \tag{3.5}$$

[2] Similar arguments can be made for other states without full rank. For example two copies of a rank-2 state in a 4-dimensional system can be written as a pure state in tensor product with a full-rank state. Such a resource therefore already contains a pure state which can then be mapped to the ground-state.

Thus, as the target temperature T_S approaches absolute zero, the number N of the resource oscillators out of equilibrium has to diverge.

The results in this chapter show that to cool a system to zero temperature, it has to effectively interact with infinitely many (or one infinitely large) resource systems. This implies that also an infinite amount of time is needed, since each such interaction takes a finite time (see [61] for a detailed discussion of this point). We thus also recover the unattainability principle in terms of the time that it takes to bring a system to its ground state exactly.

In the following, I first properly define the vacancy, establish its basic properties and derive the necessary condition for cooling expressed in terms of the vacancy. After that I state the main result, which shows that the vacancy also provides the sufficient condition for cooling. Then I discuss special cases of particular physical interest and discuss how the vacancy scales with various parameters. Finally, I discuss in more detail the role of the catalysts and come back to the problem of Landauer erasure. All proofs can be found in section 14.2 unless they are provided in the main text of the chapter.

3.1   Definition and basic properties of the vacancy

The vacancy is simply defined as

$$\mathcal{V}_\beta(\rho, H) := D(\omega_\beta(H)\|\rho), \tag{3.6}$$

where D is the quantum relative entropy defined in (2.21). From the definition of the relative entropy, it follows immediately that $\mathcal{V}_\beta(\rho, H) = +\infty$ if ρ does not have full rank. Thus the vacancy is in fact a highly discontinuous function when considered as a function of arbitrary quantum states. It is, however, continuous over the set of full-rank states on a finite dimensional Hilbert-space.

Let us now show that the vacancy is an additive monotone of catalytic thermal transitions.

Lemma 3.1 (Monotonicity). *The vacancy is an additive monotone under catalytic thermal transitions.*

Proof. First note that the vacancy is only defined on full-rank states. From the additivity of the relative entropy under tensor-products it follows that the vacancy is additive over uncorrelated and non-interacting subsystems. Now let ρ be any full-rank state with associated Hamiltonian H_S and consider a thermal operation involving a catalyst that maps ρ to ρ'.

From the results in chapter 2, we know that the Rényi-divergences for $0 \leq \alpha \leq 2$ can only decrease under such an operation:

$$D_\alpha(\rho\|\omega_\beta(H_S)) \geq D_\alpha(\rho'\|\omega_\beta(H_S)), \quad 0 \leq \alpha \leq 2. \tag{3.7}$$

We can now expand this equation around $\alpha = 0$. A direct calculation shows that

$$\lim_{\alpha' \searrow 0} \frac{\mathrm{d}D_\alpha(\rho\|\omega_\beta(H_S))}{\mathrm{d}\alpha}\bigg|_{\alpha=\alpha'} = D(\omega_\beta(H_S)\|\rho) = \mathcal{V}_\beta(\rho, H_S). \tag{3.8}$$

Since $D_0(\rho\|\omega_\beta(H_S)) = 0$ by the assumption that ρ has full rank, we then get

$$\mathcal{V}_\beta(\rho, H_S)\alpha + O(\alpha^2) \geq \mathcal{V}_\beta(\rho', H_S)\alpha + O(\alpha^2). \tag{3.9}$$

Dividing by α and taking the limit $\alpha \to 0$, we then obtain $\mathcal{V}_\beta(\rho, H_S) \geq \mathcal{V}_\beta(\rho', H_S)$ as required. $\square$

Note that the monotonicity holds for arbitrary quantum states, not only states that are diagonal in the energy basis. We can now easily derive a necessary condition for cooling by applying the vacancy to the resource and system together:

$$\mathcal{V}_\beta\left(\rho_R \otimes \omega_\beta(H_S), H_R + H_S\right) \geq \mathcal{V}_\beta\left(\rho'_{RS}, H_R + H_S\right)$$
$$\geq \mathcal{V}_\beta\left(\omega_\beta(H_R) \otimes \rho_S, H_R + H_S\right).$$

The last inequality follows from the fact that one can always replace the state on any system by an uncorrelated thermal state at the heat bath's temperature using a thermal operation. Using additivity of the vacancy and the fact that $\mathcal{V}_\beta(\omega_\beta(H), H) = 0$, we then find that a completely general necessary condition for cooling is given by

$$\mathcal{V}_\beta(\rho_R, H_R) \geq \mathcal{V}_\beta(\rho_S, H_S). \tag{3.10}$$

The vacancy $\mathcal{V}_\beta$ was in fact first introduced (under a different name) in Ref. [46], where its use for cooling was discussed in the special case of a two-level system as a target system. In this chapter, we go beyond the results in Ref. [46] by treating arbitrary systems.

From the necessary condition (3.10) the relevance of the vacancy for the unattainability principle is clear: Since a ground state does not have full support and the vacancy diverges in this case, it shows that infinite resources, as measured by the vacancy, are necessary to cool a system to zero temperature. In particular, if the resource is given by n independent systems, then n has to diverge as long as each of the systems has a finite vacancy.

We will now see that the decrease in vacancy is not only a necessary condition, but that in many cases it is also a *sufficient* condition for cooling. It thus completely characterizes the problem of approximate ground state cooling.

3.2   Sufficiency: A general theorem

As alluded to before, the vacancy not only captures the necessary but also the sufficient condition for cooling in the case where the target temperature is extremely low. Furthermore, we need the additional assumption that the resource state is diagonal in the energy-eigenbasis. This is made precise in the following general theorem.

Theorem 3.2 (General sufficient condition for cooling). *For every choice of β and H_S there is a critical $\beta_{\mathrm{cr}} > 0$ such that for any $\beta_S > \beta_{\mathrm{cr}}$ and any diagonal full-rank resource (ρ_R, H_R) the condition*

$$\mathcal{V}_\beta(\rho_R, H_R) - K(\beta_S, \beta, \rho_R, H_R, H_S) \geq \mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S) \tag{3.11}$$

is sufficient *for cooling. The function K is positive semi-definite and has the property $K(\beta_S, \beta, \rho_R, H_R) \to 0$ as $\beta_S \to \infty$ for any fixed $\beta, H_R, \rho_R > 0$ and H_S.*

The actual proof of this result is somewhat involved and can be found in section 14.2.1, but a sketch of the main ideas involved in the proof is given at the end of this section. The function K is given by

$$K(\beta_S, \beta, \rho_R, H_R, H_S) = \max\left\{0, -\min_{\alpha \leq \delta(\beta_S)} D''_\alpha(\rho_R \| \omega_\beta(H_R))\right\} \delta(\beta_S),$$

where

$$\delta(\beta_S) := \log(Z_\beta(H_S)) / \mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S).$$

The bound (3.11) is completely general and applies for any diagonal resource state. In practice, however, it can be difficult to compute $K(\beta_S, \beta, \rho_R, H_R, H_S)$ since it involves an optimization over a small range of α.

It is therefore important to study whether one can find general physical conditions for resource systems (ρ_R, H_R) such that $K(\beta_S, \beta, \rho_R, H_R, H_S) = 0$. In such a case, the general sufficient condition (3.11) taken together with the necessary condition (3.10) implies that a *necessary* and *sufficient* condition is given by

$$\mathcal{V}_\beta(\rho_R, H_R) \geq \mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S), \tag{3.12}$$

In the next section, I will show that this indeed holds true for large classes of thermal resources.

From the additivity of the Rényi-divergences, it follows that $K(\beta_S, \beta, \rho_R, H_R, H_S)$, just as the vacancy, is additive over non-interacting and uncorrelated resources. This property will be important in section 3.3, where I discuss the setting of identically and independently distributed (i.i.d.) resources.

The result given above is tailored to thermal target states. This is in fact not necessary. With much more technical effort, it is possible show (see Ref. [8]) that a completely analogous result holds for states of the form

$$\rho_\epsilon = (1 - \epsilon)\,|0\rangle\langle 0| + \epsilon\rho^\perp, \quad \epsilon \ll 1, \tag{3.13}$$

where $\rho^\perp$ is any density matrix which has full rank on the subspace orthogonal to the ground state $|0\rangle$ and commutes with H_S. Nevertheless, for the rest of the chapter I will for simplicity only consider thermal target states.

3.2.1   Sketch of the proof of theorem 3.2

Let us now sketch the proof of theorem 3.2. The essential idea behind the proof is that for sufficiently low target temperatures ($\beta_S > \beta_{\rm cr}$) the general sufficient conditions in (3.3) only give non-trivial constraints for very small values of α: for $\beta_S > \beta_{\rm cr}$ the right-hand side of (3.3), given by $D_\alpha(\omega_{\beta_S}(H_S)\|\omega_\beta(H_S))$, quickly saturates to its maximum value as we increase α and it is concave (see Fig. 3.1). To check the conditions (3.3) we can therefore perform a Taylor expansion around α to get

$$\begin{aligned} D_\alpha(\rho_R\|\omega_\beta(H_R)) &\approx D_0(\rho_R\|\omega_\beta(H_R)) + \left.\frac{\partial D_\alpha(\rho_R\|\omega_\beta(H_R))}{\partial\alpha}\right|_{\alpha=0}\alpha \\ &\quad + k\alpha^2 \\ &= \mathcal{V}_\beta(\rho_R, H_R)\alpha + k\alpha^2. \end{aligned} \tag{3.14}$$

On the other hand, one can show that the divergences $D_\alpha(\omega_{\beta_S}(H_S)\|\omega_\beta(H_S))$ are concave in α for small enough β_S. One can then upper bound them simply using the linear approximation at the origin, which is given by $\mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S)\alpha$. Properly taking account of the errors made in the Taylor approximation then yields the theorem.

3.3   i.i.d. *resources and the scaling of the vacancy*

So far results did not depend explicitly on any structure of correlations in the resource systems. In this section, I will discuss the paradigmatic case where an experimenter simply has a large number n of uncorrelated copies of some resource system. This setting also allows us to study in a simple way how the vacancy scales with the system size of the resource and what are other important scales in the problem.

Thus, in this section the resource takes the form $\rho_R = \varrho_R^{\otimes n}$ with Hamiltonian $H_R = \sum_i h_R^i$ where $h_R^i = \mathbf{1}_1 \otimes \cdots \otimes \mathbf{1}_{i-1} \otimes h_R \otimes \mathbf{1}_{i+1} \otimes \cdots \otimes \mathbf{1}_n$. Let us then consider the following task: Given fixed $\varrho_R, H_R, \beta, H_S$, find the minimum number of copies n so that it is possible to cool down the target state to inverse temperature β_S. Already the necessary condition for cooling (3.10) together with additivity of $\mathcal{V}_\beta$ poses lower bounds on the necessary number of copies $n^{\rm nec}(\beta_S)$ as

$$n^{\rm nec}(\beta_S) \geq \frac{\mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S)}{\mathcal{V}_\beta(\varrho_R, h_R)}. \tag{3.15}$$

We can now use the sufficient condition obtained from theorem 3.2

$$n[\mathcal{V}_\beta(\varrho_R, h_R) + K(\beta_S, \beta, \varrho_R, h_R, H_S)] \geq \mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S)$$

to also obtain a sufficient number of copies n^{suff}. This condition is sufficient, but not always necessary. We therefore obtain

$$n^{\text{suff}}(\beta_S) \leq \frac{\mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S)}{\mathcal{V}_\beta(\varrho_R, h_R) + K(\beta_S, \beta, \varrho_R, h_R, H_S)}. \tag{3.16}$$

Since the correction K goes to zero as $\beta_S \to \infty$ (the target temperature going to zero), we see that

$$\lim_{\beta_S \to \infty} \frac{n^{\text{suff}}(\beta_S)}{n^{\text{nec}}(\beta_S)} = 1. \tag{3.17}$$

It is then interesting to know how n^{nec} scales with the target temperature. I will now show that n^{nec} and n^{suff} scale as β_S for large β_S and calculate the prefactors.

In the special case of thermal resource and target states, we can reformulate the vacancy, and hence the necessary and sufficient number of copies, also in terms of non-equilibrium free energies. The vacancy of a thermal state at temperature β_S then takes the form

$$\mathcal{V}_\beta(\omega_{\beta_S}(H), H) = \beta_S \Delta F_{\beta_S}(\omega_\beta(H_S), H_S). \tag{3.18}$$

The necessary condition (3.10) then reads:

$$\Delta F_{\beta_R}(\omega_\beta(H_R), H_R) \geq \frac{\beta_S}{\beta_R} \Delta F_{\beta_S}(\omega_\beta(H_S), H_S). \tag{3.19}$$

From (3.18) we see that for large β_S we have (assuming vanishing ground state energy)

$$\mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S) = \beta_S E_\beta - S_\beta, \quad \text{as } \beta_S \to \infty. \tag{3.20}$$

Here, E_β and S_β denote the average thermal energy and entropy of the system S at inverse temperature β. Assuming, as above, a resource system of n non-interacting identical particles each described by $(\omega_{\beta_R}(h_R), h_R)$, we obtain that the minimum achievable temperature scales as

$$n k_B T_S = \frac{E_\beta^S}{\mathcal{V}_\beta(\omega_{\beta_R}(h_R), h_R)}, \quad T_S \ll 1. \tag{3.21}$$

This result is similar to the asymptotic result of Janzing et al. [46] in the special case of a target qubit. We thus see that the minimum target temperature scales as $1/n$ to leading order.

In applications related to quantum information processing, a figure of merit that is expected to be more important than the temperature is the final probability p to find the system in the ground state upon an energy measurement. The above scaling relation implies that p increases exponentially to 1 with n. For example, if the target system is a $d + 1$ dimensional system with gap Δ above a unique ground state, we have (for large n)

$$p \geq 1/(1 + d e^{-\beta_S \Delta}) \approx 1 - d e^{-n \mathcal{V}_\beta(\omega_{\beta_R}(h_R), h_R) \Delta / E_\beta^S}.$$

Thus, the optimal ground state probability converges exponentially to unity in the number of resource systems.

The above analysis fundamentally rests on the fact that the resource systems are uncorrelated. However, it seems reasonable to expect that similar results hold true in the case where the resource system is a many-body system with exponential decay of correlations in space. To rigorously proof this, one would need to show that the error term K also scales extensively with the system size if the subsystems are weakly correlated. I leave this problem for future work.

3.4   Thermal resources

While the setting of i.i.d.-resources is very relevant, our results allow us to consider settings far outside this regime. In general, the error term K in theorem 3.2 can be quite difficult to analyze in concrete set-ups. It would therefore be nice to know that it vanishes for general classes of resources. I now discuss general results that allow us to decide whether K vanishes by evaluating more physically accessible quantities. In the following, I always consider resources that are initially in a thermal state at inverse temperature β_R. In this case we can derive a general sufficient condition which ensures that K vanishes identically. This is based on showing that the function

$$\alpha \mapsto D_\alpha(\omega_{\beta_R}(H_R)\|\omega_\beta(H_R)) \tag{3.22}$$

is convex for a range of values of $\alpha < 1$, which implies that K vanishes at low enough target temperatures. It is possible to show convexity of (3.22) by checking whether the average energy

$$x \mapsto E_x^R = \mathrm{Tr}(H_R\omega_x(H_R)) \tag{3.23}$$

is convex as a function of the inverse temperature. In particular, if $\beta_R < \beta$ and the function $x \mapsto E_x^R$ is convex in $[\beta_R, \beta]$, the function $D_\alpha(\omega_{\beta_R}(H_R)\|\omega_\beta(H_R))$ is convex for all $\alpha < 1$.

Theorem 3.3. *For resources of the form $(\omega_{\beta_R}(H_R), H_R)$ that are hotter than the bath, that is with $\beta_R \leq \beta$, if $E_x^R = \mathrm{Tr}(\omega_x(H_R)H_R)$ is convex in the range $x \in [\beta_R, \beta]$, then (3.10) is a sufficient and necessary condition for low temperature cooling.*

Proof. The proof is given in section 14.2.3. □

This theorem drastically simplifies showing that the error-term K vanishes in many cases, since the average energy is a much more accessible quantity then the Rényi-divergences. In particular, we can formulate the condition on E_x^R in terms of the heat-capacity of the resource system. To do that, let us define the heat capacity $C_x := \frac{\mathrm{d}E_x^R}{\mathrm{d}T}$, with $T = 1/x$. Then the condition on E_x^R in theorem 3.3 is equivalent to

$$\frac{1}{\beta_R^2}C_{\beta_R} \geq \frac{1}{\beta'^2}C_{\beta'} \tag{3.24}$$

for all $\beta_R \leq \beta' \leq \beta$. Since in most thermodynamic systems the heat capacity increases monotonically with the temperature, this inequality is usually fulfilled for thermodynamic systems. However, it is not necessarily fulfilled for small systems with non-generic density of states. In the next section 3.4.1, I will discuss a large class of models for which the energy is indeed always convex.

Before we come to that, it is important to mention a different way to make sure that K vanishes. Indeed, for a fixed value of β_R, one can always find a critical value of β such that K vanishes if the environment is colder than the temperature given by that critical value. Thus, if an experimenter manages to pre-cool the environment to sufficiently low temperatures, the error term vanishes. This is proven in section 14.2.3. Unfortunately, however, it is not easy to give precise bounds on *how* cold the critical temperature for the heat bath is if we are given some resource. Finding such bounds which can be expressed in terms of meaningful physical quantities is an interesting open problem.

3.4.1   Systems for which the average energy is convex

I will now present a large class of models for which the average energy is a convex function of the inverse temperature. Thus, for all such resources (3.10) is both a necessary and sufficient condition. The examples are based on the following simple result (the proof is given in section 14.2.4).

Lemma 3.4 (Equidistant levels)**.** *Consider any Hamiltonian with equidistant and non-degenerate energy levels. Then the function $\beta \mapsto E_\beta$ is convex.*

Examples of systems with equidistant energy levels are given by two-level systems and harmonic oscillators. This seems at first like a very small class of systems. However, we can make use of the fact that the vacancy is unitarily invariant (does not depend on a choice of basis in the Hilbert-space) and additive over uncorrelated and non-interacting subsystems. Then the following corollary follows immediately from the existence of normal-mode decompositions in quasi-free systems:

Corollary 3.5. *The average thermal energy E_x^R is convex for any non-interacting (quadratic) fermionic or bosonic Hamiltonian. Thus, for any such resource with $\beta_R < \beta$, the condition (3.10) is both sufficient and necessary.*

While these models formally only cover fairly simple systems, they nevertheless can show interesting physical properties. For example, their thermal states can be arbitrarily correlated and even entangled. In fact, as is widely known, they can serve as effective models for many interesting physical effects, such as superconductivity or topological phases of matter [146]. Thus, we can expect that the condition (3.10) is both sufficient and necessary for large classes of physical resources. A particularly interesting resource that is included by these results is that of hot thermal light, which has been considered before as a valuable resource for cooling [135]. Also note that these results hold without any thermodynamic limit. They therefore apply far outside the asymptotic limit of i.i.d.-resources considered in Ref. [46].

Finally, many interacting many-body systems have a density of states which, in the bulk of the spectrum, can be well approximated by a Gaussian for large system sizes [147]. It is easy to check that if one considers an energy density of the form $\mu(\epsilon) \simeq e^{\gamma\epsilon - \alpha\epsilon^2}$, the average thermal energy of the corresponding Gibbs-state is convex as a function of the inverse temperature.

3.5   Exact and approximate catalysts

In this chapter, I have assumed catalytic thermal operations, where the catalyst has to be returned with an arbitrarily small error. This allowed us to consider only Rényi-entropies for positive α. In general, this assumption is very reasonable, since in practice it is impossible to distinguish between arbitrarily small errors and zero errors. However, as explained in section 2.3, below theorem 2.1, to reduce the sufficient conditions to positive α one formally introduces a catalyst that does not have full rank but is returned with full rank. Therefore the change of vacancy on this catalyst is infinite. It is thus desirable to develop an understanding about whether demanding absolutely exact catalysts changes the picture. First note that only the sufficient condition can change by changing this assumption since the necessary assumption was derived from the monotonicity of Rényi-divergences with positive α, which holds in both cases.

Regarding the sufficient condition, it turns out that one can indeed prove a similar theorem as Theorem 3.2:

Theorem 3.6 (Sufficient condition for absolutely exact catalysts)**.** *Assume thermal operations with exact catalysts. Then for every choice of β and H_S there is a critical $\beta_{cr} > 0$ such that for any $\beta_S > \beta_{cr}$ and full-rank resource (ρ_R, H_R) (diagonal in the energy eigenbasis) the condition*

$$\mathcal{V}_\beta(\rho_R, H_R) - K(\beta_S, \beta, \rho_R, H_R, H_S) \geq r(\beta, H_S)\mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S) \qquad (3.25)$$

is sufficient for cooling. The positive semi-definite function K is identical to that in theorem 3.2 and the constant $r(\beta, H_S)$ is independent of ρ_R, H_R and β_S and given by

$$r(\beta, H_S) = 1 + 2\frac{E_{\max} - E_\beta}{E_\beta}, \qquad (3.26)$$

where $E_{\max}$ is the largest eigenvalue of H_S and we assume that the ground state energy of H_S is zero.

Proof. The proof is given in section 14.2.5. □

The implication of this theorem is most easily understood in the case of n independent resource systems. In this case it simply says that instead of the $n^{\text{suff}}(\beta_S)$ resource systems that are sufficient to cool a system to temperature β_S using catalytic thermal operations given in (3.16), in the case of exactly catalytic operations one requires at most $r(\beta, H_S)n^{\text{suff}}(\beta_S)$ of the same resource systems. Note that the factor $r(\beta, H_S)$ is independent of the target-temperature and the resource. While quantitatively, this result is different, qualitatively the picture does not change much: The vacancy still determines the resource scaling for approximate ground state cooling. Furthermore, there is good reason to believe that the factor $r(\beta, H_S)$ can be made much closer to unity by more elaborate proof techniques.

One might be worried that also in the case of exact catalysis an infinite vacancy might be required on the catalyst. This is not true and can be seen by the following argument: Let σ_C denote the state of the catalyst in a given process. Then, in the case of exact catalysis, the unitary that implements the process has to map the subspace $\mathscr{H}_{BRS} \otimes \operatorname{supp}(\sigma_C)$ into itself. The result on the system S is then completely unchanged if we simply consider as catalyst only the state σ_C restricted to its support, which is then a state with full rank and hence a finite vacancy.

A further natural question to ask is what happens if we go in the opposite direction and allow for a fixed finite error in the catalyst instead of an arbitrarily small error. In this case, the first observation to make is that it is then crucial how we measure this error. It has been shown in Ref. [49] that if we allow the state of the catalyst to change by a finite amount ϵ as measured by the trace-distance, then no matter how small we fix this error, one can in fact prepare *any state* with such an approximately catalytic thermal operation. This is due to the fact that the size of the catalyst is not bounded. It then would be sensible to allow the error on the catalyst in terms of trace-distance to be of the order $\epsilon / \log(d_C)$, where d_C is the Hilbert-space dimension of the catalyst. In this case one finds that one can prepare any state ρ from a resource ρ_R to arbitrary precision if the free energy of ρ is smaller than that of ρ_R. This would imply that one can cool a system to arbitrarily small temperatures with finite resources and hence would imply, in a sense, a violation of the unattainability principle in terms of resources.

In Ref. [49], the authors also suggest a potential solution to this problem. The idea is that the error on the catalyst should be measured by the natural measure associated to the operational task at hand. In our example, this would suggest to allow that the vacancy of the catalyst changes by at most ϵ. If we adopt this notion of approximate catalyst, we see that the necessary condition for cooling is indeed stable: If the initial state of the catalyst is ρ_C and the final state is ρ'_C, we obtain the necessary condition

$$\mathcal{V}_\beta(\rho_R, H_R) \geq \mathcal{V}_\beta(\rho_S, H_S) + (\mathcal{V}_\beta(\rho'_C, H_C) - \mathcal{V}_\beta(\rho_C, H_C)), \qquad (3.27)$$

where ρ_R and ρ_S denote the initial state of the resource and final state of the target, respectively. Thus, $\mathcal{V}_\beta(\rho_R, H_R) \geq \mathcal{V}_\beta(\rho_S, H_S) \pm \epsilon$. Regarding the sufficient condition, note that allowing for a change of ϵ in the vacancy of the catalyst in principle allows for much more freedom. It is therefore conceivable that the sufficient condition in this case exactly simplifies to $\mathcal{V}_\beta(\rho_R, H_R) \geq \mathcal{V}_\beta(\rho_S, H_S)$. Proving this statement would potentially allow us to get rid of all the assumptions on the resource system. It constitutes a very interesting open problem for future work to try to prove this relation.

3.6   *Coming back to Landauer erasure and energy measurements*

In chapter 2 we derived the *necessary* condition

$$\Delta F_\beta(\rho_R, H_R) - \Delta F_\beta(\rho'_R, H_R) \geq \Delta F_\beta(\,|0\rangle\langle 0|\,, H_S) \qquad (3.28)$$

for resetting a system S to $|0\rangle$, independent of the initial condition and using access to a heat bath at inverse temperature β. Here, the system R serves as a thermodynamic resource, just as in this chapter. From a formal point of view, the problem of resetting a qubit into an energy-eigenstate and cooling the qubit to the ground state are completely equivalent. Hence, a further necessary condition for information erasure is provided by the vacancy

and we see that *perfect* erasure without side information is in fact *impossible* under the assumptions of our framework. As shown in section 3.3, approximate erasure can in fact be done with an exponentially small error, explaining why this result is not in conflict with the obvious fact that in practice we can erase bits. However, for small error probabilities, erasure is governed not by the free energy, but by the vacancy.

How is the setting where we have side-information compatible with this statement? Consider the setting of a qubit that is correlated with a memory system M:

$$\rho_{SM} = p \, |00\rangle\langle00| + (1 - p) \, |11\rangle\langle11| \tag{3.29}$$

and assume that $H_{SM} = 0$. In this case, we can erase the system S using the conditional unitary

$$U_{\text{erasure}} = \mathbf{1} \otimes |0\rangle\langle0| + \sigma^x \otimes |1\rangle\langle1|. \tag{3.30}$$

In the process, the state on M is unaltered. Similar conclusions hold if the initial state on SM is an entangled state, although in this case the state on M changes in the process. From our necessary condition in terms of the vacancy, we then have

$$\mathcal{V}_\beta(\rho_{SM}, H_{SM}) \geq \mathcal{V}_\beta(|0\rangle\langle0| \otimes \rho_M) = \mathcal{V}_\beta(|0\rangle\langle0|, H_S) + \mathcal{V}_\beta(\rho_M, H_M) = +\infty.$$

Thus, the state of SM needs to have infinite vacancy if it allows for erasure of S, as expected. Indeed this can be seen explicitly by noting that ρ_{SM} cannot have full rank in such a case.

Finally, let me discuss briefly an apparently obvious way to cool a system to the ground state from an abstract point of view. First note that the sufficient condition (3.3) allows to map any energy-eigenstate to the ground state by a (catalytic) thermal operation. It would thus seem that we could simply perform an energy-measurement and bring the system to the ground state by a thermal operation that depends on the measurement outcome. However, from a thermodynamic perspective, energy measurements are very costly operations and cannot be done as thermal operations. Suppose to the contrary that we *could* do energy-measurements "for free" by somehow utilizing the thermal environment in an energy-preserving and entropy-preserving way. If we would perform a projective measurement (with post-selection) of the energy of a thermal system, the system would be mapped to an eigenstate $|E_i\rangle$ and we would be sure about the system's state after the measurement (by the definition of a measurement). We could then use a thermal operation to extract an amount $\Delta F_\beta(|E_i\rangle\langle E_i|, H_S) > 0$ of *work* into an auxiliary system (see chapter 9 to see concrete ways to do this). After this operation, the system would again be in thermal equilibrium and we could repeat this process. Since the extracted work would be positive in every run of the process, we would cyclically extract positive work form a heat bath – this is clearly impossible by the second law of thermodynamics.

Using the vacancy, we can understand this more clearly: The vacancy on the system S changes by an infinite amount upon the energy-measurement. Thus the vacancy of an ideal measurement device has to be infinite. This can be understood by observing that an ideal measurement device requires an empty memory or pointer in a perfectly known, deterministic state to record the measurement outcome. This requires that the initial state of the measurement device does not have full rank and hence infinite vacancy. Indeed, once such a state $|0\rangle$ is available, an energy-measurement and later erasure can be done in the following way: If ρ_S is the initial state, we can first use a thermal operation to dephase ρ_S in the energy-eigenbasis. This is always possible using a thermal operation. Once ρ_S is diagonal, we can "copy" it into the register of the measurement device using the unitary

$$U_{\text{measure}} = \sum_i |E_i\rangle\langle E_i| \otimes V_i, \tag{3.31}$$

where V_i are unitaries such that $V_i \, |0\rangle = |i\rangle$ are ortho-normal states. This produces a state of the form (3.29) and we can use the procedure explained above to erase the system. The overall outcome of this procedure is that we simply swapped the known state $|0\rangle$ from the memory with the dephased version of the state ρ_S on the system – an operation which is obviously possible using a thermal operation directly.

3.7   Conclusion and outlook

In this chapter, I have shown how to derive novel necessary and sufficient conditions for cooling a system close to the ground state. The conditions are expressed in terms of a new monotone of catalytic thermal operations, called vacancy, and relate the amount of non-equilibrium resources (measured in terms of the vacancy) to the vacancy of the system to be cooled in its target state. Since the vacancy diverges as the target temperature approaches zero, the unattainability principle in terms of non-equilibrium resources is recovered naturally.

Importantly, for a large class of resource states, the necessary and sufficient condition coincide. This shows that the vacancy is the unique measure of non-equilibrium that controls the problem of ground state-cooling. However, there remain many open questions. Firstly, it can be expected that it can be proven that the necessary and sufficient condition coincide for a much larger class of systems then shown here by introducing more concrete physical assumptions. Secondly, it is necessary to study the scaling of the vacancy in more detail under concrete physical assumptions on the non-equilibrium resources. For example, it should be investigated if the vacancy is extensive for any resource controlled by a local Hamiltonian and a state that fulfills clustering of correlations. There are good reasons to believe that this is true. As an example, if the resource is a thermal resource at inverse temperature β_R, then the relation

$$\mathcal{V}_\beta(\omega_{\beta_R}(H_R), H_R) = \Delta F_{\beta_R}(\omega_\beta(H_R), H_R) = E_\beta - E_{\beta_R} - \frac{1}{\beta_R}(S_\beta - S_{\beta_R}) \qquad (3.32)$$

shows that the vacancy is extensive if the energy and entropy are extensive for both the temperatures β and β_R and the different terms do not cancel. Nevertheless, a general proof that the vacancy is an extensive quantity remains as an open problem.

The strategy that we used in this chapter to study the problem of ground state cooling is a quite general one in any resource-theoretic approach: One simply tries to find the right monotone that describes the problem at hand. In chapter 8, we use the same strategy to study the problem of equilibration of many-body systems from the resource theoretic point of view.

In many resource theories, a natural family of monotones is provided by Rényi-divergences of the form $D_\alpha(\sigma\|\tau)$, where τ is a "free state" that is associated to the resource theory. For example, in the resource theory of asymmetry [107–109, 111], τ are symmetric states. In this chapter, we have seen that strengthened conditions on state-transitions can be derived by also considering monotones of the form $D_\alpha(\tau\|\rho)$ (provided that they are well-defined). In particular, $D(\tau\|\rho)$ controls the behaviour of the monotones $D_\alpha(\rho\|\tau)$ for small α. We can expect that this quantity becomes important in physical processes where ρ is close to a pure state, but τ is not. As a further example, a standard measure of correlations in quantum information theory is given by the mutual information [62]

$$I(A:B)_{\rho_{AB}} := D(\rho_{AB}\|\rho_A \otimes \rho_B), \qquad (3.33)$$

as it cannot be increased by local operations. In terms of the resource theory of local operations, states of the form $\rho_A \otimes \rho_B$ are free states since they are uncorrelated. However, any "Rényi mutual information" $D_\alpha(\rho_{AB}\|\rho_A \otimes \rho_B)$ also provides a measure for correlations since it cannot be increased under local operations. Then for small α these quantities can be expected to be controlled by $D(\rho_A \otimes \rho_B\|\rho_{AB})$, which indeed also cannot be increased by local operations and thus provides a correlation measure. In chapter 11, we will find that a quantity similar to this quantity plays an important role in thermodynamics in the strong-coupling regime. It is an interesting open problem to study such "reversed relative entropies" in the context of other resource theories than thermodynamics in more detail.

4 What is work?

IT IS VERY SIMPLE to check whether a given macroscopic machine is useful from a thermodynamic point of view: We just see whether we can somehow make it lift a weight. We might have to add some extra gear, but in principle it is always possible if the machine is able to do work on some system in some sense. If we consider smaller and smaller machines, at some point this will no longer be true. First, the energy scale at which the machine operates will at some point be no longer comparable with that of a macroscopic body in a gravitational potential. Second, at some point energy will be a fluctuating quantity, first due to thermal noise and ultimately due to quantum mechanical fluctuations.

For example, while it is quite clear that a biological cell in a human being must do something akin to "thermodynamic work" to produce a gradient in the concentration of certain molecules, which in turn makes sure that the transcription (copying) of (parts of) the DNA into messenger RNA occurs in the right direction[1], it is not so easy to connect this with raising a weight in a gravitational potential. Once we get to the quantum scale, where only few atoms are involved, and which is happening in experiments nowadays, it becomes even less clear when we can claim that a quantum thermal machine does something thermodynamically "useful" and how to quantify this.

In this chapter [2], I want to discuss the problem of how one can possibly define a notion of "thermodynamic work" in processes where it is not obvious how to isolate a deterministic degree of freedom in a potential, but which occur in a thermal environment of some temperature. The copying of information in the example above, or the erasure of information like in Landauer erasure (see chapter 2) are examples of such processes. The results presented in this chapter will not only be useful in the microscopic domain, but in principle always when it is not obvious how to single out "work-like" degrees of freedom. For example, the process of changing the state in some system from an energy-eigenstate to a coherent super-position of two energy-eigenstates is a very important and useful basic operation in a quantum computer, but is difficult to connect to the classical definition of mechanical or chemical work. It is thus difficult to associate a "thermodynamic cost" with such a process. In particular it is impossible to associate a fluctuating work-variable in the sense of fluctuation relations to such a process [79].

We will see that it is indeed possible in great generality to find quantifiers for such thermodynamic usefulness, and, importantly, these can be derived from very general principles, which in fact do not need to make reference to the underlying precise physical theory. They thus can be applied to the case of quantum mechanics, but can also be applied to macroscopic systems, recovering the notion of work as energy stored in a deterministic degree of

[1] See, for example, Refs. [35, 148, 149] for a detailed discussion of this process from a thermodynamic point of view and its relation to reversible computing.

[2] This chapter is based on work together with Rodrigo Gallego and Jens Eisert, published in [2].

Figure 4.1: A work-storage device A with a unique ground state and a g-fold degenerate excited level undergoing a transition which results in a deterministic increase of energy, but an increase in entropy. Such a protocol exists for every value of Δ and β if g is large enough. (Figure taken from Ref. [2].)
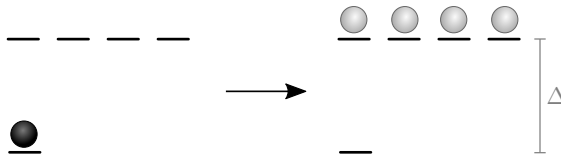
freedom under the influence of a conservative force.

Before explaining our approach to this problem, let me first explain in an example and in more detail why it is difficult to define work in the quantum regime and why even a deterministic change of energy cannot always be considered as thermodynamic work.

4.1   Deterministic changes of energy as heat

Let us consider a simple Gedankenexperiment. Suppose you had some small quantum mechanical system, initially in the ground state with energy zero. Then you hand it to a friend and some time later she returns it. Furthermore she promises that its probability distribution of energy is now completely deterministic, that is, it is completely concentrated on a single energy-eigenvalue Δ. In such a situation, it would be tempting to say that from a thermodynamic point of view, your friend had performed an amount of work given by Δ.

I will now argue why this need not be so. Indeed, there are situations of this form where all the change of energy should be attributed to heat and not work. The argument is very simple. Suppose your system only has two energy levels. Assume that the ground state is unique, but the excited state has a large degeneracy $g \gg 1$ (see Fig. 4.1). In such a case, if your friend simply lets the system thermalize with the environment at temperature β, the system is almost surely in the excited state at the end. The probability is given by $ge^{-\beta\Delta}/(1 + ge^{-\beta\Delta})$, which can be made as close to 1 as one wishes by increasing g.

In fact one can achieve more. I will now show that by acting with an energy-preserving unitary on the system and heat-bath together (and possibly a catalyst), one can remove all uncertainty in energy from the system and put it into a state with a deterministic energy-distribution supported on the excited energy-level.

To do that we will use theorem 2.1 and the fact that the Rényi-divergences are monotonic in α [69]: $D_\alpha(\rho\|\sigma) \geq D_{\alpha'}(\rho\|\sigma)$ if $\alpha \geq \alpha' \geq 0$. If the final state is ρ', the theorem thus tells us that it is sufficient to check that $D_0(\,|0\rangle\langle0|\,\|\omega_\beta(H_S)) > D_\infty(\rho'\|\omega_\beta(H_S))$. A simple calculation shows that

$$D_0(\,|0\rangle\langle0|\,\|\omega_\beta(H_S)) = \log(Z_\beta),$$

$$D_\infty(\rho'\|\omega_\beta(H_S)) = \lim_{\alpha\to\infty} \frac{1}{\alpha - 1} \log\left(g^{1-\alpha}\frac{e^{-\beta\Delta(1-\alpha)}}{Z_\beta^{1-\alpha}}\right)$$

$$= \log\left(\frac{e^{\beta\Delta}Z_\beta}{g}\right), \tag{4.1}$$

with $Z_\beta = 1 + ge^{-\beta\Delta}$. This shows that as long as $g > e^{\beta\Delta}$, the transition is possible exactly. At the same time the Shannon entropy of the system increases from 0 to $\log(g) > \beta\Delta$. Thus we have $k_B T\Delta S > \Delta$. In particular the free energy decreases, as it has to.

This simple example shows that for small quantum systems, even a purely deterministic increase of energy cannot always be accounted for as "work" as it can sometimes in principle be reliably achieved by only having access to a heat bath. Of course, the example is very special in that it assumes a small system with a very particular density of states. But we cannot in general rule out such density of states for small individual systems.

We should therefore see the example as a motivation to explore new ways of how we should quantify work. In the following sections, we will present exactly such an attempt which is based on operational principles. The basic guiding principle that we will have in the back of our mind is the following, which we already implicitly used above:

Consider a machine inducing a transition on some system A. If such transition could have been also done by having the machine only using systems in thermal equilibrium with the

*environment and leaving any other system involved in the process in the same state as it was at
the beginning; then we conclude that the machined did not induce a thermodynamically useful
transition on A. In this case, we say that the machine did not perform positive work on A.*

This statement paraphrases the second law of thermodynamics, which states that a machine
cannot perform positive work in a cyclic process (leaving any other system in the same
state in the formulation above) by using a single heat bath (systems in thermal equilibrium
with the environment). We will thus essentially use the second law of thermodynamics to
define what we mean by "work" instead of seeing the second law of thermodynamics as a
constraint on possible processes in terms of a pre-defined notion of what "work" is.

4.2   *Thermodynamic tasks and work-storage devices*

Motivated by the previous discussions and the example in the last section, I will now present
an operational approach to quantify the "thermodynamic usefulness" of state transitions
that a machine implements. Here, I will use the word machine in a very general sense
as a system that acts (or is made to act) on a different system and thus induces a state-
transition on this additional system. For example, in macroscopic thermodynamics, the
machine might perform mechanical work on a suspended weight, it might charge a battery
or it might cool down or heat up some system. In the following, I will, to simplify the
language, in all such cases say that the machine performs *work*. Similarly, I will call the
system that the machine acts upon (e.g., the weight or the battery) a *work-storage device*.

I will use Latin letters $p = (\rho, H)$ to describe the pair of state and Hamiltonian of
systems [3] and denote a state-transition by an arrow such as $p \to q$. Furthermore, I will call
the pairs *p objects* to make clear that they contain more information than just the state of
the system.

Typically, we will be interested only in transitions between objects in a given set of ob-
jects which encode an *operational task*. We will call this set $\mathcal{P}$. In the *macroscopic* world,
the set $\mathcal{P}$ would include arbitrary states with a well-defined temperature with respect to the
Hamiltonian when we are interested in the task of cooing/heating. Or we could be inter-
ested in the task of performing mechanical work, in which case the set $\mathcal{P}$ would include, for
example, deterministic states of mechanical systems in a gravitational or electric potential.

For microscopic systems, it is less clear what $\mathcal{P}$ should look like. First, clearly one can
incorporate microscopic analogues of tasks such as cooling/heating or performing mechan-
ical work in this framework. As an example, recently, the so-called *single-shot* approach
to work extraction in small quantum systems has received a lot of attention [47, 49, 50, 52,
60, 150–153]. Here, the idea is that one only allows for processes where the fluctuations in
the extracted work are arbitrarily small. One can capture this by saying that the probability
that the extracted work (which is here defined as a transition between energy-eigenstates)
deviates by more than δ from a given value is less than ϵ. In the framework of this chapter,
this can be captured by letting $\mathcal{P}$ only contain states of the form

$$\rho_\delta^\epsilon(E) := p_\epsilon \sigma + (1 - p_\epsilon)\rho_\delta(E), \tag{4.2}$$

where $p_\epsilon < \epsilon$ and $\rho_\delta(E)$ is supported on energy-eigenstates whose energy deviates at
most δ from the energy E. Since in our framework thermodynamic protocols have to map
states from $\mathcal{P}$ into states from $\mathcal{P}$, such a choice constitutes a model for single-shot work-
extraction as discussed above.

Nevertheless, as argued above, tasks such as work-extraction will not necessarily be the
important task for a microscopic machine. In the following let us hence leave the set $\mathcal{P}$
completely unspecified. All the results will either hold for arbitrary such sets or will be a
function of this set.

Hence, our task now is to quantify, from a thermodynamic perspective, the value of a
transition $p \to q \in \mathcal{P}$ in an environment of inverse temperature β.

4.3   *An operational approach*

To introduce our approach to the problem of quantifying work, I will use a language that
might be unfamiliar from the usual physics literature. It is inspired from computer science

[3] It will be clear that additional
information such as additional
conserved quantities can be in-
corporated in this framework.
For simplicity I leave them out
of the discussion.

and tries to describe a problem as an interaction between two agents, one of which, called Merlin, is very powerful and the other one, called Arthur, is limited in his capabilities. I want to stress, however, that the language captures the usual thermodynamic setting and bounds the very same quantities usually under consideration in thermodynamics. This will become clear in the later sections.

Consider the two players, Arthur and Merlin living in a world of inverse temperature β. They both agree upon the fact that the environment has inverse temperature β. They also agree that this fact implies the existence of a set of "(catalytic) free transitions", i.e., a set of transitions on physical systems that can be implemented without additional resources. This should be understood in the sense of chapter 2, but in fact all our results will be a function of the choice of model for the free transitions and therefore other models than thermal operations can be considered as well (in section 14.3 I discuss in detail which are the minimal properties of the free operations that are explicitly needed to derive our results). For the purpose of this chapter, let us introduce the following notation. We will denote by $\mathcal{F}(p)$ the set of objects that can be reached from p by any free transition. Similarly, we will denote by $\mathcal{F}_C(p)$ the set of objects that can be reached by a catalytic free transition. In mathematical terms, this means that $q \in \mathcal{F}_C(p)$ whenever there is a catalyst c such that $q \otimes c \in \mathcal{F}(p \otimes c)$. Here, we introduced the notation

$$(\rho, H) \otimes (\sigma, K) := (\rho \otimes \sigma, H \otimes \mathbf{1} + \mathbf{1} \otimes K). \tag{4.3}$$

Merlin now claims to own a machine which can perform thermodynamically useful tasks and trades quantum systems for a living. Arthur is one of his customers whose system takes the role of the work-storage device.

We then imagine that Arthur hands to Merlin his system described by $p = (\rho, H) \in \mathcal{P}$ and leaves again. Later, Arthur comes back and Merlin returns the system described by a different object $q \in \mathcal{P}$ (along with the classical description of the state). The new state might be more or less useful for Arthur. In general Arthur will have to pay some money to Merlin (or vice-versa if Merlin made the system less useful for Arthur). The two thus have to find a fair agreement on how they value such state-transitions. We will denote this price by

$$\mathcal{W}(p \to q, \beta) \in \mathbb{R}. \tag{4.4}$$

Note that $\mathcal{W}$ is a function that associates a real number only to a transition between valid work-storage devices, that is objects in $\mathcal{P}$. No price is associated if Merlin does not return a valid work-storage device. We will make the convention that $\mathcal{W} \geq 0$ if Arthur has to pay money to Merlin and will refer to $\mathcal{W}$ as a *work quantifier*.

Importantly, Arthur does not know how exactly Merlin performed the transition and it is completely unimportant to him as he is only interested in what happens to his system. This is in analogy to the case of mechanical work or cooling: One can check whether (and how much) work a machine performs without knowing the internal details of the machine by simply looking at how the machine acts on auxiliary systems. It is important to understand that this is not in contradiction with the fact that thermodynamic work is a path-dependent quantity from the viewpoint of the machine, as is explained in Fig. 4.2.

As stated above, Arthur and Merlin have to agree on a "fair" agreement on the work-quantifier $\mathcal{W}$. Such an agreement is of course subjective in principle and hence has to be assumed. The rest of this chapter will be devoted to establishing natural and mathematically precise Axioms that any valid work-quantifier should fulfill and derive from these Axioms a series of properties for any such work-quantifier. Before we come to the precise Axioms, however, we have to introduce a few more important concepts.

4.3.1   Work of transition, work cost and work value

The work-quantifier $\mathcal{W}$ evaluates transitions on Arthur's work-storage device. It is easy to confuse this quantity with other quantities in thermodynamics. After having defined $\mathcal{W}$ one is often interested in the optimal transition (as measured by $\mathcal{W}$) that one can implement on a work-storage device given some non-equilibrium resource $p_M^{(i)}$. In other words, suppose that Merlin has, inside of his lab, a system described by the object $p_M^{(i)}$ but no other
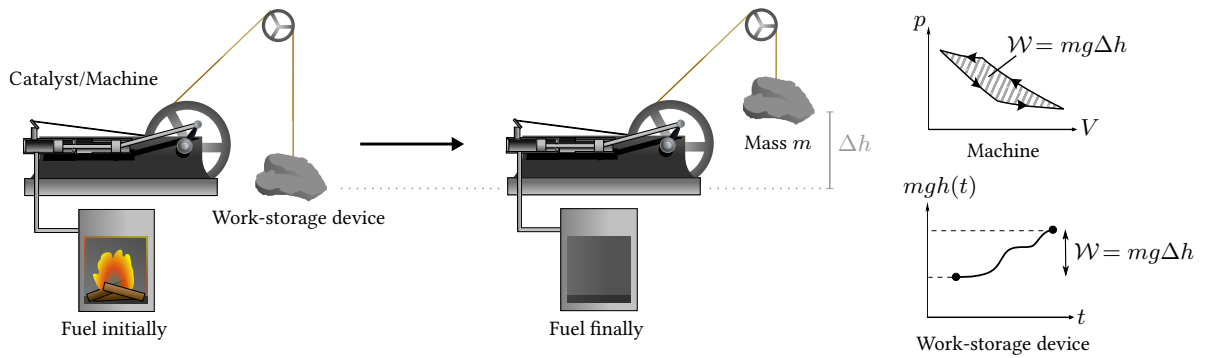
Figure 4.2: *Left:* Phenomenological analogy of the setting in the case of a machine that burns fuel to perform mechanical work. The catalyst corresponds to the machine that returns to its initial state, using up burning fuel to lift a weight. The burning fuel corresponds to Merlin's non-equilibrium system and the lifted weight corresponds to Arthur's work-storage device. *Right:* Work from two points of view. Path-dependent work obtained by looking at the time-dependent thermodynamic state of the thermal machine at the top and operational path-independent work obtained by looking at the weight (work-storage device) at the bottom. All processes happen at some background-temperature T. The work of transition W_{trans} of the fuel corresponds to the maximal height that the weight can be lifted by arbitrary machines leaving the fuel in the corresponding final state and operating at background-temperature T. (Figure from Ref. [2].)

resources. Then it is interesting to know how much money he can earn if a customer comes with a work-storage device. This corresponds to what is often called the "extractable work" or "work value" in the literature [47, 49, 50, 70, 150]. Similarly, we could not only fix the initial, but also the final object $p_M^{(f)}$ and ask how much Merlin can in principle earn from a transition between the two states. We will call this quantity the *work of transition* [47].

Definition 4.1 (Work of transition). *Given a work-quantifier $\mathcal{W}$, an inverse temperature β, a class of work-storage devices $\mathcal{P}$, and initial and final objects $p_M^{(i)}$ and $p_M^{(f)}$, the* work of transition *is defined as*

$$W_{\text{trans}}(p_M^{(i)} \to p_M^{(f)}, \beta) := \sup_{\substack{p_A^{(i)}, p_A^{(f)} \in \mathcal{P}; \\ p_M^{(f)} \otimes p_A^{(f)} \in \mathcal{F}_C(p_M^{(i)} \otimes p_A^{(i)})}} \mathcal{W}(p_A^{(i)} \to p_A^{(f)}, \beta). \qquad (4.5)$$

According to our convention, if the work of transition is positive, then Merlin receives money from Arthur. It is important to realize that the work of transition depends on all possible catalytic free transitions with all possible work-storage devices that can be implemented in such a way that the transition on Merlins non-equilibrium resource matches the given one. In contrast to $\mathcal{W}$, the work of transition W_{trans} is *not* defined on the work-storage device, but on the transition on Merlin's system. Its input objects thus do *not* have to be in $\mathcal{P}$. In fact the work (as measured by $\mathcal{W}$ on the work-storage device) associated to some process can never be deduced from the initial and final state on M alone. Thus it is impossible to define $\mathcal{W}$ as a function on transitions $p_M^{(i)} \to p_M^{(f)}$. One either needs to specify the precise process including the work-storage device or, if one has only access to M, consider the optimal possible value as in (4.5). It is precisely in this sense that work, as a function of transitions on M, is a path dependent quantity when evaluated in transitions on M, and a path-independent quantity when evaluated in transitions on A. This is also the case in phenomenological thermodynamics: work can be specified by knowing only the initial and final height of the lifted weight, however it is path-dependent as function of the state of the machine (see Fig. 4.2).

After having defined the work of transition, we can also define the *work value* and the *work cost* as the maximum possible value that Merlin can obtain from p_M and the minimum possible value that he has to pay to be able to create p_M, respectively:

$$W_{\text{value}}(p_M, \beta) := W_{\text{trans}}(p_M \to \omega_\beta, \beta), \qquad (4.6)$$

$$W_{\text{cost}}(p_M, \beta) := -W_{\text{trans}}(w_\beta \to p_M, \beta), \qquad (4.7)$$

where w_β denotes an object representing a system in thermal equilibrium. While both these quantities are important in thermodynamics, it is clear that they can only be defined once a work-quantifier $\mathcal{W}$ has been established, which is the primary concern in this chapter. We will however see, that our Axioms imply $W_{\text{value}} \leq W_{\text{cost}}$, which can be seen as an expression of the second law of thermodynamics.

Finally, note that in (4.5), it is assumed that the work-storage device and the system of Merlin are uncorrelated both in the beginning and the end of the transition. This is in fact important and in section 4.7 we will discuss in detail the role of correlations between M and A.

4.4   The Axioms

We are finally in position to formulate and discuss the two Axioms that will be the basis for the rest of the analysis. While they might seem at first technical, they are very natural when viewed in the language of Arthur and Merlin. In that language they simply say that neither Arthur nor Merlin can get arbitrarily rich without spending resources. However, below I will also discuss the meaning and implications in more physical terms.

Axiom 1 (Cyclic transitions of the work-storage device). *For any cyclic sequence of transitions of the work-storage device* $p_A^{(1)} \to p_A^{(2)} \to \cdots \to p_A^{(n)} = p_A^{(1)}$, *such that* $p^{(i)} \in \mathcal{P}\ \forall i$, *the sum of the work-values of the individual transitions is larger than or equal to zero,*

$$\sum_{i=1}^{n-1} \mathcal{W}(p_A^{(i)} \to p_A^{(i+1)}, \beta) \geq 0. \tag{4.8}$$

Remembering that a positive value of $\mathcal{W}$ means that Arthur pays money to Merlin, the previous axiom ensures for example, in the simplest case, that Arthur cannot get rich by demanding Merlin to first do a transition $p^{(1)} \to p^{(2)}$ and then asking from him to undo the transition. If this principle was violated, Arthur could get infinitely rich just by repeatedly interacting with Merlin. Note that since it is always Merlin that actually implements transitions in our framework, Arthur, by definition, never spends any resources.

Axiom 2 (Cyclic transitions of the fuel). *For any cyclic sequence of transitions of the fuel (Merlin's system)* $p_M^{(1)} \to p_M^{(2)} \to \cdots \to p_M^{(n)} = p_M^{(1)}$, *the sum of the optimal work that Merlin can obtain in each sequence (this is given by* W_{trans} *in (4.5)) is smaller or equal to zero,*

$$\sum_{i=1}^{n-1} W_{\text{trans}}(p_M^{(i)} \to p_M^{(i+1)}, \beta) \leq 0. \tag{4.9}$$

This second Axioms says, again in the simplest case where $n = 2$, that Merlin cannot get money from Arthur if he simply does not use up any resource, for in that case $W_{\text{trans}} \leq 0$. Thus also Merlin cannot get rich without "burning" resources. While the two Axioms seems completely symmetric, it is important to note that the first Axioms is concerned with the work-quantifier $\mathcal{W}$, which is only evaluated on valid work-storage devices, while the second Axiom is concerned with the function W_{trans}, which can be evaluated on all objects[4].

Summarizing, the two Axioms together provide a precise formulation of the second law of thermodynamics in terms of work. As noted before, however, unlike its usual interpretation as a restriction on possible processes with a pre-defined notion of work, here we use the Second Law of Thermodynamics to figure out what could be a sensible measure of work. In this sense it is *impossible* to violate the Second Law in terms of work in our set-up: If somebody claimed to violate the Second Law by extracting work in a cyclic proccess from a single heat bath, whatever the person considered as work would not be accounted for as work in this approach.

[4] Of course, W_{trans} implicitly also depends on $\mathcal{P}$, since it depends on $\mathcal{W}$.

4.5   Functions that fulfill the Axioms

After having stated the two Axioms, it is time to discuss what kind of functions constitute valid work-quantifiers, i.e., fulfill the Axioms. From the generality of the Axioms, it may seem that they should not provide strong conditions on possible work-quantifiers. In the following, we will see however that they in fact provide strong conditions. Under certain additional assumptions they in fact allow to essentially uniquely specify possible work-quantifiers.

In the following chapters, I will try to keep the discussion as informal as possible, deferring essentially all technical proofs to section 14.3. Nonetheless, some level of technicality is necessary to state and discuss the results.

In the following, we assume that not having any work-storage device at all constitutes a valid work-storage device as well. From the point of view of Merlin and Arthur, this just means that Arthur does not have any system, for example because he handed it to Merlin. Formally, we introduce the *empty* object $\varnothing$. It has the properties $p \otimes \varnothing = \varnothing \otimes p = p$ for all $p \in \mathcal{P}$ [5]. Also recall that $\mathcal{F}(p)$ and $\mathcal{F}_C(p)$ denote the set of states that can be reached from p by free and catalytic free transitions, respectively. It is then possible to prove the following theorem.

Theorem 4.2 (Form of work quantifiers). *A function $\mathcal{W}$ respects Axioms 1 and 2 if and only if it can be written as*

$$\mathcal{W}(p \to q) = M(q) - M(p), \tag{4.10}$$

for a function M such that $M(\varnothing) = 0$ and that fulfills the property of additive monotonicity: *For all $p^{(1)}, \ldots, p^{(m)}$ and $q^{(1)}, \ldots, q^{(m)}$ in $\mathcal{P}$ such that $\bigotimes_{i=1}^m q^{(i)} \in \mathcal{F}_C(\bigotimes_{i=1}^m p^{(i)})$*

$$\sum_{i=1}^m M(q^{(i)}) \le \sum_{i=1}^m M(p^{(i)}). \tag{4.11}$$

The above result has many implications. First, it implies that the work, *as measured on the work-storage device*, is a path-independent quantity, in the sense that

$$\mathcal{W}(r \to p) + \mathcal{W}(p \to q) = \mathcal{W}(r \to q), \tag{4.12}$$
$$\mathcal{W}(p \to q) = -\mathcal{W}(q \to p). \tag{4.13}$$

Since this might seem at first surprising, given that work is usually thought of as a path-dependent quantity, it is important to emphasize that this by no means implies that the function W_{trans} also fulfills properties 4.12 and 4.13. In particular, Eqs. (4.12) and (4.13) are perfectly compatible with *irreversibility*, i.e., the fact that $W_{\text{value}}(p) < W_{\text{cost}}(p)$.

Second, the theorem implies that no work can be done without investing resources:

$$\mathcal{W}(p \to q) \le 0, \quad \text{if} \quad q \in \mathcal{F}_C(p). \tag{4.14}$$

In other words, this means that the function M appearing in $\mathcal{W}$ is a *monotone* of the free catalytic transitions, in the sense of Chapter 2: $M(q) \le M(p)$, if $q \in \mathcal{F}_C(p)$.

Both properties will be discussed in more detail later. For now, it is useful to note the similarity with a weight in a gravitational potential. The properties highlighted above indeed say that the work-storage device can be treated similarly to the case of a massive body under the influence of a conservative force in classical mechanics: There is a state-variable M and its difference along a transition determines the work-value of the transition. Using catalytic free operations, which generalise the concept of putting a system in contact with a heat bath in phenomenological thermodynamics, this state-variable cannot be increased. This is even true when considering tasks such as heating or cooling in which the system undergoes a transition between two states with fluctuating energy and possibly large amounts of entropy.

Let us now discuss in more detail the consequenecs of the property of *additive monotonicity* as stated in the theorem. By considering particular free transitions, it is easy to derive the following properties of M.

[5] Technically, it can be represented by the state given by the real number 1 with Hamiltonian 0 on the Hilbert-space $\mathbb{C}$.

Lemma 4.3 (Consequences of additive monotonicity). *If a function M fulfills additive monotonicity and $M(\varnothing) = 0$, then it also fulfills the following properties.*

- Monotonocity: $M(q) \leq M(p) \ \forall \ p, q \in \mathcal{P}$, such that $q \in \mathcal{F}(p)$.

- Additivity: $M(p_A \otimes p_B) = M(p_A) + M(p_B) \ \forall \ p_A, p_B, p_A \otimes p_B \in \mathcal{P}$.

- Positivity: $M(p) \geq 0 \ \forall \ p \in \mathcal{P}$.

These properties only hold true for objects that describe valid work-storage devices, i.e., objects $p \in \mathcal{P}$. In particular, if $\mathcal{P}$ imposes restrictions on what constitutes a valid work-storage device, they are in general not sufficient to guarantee that a function M provides a valid work-quantifier. In the special case that $\mathcal{P}$ contains all quantum states and Hamiltonians, however, Monotonicity and Additivity do indeed imply that M induces a valid work-quantifier. Furthermore, it should be clear then that if a function M fulfills Monotonicity and Additivity for all quantum states and Hamiltonians, then it also fulfills these properties for any subset $\mathcal{P}$. This allows us to make connection to previous results from the resource theory of quantum thermodynamics and see that many of the monotones of catalytic thermal transitions and catalytic Gibbs-preserving transitions induce work-quantifiers:

Lemma 4.4 (Work quantifiers from Rényi-divergences). *If the free transitions $\mathcal{F}$ are given by Gibb-preserving transitions or any subset of them, then the work quantifier $\mathcal{W}_\alpha (p \rightarrow q, \beta) := \Delta F_\beta^\alpha(q) - \Delta F_\beta^\alpha(p)$ with*

$$\Delta F_\beta^\alpha (\rho, H) = \frac{1}{\beta} D_\alpha \left(\rho \| \omega_\beta(H) \right), \tag{4.15}$$

where D_α are the quantum Rényi-divergence (see chapter 2), fulfills Axioms 1 and 2 for any set $\mathcal{P}$ of quantum states and Hamiltonians that admit finite-temperature Gibbs-states.

From all the functions ΔF_β^α, the usual non-equilibrium free energy $\Delta F_\beta = \Delta F_\beta^1$, will play a crucial role later. It can, in a sense, be seen as a "universal" work-quantifier in this framework (if we apply it to quantum mechanical systems or work in classical statistical mechanics). Nevertheless, it is crucial to understand that the functions ΔF_β^α are *not* the only possible work-quantifiers. For example, if $\mathcal{P}$ consists of all possible Hamiltonians, but only accepts energy-eigenstates as quantum-states, then the energy would provide a valid function M, independent of the value of β. The precise set of valid work-quantifiers thus always depends on the precise structure of $\mathcal{P}$.

4.6   Irreversibility and the second law

When introducing the Axioms, I wrote that they encode the second law of thermodynamics. Let us now turn to how exactly a quantitative Second Law emerges from the Axioms. In the language of this chapter, a Second Law should be expressed in terms of the quantities W_{value} and W_{cost} as $W_{\mathrm{value}} \leq W_{\mathrm{cost}}$. Indeed, we can easily derive this inequality from the Axioms and the definition of the two quantities (eqs. (4.6) and (4.7)). To do that, simply consider a cyclic sequence of the form $p \rightarrow \varnothing \rightarrow p$ in Axiom 2. The empty object can of course also be replaced by any object in thermal equilibrium, since transitions of the form $\varnothing \rightarrow w_\beta$ and $p \rightarrow w_\beta$ are "for free" by assumption. By Axiom 2, we then have

$$0 \geq W_{\mathrm{trans}}(p \rightarrow \varnothing, \beta) + W_{\mathrm{trans}}(\varnothing \rightarrow p, \beta). \tag{4.16}$$

Applying the definitions of the work value and work cost, Eqs. (4.6) and (4.7), we then obtain

$$W_{\mathrm{value}}(p) \leq W_{\mathrm{cost}}(p). \tag{4.17}$$

Importantly, this inequality holds for all objects and not just valid work storage devices $p \in \mathcal{P}$. Let us now discuss when equality can be reached in this inequality. Unravelling the definitions of the work value and work cost, we see that the two can only be equal if it

is possible to find a catalytic free transition which takes the system described by p together with an uncorrelated work-storage device, extracts the maximum amount of work possible into a work-storage device while bringing p to thermal equilibrium and then *reverses* this transition using a free catalytic transition while creating p from thermal equilibrium again. In other words there has to be a physical process in the free transitions that reversibly extracts all "non-equilibriumness" of p into a valid work-storage devices. This is always possible if $p \in \mathcal{P}$ from the beginning, because in this case Merlin can simply swap p with work-storage device. Thus, we see that

$$W_{\text{value}}(p) = W_{\text{cost}}(p), \quad \text{if} \quad p \in \mathcal{P}. \tag{4.18}$$

On the other hand, if $p \notin \mathcal{P}$ we expect that such a process is not possible: By trying to convert the "non-equilibriumness" of p into some valid work-storage device, there will be unavoidable dissipation, which results in *irreversibility*:

$$W_{\text{value}}(p) < W_{\text{cost}}(p). \tag{4.19}$$

Such irreversibility has indeed shown up in many analyses of deterministic work extraction (or single-shot work extraction) [50, 52, 70, 150, 151, 154]. From a heuristic point of view, this is obvious: If Merlin had some sort of "burning fuel" (or other non-equilibrium resource), he couldn't simply give this burning fuel to Arthur, if Arthur demanded that Merlin raises Arthur's weight. Merlin had to actually use the burning fuel to lift the weight, which would result in at least some amount of dissipation.

On the other hand, if the set $\mathcal{P}$ would be completely unrestricted, Arthur would accept the burning fuel as a work-storage device and the best strategy for Merlin would indeed be to simply hand the burning fuel to Arthur without doing anything with it, because any free transition can only reduce the non-equilibriumness in a system.

These points highlight again the importance of formulating precise operational tasks, encoded in non-trivial sets $\mathcal{P}$, for the emergence of thermodynamic properties.

Finally, observe that the fact of whether irreversibility occurs is not a property of the work-quantifier, but of the set of free transitions and $\mathcal{P}$. The work-quantifier merely quantifies how much reversibility is violated.

4.7   The role of correlations and super-additivity

We now turn to discussing the role of correlations between the fuel (Merlin's system) and the work-storage device (Arthur's system). The correct treatment of correlations is subtle, but will turn out to be essential for the validity of the second law. In the definition of the work of transition (Definition 4.1), it is assumed that Merlin's and Arthur's systems end up uncorrelated. To investigate what effect such an assumption has, let us now define a *correlated work of transition* in the following way.

Definition 4.5 (Correlated work of transition)**.** *Given a work quantifier $\mathcal{W}$ and inverse temperature β, a set of restrictions $\mathcal{P}$, and initial and final objects of M, denoted by $p_M^{(i)}$ and $p_M^{(f)}$, respectively, the* correlated work of transition $W_{\text{trans}}^{\text{corr}}(p_M^{(i)} \to p_M^{(f)}, \beta)$ *is defined as*

$$W_{\text{trans}}^{\text{corr}}(p_M^{(i)} \to p_M^{(f)}, \beta) := \sup_{\substack{p_A^{(i)}, p_A^{(f)} \in \mathcal{P}; \\ p_{MA}^{(f)} \in \mathcal{F}_C(p_M^{(i)} \otimes p_A^{(i)})}} \mathcal{W}(p_A^{(i)} \to p_A^{(f)}, \beta).$$

The only difference to the previous definition is that the fuel and the work-storage device are now allowed to become correlated in the process implemented by Merlin. We can also define the correlated work cost and value as

$$W_{\text{value}}^{\text{corr}}(p_M, \beta) := W_{\text{trans}}^{\text{corr}}(p_M \to \omega_\beta, \beta), \tag{4.20}$$

$$W_{\text{cost}}^{\text{corr}}(p_M, \beta) := -W_{\text{trans}}^{\text{corr}}(w_\beta \to p_M, \beta). \tag{4.21}$$

These modifications might seem a bit pedantic in the beginning. But we will shortly see that by allowing the built-up of correlations one can violate the "correlated" second law in the sense that there are work-quantifiers that fulfill Axioms 1 and 2, but $W_{\text{value}}^{\text{corr}}(p, \beta) > W_{\text{cost}}^{\text{corr}}(p, \beta)$ for some objects p. To see this, we make use of the absence of *super-additivity* in work-quantifiers. Super-additivity is concerned with the value of a function on bipartite objects with non-interacting Hamiltonians, but possibly correlated states. A function is called super-additive if it fulfills

$$M(p_{AM}, \beta) \geq M(p_A, \beta) + M(p_M, \beta) \tag{4.22}$$

for all objects such that $p_A, p_M, p_{AM} \in \mathcal{P}$. Here p_A and p_M denote the local objects of systems A and M, when the combined system is described by object p_{AM}. While the non-equilibrium free energy ΔF_β fulfills super-additivity, the Rényi-divergences for $0 \neq \alpha \neq 1$ do not.

Now assume a work-quantifier such that M does not fulfill super-additivity for some bipartite object p_{AM}, i.e., $M(p_{AM}, \beta) < M(p_A, \beta) + M(p_M, \beta)$. Furthermore assume that Arthur's system is initially described by p_{AM} and he simply gives subsystem M to Merlin. This puts an upper bound on the work cost of p_M as

$$W_{\text{cost}}^{\text{corr}}(p_M, \beta) \leq M(p_{AM}, \beta) - M(p_A, \beta). \tag{4.23}$$

On the other hand, assume that Merlin initially has p_M (as a standalone) and Arthur does not have any system. If Merlin simply gives p_M to Arthur, we obtain a lower bound on the work value as

$$W_{\text{value}}^{\text{corr}}(p_M, \beta) \geq M(p_M, \beta) - M(\varnothing, \beta) = M(p_M, \beta). \tag{4.24}$$

But using the assumption $M(p_{AM}, \beta) < M(p_A, \beta) + M(p_M, \beta)$ we then obtain

$$W_{\text{cost}}^{\text{corr}}(p_M, \beta) < W_{\text{value}}^{\text{corr}}(p_M, \beta). \tag{4.25}$$

Let us now discuss what would happen if Merlin would like to use this result to become arbitrarily rich, or in other words, create a *perpetuum-mobilé*. Clearly, he can do this if he has access to a "stream of Arthurs", each of whom possesses a copy of p_{AM}. First he obtains a copy of p_M, which he sells and earns $W_{\text{value}}^{\text{corr}}(p_M, \beta)$. He then simply buys p_M from one of the Arthurs who own a correlated system p_{AM}, which finishes the transaction cycle. Merlin's system is thus returned to its initial state after each cycle. In this sense Merlin is not using up resources. On the other hand, Merlin is responsible for correlating both his system with that of the Arthurs and the system of the different Arthurs among themselves. Even though in this particular example he merely "shifts around" the correlations, since he only swaps subsystems, in more general scenarios he might in fact create new correlations. This means that he is spending the "absence of correlations". A similar, but non-equivalent, effect has been discussed in Ref. [105], where the correlations are established among different parts of the catalyst. Thus, one possible viewpoint is to state that in order to account properly for resources, correlations cannot be created. Hence, the second law would take the form of the "uncorrelated" Second Law (4.17) which is indeed fulfilled for any work quantifier satisfying Axioms 1 and 2.

One can, however, also consider a complementary view in which it is allowed to create correlations between a catalyst (or Merlin's system) and the work-storage device. In this case, the possible work-quantifiers have to be modified accordingly and the second law would take the form $W_{\text{cost}}^{\text{corr}}(p_M, \beta) \geq W_{\text{value}}^{\text{corr}}(p_M, \beta)$. This can be achieved by introducing a reformulation of the second Axiom in terms of the correlated work of transition:

Axiom 3 (Correlated cyclic transitions of the fuel). *For any cyclic sequence of transitions of the "fuel" (Merlin's system) $p_M^{(1)} \to p_M^{(2)} \to \cdots \to p_M^{(n)} = p_M^{(1)}$, the sum of the optimal work that Merlin can obtain in each sequence when correlations with the work-storage device are allowed (this is given by $W_{\text{trans}}^{\text{corr}}$ in (4.5)) is smaller or equal to zero,*

$$\sum_{i=1}^{n-1} W_{\text{trans}}^{\text{corr}}(p_M^{(i)} \to p_M^{(i+1)}, \beta) \leq 0. \tag{4.26}$$

The intuition behind this Axioms is exactly the same as that behind Axiom 2, with the only difference that Merlin is now allowed to create correlations with the work-storage device. As expected, one can now easily show that any valid work-quantifier fulfills the second law in the form

$$W_{\text{cost}}^{\text{corr}}(p_M, \beta) \geq W_{\text{value}}^{\text{corr}}(p_M, \beta). \tag{4.27}$$

Let us explore how this comes about. From Axiom 3 it follows directly that $W_{\text{trans}}^{\text{corr}}(p_M \to p_M) \leq 0$. Unravelling definition 4.5, we can see that this implies

$$\mathcal{W}(p_A \to q_A, \beta) \leq 0 \tag{4.28}$$

for all q_A, p_A in $\mathcal{P}$ such that $q_A \in \mathcal{F}_C^{\text{Corr.}}(p_A)$. Here, we define $\mathcal{F}_C^{\text{Corr.}}(p)$ to be the set of objects that can be reached from p by using thermal baths and an auxiliary system that is left, after the interaction with the bath, with the same marginal state and Hamiltonian, but possibly correlated with the system. We will refer to this transitions as *correlated catalytic free transitions*. It is easy to see that $p_A \otimes p_B \in \mathcal{F}_C^{\text{Corr.}}(p_{AB})$ for $p_A, p_B, p_{AB} \in \mathcal{P}$. We can now use Additivity together with Axioms 1 and 3 to show that valid work-quantifiers have to take the form $\mathcal{W}(p \to q, \beta) = M(q, \beta) - M(p, \beta)$ with M, apart from the properties stated in theorem 4.2, also fulfilling super-additivity:

$$M(p_{AB}, \beta) \geq M(p_A, \beta) + M(p_B, \beta), \tag{4.29}$$

where p_{AB} describes a bipartite and non-interacting, but possibly correlated system. This property has important consequences. From all the monotones of catalytic thermal transitions induced from Rényi-divergences (see again chapter 2), only the non-equilibrium free energy ΔF_β remains to be a valid and non-trivial monotone that induces a valid work-quantifier for arbitrary sets $\mathcal{P}$.

Indeed, in the next chapter, we will show the following much stronger result:

Theorem 4.6. *If we consider as free transitions those induced by* Gibbs-preserving maps*, then the non-equilibrium free energy ΔF_β is (up to a constant) the* only *possible function that is* continuous *on the quantum state and induces a valid work-quantifier for arbitrary sets $\mathcal{P}$ of quantum states and Hamiltonians under respecting Axioms 1 and 3.*

Properly accounting for the built-up of correlations therefore allows us to essentially show that there is only one sensible *universal* work-quantifier, namely the non-equilibrium free energy. Importantly, we have singled out the free energy without making any reference to a *thermodynamic limit*, but it followed from considerations about how to take into account correlations.

4.8   Probabilistic work

From a general point of view, the results in this chapter can be used to constrain possible work-quantifiers in almost arbitrary resource theories once operational tasks can be described as state-transitions within specified sets of states $\mathcal{P}$. In this last section, let me briefly comment on how the notion of probabilistic work fits into the framework.

Probabilistic work, with which fluctuation theorems are concerned (see, e.g., section 2.5), is concerned with associating a probability distribution of work to a given process. This requires that there is some probabilistic component to this process and it furthermore requires that in every run of the process a definite work-value can be associated. In the usual example of fluctuation relations, the probabilistic component is given by the energy-measurements that are being done before and after a previously specified protocol is implemented on a system. The definite work-value that is associated to a single run is simply the difference of the outcomes of the measurements.

In the language of this chapter, such a scenario can be translated as Arthur giving to Merlin a system in a known state ρ_α with some probability $P(\alpha)$ and Merlin returning to Arthur the system in state ρ_γ with conditional probability $P(\gamma|\alpha)$. For example, Merlin could randomly choose a process that he implements in his lab, with the probability for the different

process depending on α. Or he could implement a process, which contains a probabilistic element such as an energy measurement. In any case, the work-value $\mathcal{W}(\rho_\alpha \rightarrow \rho_\gamma)$ occurs with probability $P(\gamma, \alpha) = P(\gamma|\alpha)P(\alpha)$ and the average work-value is given by

$$\langle \mathcal{W} \rangle = \sum_{\alpha, \gamma} P(\gamma, \alpha) \mathcal{W}(\rho_\alpha \rightarrow \rho_\gamma). \tag{4.30}$$

Probabilistic work is thus incorporated completely naturally in the framework independent of the set $\mathcal{P}$. Note however, that the average work-value $\langle \mathcal{W} \rangle$ *cannot* be calculated as an "average transition". In particular

$$\langle \mathcal{W} \rangle \neq \mathcal{W} \left(\sum_\alpha P(\alpha)\rho_\alpha \rightarrow \sum_\gamma P(\gamma)\rho_\gamma \right). \tag{4.31}$$

In general, the right-hand side need not even be defined since $\mathcal{P}$ is in general not a convex set and therefore $\sum_\alpha P(\alpha)\rho_\alpha$ is in general not contained in $\mathcal{P}$. As an example, suppose $\mathcal{P}$ only contains energy-eigenstates and Arthur always comes with the ground state $|E_0\rangle$ to Merlin. Then Merlin returns the system in the energy-eigenstate $|E_i\rangle$ with probability $w_i := \mathrm{e}^{-\gamma E_i}/Z_\beta$ [6]. Furthermore suppose that the work-quantifier is simply given by the non-equilibrium free energy ΔF_β. Then in every run, Arthur pays an amount of money E_i to Merlin with probability w_i. In particular, on average Arthur pays to Merlin the amount of money $E_\gamma = \sum E_i w_i$. However, the *average outcome state* of Merlin's action is simply a thermal state at the environment temperature. Thus, for Merlin to be able to claim the money, he has to make sure that Arthur's system is in a known energy-eigenstate at the end of each run of the process, which forces him to implement a projective energy measurement. Otherwise, Arthur would demand compensation for the thermalization of his valuable work-storage device. Since projective energy measurements are costly from a thermodynamic point of view, Merlin cannot get rich using this strategy (see section 3.6).

[6] For example, Merlin simply thermalizes the system and then measures the energy.

4.9   Summary

As miniaturization of machines progresses forward it is becoming increasingly interesting to ask how to quantify the performance of machines on the micro-scale. In this chapter, I discussed an operational approach to this problem by formulating it as a transaction between two players in the language of a resource theory. To make the discussion more concrete I worked in the resource theory of thermodynamics and considered the problem of measuring "thermodynamic work" from the microscopic perspective. Demanding that a possible work-quantifier fulfills natural Axioms then allowed us to derive surprisingly stringent conditions on such work-quantifiers, such as additivity and monotonicity under free transitions. A particular example of a possible work-quantifier in the setting of thermodynamics is given by the non-equilibrium free energy. Nevertheless, this does by no means imply that it is the only such quantifier. Indeed, possible quantifiers depend crucially on the set $\mathcal{P}$ that encodes the operational task at hand.

We discussed in detail the role of correlations for the emergence of the second law of thermodynamics. In the next chapter, we will see that this approach yields a new axiomatic derivation of the relative entropy and free energy. This discussion will also become important again in chapter 8, where we discuss quantitative arguments that show that it is difficult to prepare complex systems in states that do not equilibrate.

The general Axioms put forward in this chapter, as well as the stringent conditions on possible quantifiers for state-transitions in resource theories, are broadly applicable also to other resource theories that allow for catalysts. It is an interesting avenue for further research to explore the results in this chapter in the context of different resource theories.

5 Uniqueness of the relative entropy and free energy

THE NON-EQUILIBRIUM FREE ENERGY ΔF_β is one of the central quantities in non-equilibrium thermodynamics. We have already seen it appearing frequently in the previous chapters, and it will continue to show up in later chapters related to work extraction and the minimum work-principle. From the perspective of resource theories, this is surprising, since it is just one of many possible monotones. Indeed, the results of chapter 3 show that other monotones play important roles for fundamental thermodynamic tasks such as low-temperature cooling.

It is sometimes argued that the non-equilibrium free energy ΔF_β only plays an important role in the thermodynamic limit. The reason is given by the following argument. Suppose one wants to prepare a large number n of copies of ρ. In such a setting it is often reasonable to allow for an arbitrary small error in the final state, i.e., allow for protocols that prepare a state ρ'_n that is ϵ close to $\rho^{\otimes n}$ in trace-norm. In such a setting, the *densities* of the Rényi-divergences D_α of the final state ρ'_n all converge to the non-equilibrium free energy density if one first takes the limit $n \to \infty$ and then the limit $\epsilon \to 0$ [49, 69]. Thus, it seems that only the non-equilibrium free energy remains as a valid monotone. Similar results are expected to hold true if one weakens the assumption that the final state is uncorrelated to assuming that it is weakly correlated. [1]

On the other hand, the discussion about the role of correlations and the second law in chapter 4 showed that the non-equilibrium free energy can be singled out from the other monotones based on Rényi-divergences by an argument on how one treats the build-up of correlations in models of free transitions. This choice in turn determines whether monotones have to be *super-additive* or not. No thermodynamic limit is required to make this argument, but it essentially rests on the subjective choice of what an experimentalist considers natural.

In this chapter[2], I will present a rigorous argument which uniquely singles out the non-equilibrium free energy based on arguments similar to those which entered the discussion of the role of correlations for valid work quantifiers. From a technical perspective, the result is based on results by Matsumoto [155] about the quantum relative entropy, which were proven using concepts from quantum hypothesis testing [156, 157]. The result by Matsumoto allows us to prove the following new axiomatic characterization of the quantum relative entropy on finite dimensional Hilbert-spaces.

Theorem 5.1 (Uniqueness theorem)**.** *Let $f(\rho, \sigma)$ be a function on pairs of density matrices such that $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$. Furthermore, let f have the following properties:*

1. Continuity*: For fixed σ, the map $\rho \mapsto f(\rho, \sigma)$ is continuous.*

[1] Strictly speaking, even for very large n it is not the densities of the monotones that determine the possibility of a transition in the framework of thermal operations, but their absolute value. It could thus in principle be the case that contributions in the monotones that scale sub-linearly with n indeed play important roles in some cases. This constitutes an interesting open problem.

[2] This chapter is based on work together with Rodrigo Gallego and Jens Eisert published in [7].

2. Data-processing inequality: *For any quantum channel T we have,*

$$f(T(\rho), T(\sigma)) \leq f(\rho, \sigma). \tag{5.1}$$

3. Additivity:

$$f(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) = f(\rho_1, \sigma_1) + f(\rho_2, \sigma_2). \tag{5.2}$$

4. Super-additivity: *For any bipartite state $\rho_{1,2}$ with marginals ρ_1, ρ_2 we have*

$$f(\rho_{1,2}, \sigma_1 \otimes \sigma_2) \geq f(\rho_1, \sigma_1) + f(\rho_2, \sigma_2). \tag{5.3}$$

Then $f(\rho, \sigma) = CD(\rho \| \sigma)$, with $C \geq 0$ and D the quantum relative entropy.

Before giving the proof of this result in section 5.4, I will discuss in detail how it can be related to thermodynamics to single out the non-equilibrium free energy. An essential step in this argument will be to broaden our class of free transitions from the model of thermal operations to the class of *correlated-catalytic transitions* using *Gibbs-preserving maps* [1, 2, 46, 158]. In the next section, I will first describe these generalized transitions as well as related transitions, which have been introduced in [56] and which I will call *marginal-catalytic transitions*. Then I will use the above result about the relative entropy to prove that the only continuous and additive monotone of such transitions is given by the non-equilibrium free energy.

5.1   Gibbs-preserving maps

So far, we have used the model of thermal operations to treat thermodynamics as a quantum mechanical resource theory. From a fundamental perspective, thermal operations are indeed very natural, since they essentially only rely on energy-conservation and unitarity of quantum mechanics together with the fact that thermal systems are well-described by Gibbs-states. As an important consequence thermal operations have the property that they cannot bring a system out of equilibrium: Any thermal operation at environment temperature $1/\beta$, represented by a quantum channel $\mathcal{G}_\beta$, has the property that it leaves the Gibbs state of the system Hamiltonian invariant:

$$\mathcal{G}_\beta(\omega_\beta(H_S)) = \omega_\beta(H_S). \tag{5.4}$$

This is indeed a necessary property for any sensible class of free operations in a resource theory that is supposed to model thermodynamics. If it were not true, but Gibbs states at the environment temperature were considered free states, it would be possible to produce an arbitrary amount of non-equilibrium states for free from the environment, obviously violating the second law of thermodynamics.

It is, however, known that thermal operations are not the largest class of free operations that are Gibbs-preserving with respect to the system Hamiltonian [158]. This is due to the fact that they commute with the time-evolution on the system and hence cannot increase the degree of coherence of the state on the system. We will now enlargen our class of free operations by simply allowing for all quantum channels that leave Gibbs-states of the system Hamiltonian invariant. Furthermore, as in chapter 4, we will allow not only changes of the quantum state but also of the Hamiltonian. From the level of thermal operations this can be motivated by considering a situation where we apply the energy-conserving unitary U on the system and bath, but instead of tracing out the bath afterwards, we trace out the system and some part of the bath, leaving us with a new state on this part of the bath, which evolves according to its own Hamiltonian.

To implement the possibility to change a Hamiltonian formally in the framework of Gibbs-preserving maps, we can define a *Gibbs-preserving map (GP-map)* as a function $(\rho, H) \mapsto (\sigma, K) = G_\beta(\rho, H)$ such that

$$G_\beta(\omega_\beta(H), H) = (\omega_\beta(K), K). \tag{5.5}$$

The function thus maps Gibbs-states to Gibbs-states, preserving the temperature but possibly modifying the Hamiltonian. This condition can also be expressed as follows: We can

define GP channels as $\mathcal{G}_\beta^H(\omega_\beta(H)) = \omega_\beta(K(H))$ for all H, and a corresponding map between Hamiltonians as $\bar{\mathcal{G}}_\beta(H) = K$ so that

$$G_\beta(\rho, H) = (\mathcal{G}_\beta^H(\rho), \bar{\mathcal{G}}_\beta(H)). \tag{5.6}$$

With this notation, condition (5.5) is given by

$$\mathcal{G}_\beta^H(\omega_\beta(H)) = \omega_\beta\left(\bar{\mathcal{G}}(H)\right). \tag{5.7}$$

GP-maps G_β are not only a natural extension of GP-channels $\mathcal{G}_\beta$ for the case where Hamiltonians are modified, but one can also see that any GP-map can be implemented if one is given access to a GP-channels and arbitrary ancillary systems in Gibbs states. This is formalized by the following Lemma.

Lemma 5.2 (Implementation of GP maps [2]). *Any map G_β fulfilling* (5.5) *acting on a system S can be implemented by adding an ancillary system A in the Gibbs state $(\omega_{\beta,K}, K)$ and applying a GP channel $\mathcal{G}_\beta$ to the entire compound. More formally, we find that*

$$G_\beta(\rho_S, H_S) := (\sigma, K) = \left(\mathrm{Tr}_S(\mathcal{G}_\beta(\rho \otimes \omega_\beta(K))), K\right). \tag{5.8}$$

For the rest of the chapter, I will say that there is a *free transition $p \rightsquigarrow q$* between two objects $p = (\rho, H)$ and $q = (\sigma, K)$ if there is a GP-map G_β such that $G_\beta(\rho, H) = (\sigma, K)$.

The concept of GP-maps in fact allows us to interpret *any* quantum channel $\mathcal{T}$ as a GP-map. Suppose a quantum channel maps the Gibbs-state $\omega_\beta(H)$ to the density matrix ρ and for simplicity also assume that ρ has full rank. We can always define the Hamiltonian

$$H_\rho^\beta := -\frac{1}{\beta} \ln \rho + C(\rho)\mathbf{1}, \tag{5.9}$$

where $C(\rho)$ is any real-valued function. With this definition, any object of the form (ρ, H_ρ^β) is a *Gibbs-object* representing a system in thermal equilibrium. Similarly, given any quantum channel $\mathcal{T}$ we can define a derived map on Hamiltonians given by

$$\bar{\mathcal{T}}(H) := H_{\mathcal{T}(\omega_\beta(H))}^\beta. \tag{5.10}$$

Then $T(\rho, H) = (\mathcal{T}(\rho), \bar{\mathcal{T}}(H))$ defines a Gibbs-preserving map. The above argument only works if $\rho = \mathcal{T}(\omega_\beta(H))$ has full rank. There clearly are quantum-channels $\mathcal{T}$ that reduce the rank of full-rank states. This is not a problem, however. To see this, we can make use of the following fact about quantum channels:

Lemma 5.3 (Rank-decreasing quantum channels). *Let $T : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ be a quantum channel and σ any full-rank state. If $T(\sigma)$ is only supported on a subspace $P \subseteq \mathcal{H}'$, then $T(\rho)$ is supported only within P for any ρ.*

Proof. The proof is given in section 14.4.1. □

By the previous Lemma, any quantum channel that maps a full-rank state σ into a state $\mathcal{T}(\sigma)$ without full rank simply maps all states to the smaller Hilbert space $P = \mathrm{supp}(\mathcal{T}(\sigma))$ and should be considered as a map from states on $\mathcal{H}$ to states on P instead. Thus, we can simply take the position that quantum channels which decrease the rank on some state map from a larger Hilbert-space to a smaller one but always map full-rank states to full-rank states.

Above we have generalized the concept of thermal operations and Gibbs-preserving channels to Gibbs-preserving maps, which also allow to change the Hamiltonian. We have also seen how any quantum channel can be interpreted as a GP-map by adjusting Hamiltonians accordingly. In the rest of the chapter, *free transitions* will always be interpreted as those induced by GP-maps. I will now go on to also broaden the notion of catalytic free transitions and to catalytic free transitions where it is allowed to build up correlations between the system and catalyst or within the catalyst. This will be an essential step to single out the non-equilibrium free energy.
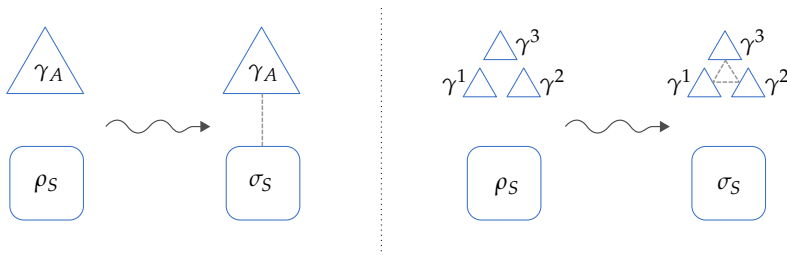
Figure 5.1: Illustration of correlated-catalytic and marginal-catalytic free transitions. Triangles denote catalysts and rectangles the system of interest. Grey, dashed lines indicate the possibility of correlations between the systems connected by the lines. *Left figure:* In the case of correlated-catalytic free transitions the system can become correlated with the catalyst, while the catalyst remains unchanged. *Right figure:* In the case of marginal-catalytic free transitions no correlations are established between the system and the catalyst, but correlations may be established between parts of the catalyst while the parts remain unchanged.

5.2   Catalysts and Correlations

Just as in previous chapters, in what follows a *catalytic free transition* is a free transition of the form $p \otimes r \rightsquigarrow q \otimes r$ and denoted by $p \underset{c}{\rightsquigarrow} q$. Here, I use the same notation as in the last chapter, i.e.,

$$(\rho, H) \otimes (\chi, R) = (\rho \otimes \chi, H \otimes \mathbf{1} + \mathbf{1} \otimes R). \tag{5.11}$$

We have already seen in the last chapter, that in some situations it might be natural to consider also a larger class of free transitions in which the catalyst can become correlated with the system. Let us now define such transitions formally (see Fig. 5.1 for an illustration).

Definition 5.4 (Correlated-catalytic free transition). *A transition*

$$(\rho_S, H_S) \rightarrow (\sigma_S, K_S) \tag{5.12}$$

is called a correlated-catalytic free transition *if there exist a GP map G and a system A described by the object (γ_A, R_A) such that*

$$G\big((\rho_S, H_S) \otimes (\gamma_A, R_A)\big) = (\eta, K_S + R_A), \tag{5.13}$$

where $\operatorname{Tr}_A(\eta) = \sigma_S$ and $\operatorname{Tr}_S(\eta) = \gamma_A$. In the following such a transition will be denoted by

$$(\rho_S, H_S) \underset{cc}{\rightsquigarrow} (\sigma_S, K_S). \tag{5.14}$$

Correlated-catalytic free transitions have the property that the catalyst remains unchanged, but there is a possible build-up of correlations between the catalyst and the system at hand. This allows for a strictly larger set of transitions, which are considered "free".

The importance of the careful treatment of correlations related to catalysts has in fact been found before in the literature. In Ref. [105], the authors introduced a notion of multipartite catalysts, in which the different parts of the catalyst remain unchanged, but can become correlated. Formally, they can be defined in the following way.

Definition 5.5 (Marginal-catalytic free transition [105]). *A transition*

$$(\rho_S, H_S) \rightarrow (\sigma_S, K_S) \tag{5.15}$$

is called a marginal-catalytic free transition *if there exist a GP map G and systems $A_1, \ldots, A_k$ described by the object $(\gamma_A, R_A) = \otimes_{i=1}^{k}(\gamma^i, R^i)$ such that*

$$G\big((\rho_S, H_S) \otimes (\gamma_A, R_A)\big) = (\sigma_S, K_S) \otimes (\tilde{\gamma}_A, R_A), \tag{5.16}$$

where $\operatorname{Tr}_{|A_i}(\tilde{\gamma}_A) = \operatorname{Tr}_{|A_i}(\gamma_A)$ for all $i \in (1, \ldots, k)$. In the following such a transition will be denoted by

$$(\rho_S, H_S) \underset{mc}{\rightsquigarrow} (\sigma_S, K_S). \tag{5.17}$$

In the case of marginal-catalytic free transitions the system A does not remain unchanged, but only its local marginals. In this sense, it is not truly a catalyst, but a catalyst on its reduced states.

In the case of correlated-catalytic transitions, on the other hand, the catalyst truly remains unchanged, however it builds up correlations with the system at hand. There is therefore no guarantee that it can be used to implement transitions on the system again. But it can be used as a catalyst for transitions implemented on different system with which it is still uncorrelated.

It is natural to expect that both kinds of transitions indeed allow for a larger set of free transitions, since the initial lack of correlations is "used up". Indeed, it was shown in Ref. [105] that marginal-catalytic free transitions provide an advantage over catalytic free transitions. By now, we know that in fact both sets of free transitions are identical. I will comment in more detail on this later.

There is one kind of transition which is both a correlated-catalytic free transition as well as a marginal-correlated catalytic free transition and which plays a crucial role in the following. Suppose we have a bi-partite, but non-interacting system which is initially in a correlated state ρ_{12}. Then

$$(\rho_{12}, H_1 + H_2) \underset{cc,mc}{\rightsquigarrow} (\rho_1 \otimes \rho_2, H_1 + H_2). \tag{5.18}$$

In the case of correlated-catalytic transitions this follows by simply swapping a copy of ρ_1 with the corresponding sub-system. In the case of marginal-catalytic operations it follows by simply having as initial catalyst $\rho_1 \otimes \rho_2$ and swapping the whole catalyst with the system. In both models of catalytic free transition it is hence for free to de-correlate a system. This has importnat consequences for possible monotones.

5.3   Monotones and the free energy

Having properly defined the free transitions that we will consider for the rest of the chapter, let us now turn to the monotones. In the end we want to argue that any reasonable monotone of correlated-catalytic free transitions and marginal-catalytic free transitions has to be proportional to ΔF_β by using theorem 5.1. So far we have defined montones M_β as functions on pairs of quantum states and Hamiltonians, which we called objects. The relative entropy is a function on pairs of quantum states instead. We therefore have to first pass to monotones defined as functions on pairs of quantum states. To do that, in the following we will only consider monotones that are *additive*, i.e., they fulfill

$$M_\beta(\rho_1 \otimes \rho_2, H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2) = M_\beta(\rho_1, H_1) + M_\beta(\rho_2, H_2).$$

Such monotones automatically have three very natural properties:

Lemma 5.6 (Normalization,positivity and gauge-invariance). *Let M_β be an additive monotone. Then:*

1. $M_\beta(\omega_\beta(H), H) = 0$ *for all H.*

2. $M_\beta(\rho, H) \geq 0$.

3. $M_\beta(\rho, H) = M_\beta(\rho, H + \lambda\mathbf{1})$ *for all $\lambda \in \mathbb{R}, \rho$ and H,*

Proof. We first prove property 1. Since tracing out and attaching thermal states are free transitions, we have by additivity that

$$M_\beta(\omega_\beta(H_1), H_1) \geq M_\beta(\omega_\beta(H_1), H_1) + M_\beta(\omega_\beta(H_2), H_2) \geq M_\beta(\omega_\beta(H_1), H_1). \tag{5.19}$$

But then

$$M_\beta(\omega_\beta(H_1), H_1) = M_\beta(\omega_\beta(H_1), H_1) + M_\beta(\omega_\beta(H_2), H_2) \tag{5.20}$$

for arbitrary H_1, H_2. We thus conclude $M_\beta(\omega_\beta(H), H) = 0$ for all H. Positivity then follows immediately, since bringing the system to thermal equilibrium is a free transition:

$M_\beta(\rho, H) \geq M_\beta(\omega_\beta(H), H) = 0$. To prove property 3 we use a similiar trick. First note that

$$(H_1 + \lambda\mathbf{1}) \otimes \mathbf{1} + \mathbf{1} \otimes H_2 = H_1 \otimes \mathbf{1} + \mathbf{1} \otimes (H_2 + \lambda\mathbf{1}), \qquad (5.21)$$

and $\omega_\beta(H) = \omega_\beta(H + \lambda\mathbf{1})$. This allows us to use additivity and $M_\beta(\omega_\beta(H), H) = 0$ to prove property 3:

$$
\begin{aligned}
M_\beta(\rho, H + \lambda\mathbf{1}) &= M_\beta(\rho \otimes \omega_\beta(H), (H + \lambda\mathbf{1}) \otimes \mathbf{1} + \mathbf{1} \otimes H) \\
&= M_\beta(\rho \otimes \omega_\beta(H), H \otimes \mathbf{1} + \mathbf{1} \otimes (H + \lambda\mathbf{1})) \\
&= M_\beta(\rho, H) + M_\beta(\omega_\beta(H + \lambda\mathbf{1}), H + \lambda\mathbf{1}) = M_\beta(\rho, H). \qquad (5.22)
\end{aligned}
$$

$\square$

We can now make use of gauge-invariance to pass from functions defined on pairs of quantum states and Hamiltonians to monotones $\mathcal{M}_\beta$ defined on pairs of quantum states. We now simply define $\mathcal{M}_\beta$ using the modular Hamiltonian (5.9) as

$$\mathcal{M}_\beta(\rho, \sigma) := M_\beta(\rho, H_\sigma^\beta). \qquad (5.23)$$

Since M_β is gauge-invariant, this mapping is one-to-one and $\mathcal{M}_\beta$ inherits additivity in the sense of theorem 5.1. So far, $\mathcal{M}_\beta(\rho, \sigma)$ is only defined on all pairs of quantum states with σ being full-rank. However, we can naturally define it whenever $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$ by setting

$$\mathcal{M}_\beta(\rho, \sigma) = \mathcal{M}_\beta(\rho|_{\mathrm{supp}(\sigma)}, \sigma|_{\mathrm{supp}(\sigma)}).$$

In the following, I will say that $\mathcal{M}_\beta$ is a monotone of (correlated,marginal)-catalytic free transitions if M_β is such a monotone.

So far we have only made use of the fact that we require our monotones to be monotonic under free transitions but not under (correlated,marginal)-catalytic free transitions. Let us now make use of that.

Lemma 5.7 (Super-addivity from correlations). *Let M_β be an additive monotone of free transitions. Then it is a monotone of (correlated,marginal)-catalytic free transitions if and only if it is super-additive, i.e., fulfills*

$$M_\beta(\rho_{12}, H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2) \geq M_\beta(\rho_1 \otimes \rho_2, H_1 + H_2). \qquad (5.24)$$

Proof. The free transition $(\rho_{12}, H_1 + H_2) \underset{cc,mc}{\rightsquigarrow} (\rho_1 \otimes \rho_2, H_1 + H_2)$ immediately implies monotonicity under (correlated,marginal)-catalytic free transitions requires super-additivity. For the converse direction, consider first the case of correlated-catalytic transitions. Suppose that $(\rho_S \otimes \rho_C, H_S + H_C) \rightsquigarrow (\rho'_{SC}, H'_S + H_C)$ with $\rho'_C = \rho_C$. Then

$$
\begin{aligned}
M_\beta(\rho_S, H_S) + M_\beta(\rho_C, H_C) &= M_\beta(\rho_S \otimes \rho_C, H_S + H_C) & (5.25) \\
&\geq M_\beta(\rho'_{SC}, H'_S + H_C) & (5.26) \\
&\geq M_\beta(\rho'_S, H'_S) + M_\beta(\rho_C, H_C). & (5.27)
\end{aligned}
$$

Cancelling $M_\beta(\rho_C, H_C)$ we thus obtain $M_\beta(\rho_S, H_S) \geq M_\beta(\rho'_S, H'_S)$. The proof for marginal-catalytic free transitions follows in complete analogy by using super-additivity on the sub-systems of the catalyst. $\square$

Again, the function $\mathcal{M}_\beta$ inherits super-additivity from M_β, but now in the sense of theorem 5.1.

In section 5.1 we have seen that any quantum channel T can be interpreted as a GP-map. Let us now calculate the consequence of this observation. Using the definition of $\mathcal{M}_\beta$ and (5.10) we obtain

$$
\begin{aligned}
\mathcal{M}_\beta(T(\rho), T(\sigma)) &= M_\beta(T(\rho), H_{T(\sigma)}^\beta) \\
&= M_\beta(T(\rho), \overline{T}(H_\sigma^\beta)) \leq M_\beta(\rho, H_\sigma^\beta) \\
&= \mathcal{M}_\beta(\rho, \sigma). \qquad (5.28)
\end{aligned}
$$

We can therefore conclude that $\mathcal{M}_\beta$ fulfills the data-processing inequality. Let us summarize our findings in a lemma.

Lemma 5.8 (Properties of monotones). *Let M_β be an additive monotone of (correlated,marginal)-catalytic free transitions. Then the associated function $\mathcal{M}_\beta$ on pairs of quantum states (with the support of the first argument contained in that of the second argument) has the following properties:*

1. Positivity: $\mathcal{M}_\beta(\rho, \sigma) \geq 0$ and $\mathcal{M}_\beta(\rho, \rho) = 0$,

2. Data-processing inequality: *For any quantum channel T, we have $\mathcal{M}_\beta(\rho, \sigma) \geq \mathcal{M}_\beta(T(\rho), T(\sigma))$.*

3. Additivity: $\mathcal{M}_\beta(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) = \mathcal{M}_\beta(\rho_1, \sigma_1) + \mathcal{M}_\beta(\rho_2, \sigma_2)$,

4. Super-additivity: *For any bipartite state $\rho_{1,2}$ with marginals ρ_1, ρ_2 we have*

$$\mathcal{M}_\beta(\rho_{1,2}, \sigma_1 \otimes \sigma_2) \geq \mathcal{M}_\beta(\rho_1, \sigma_1) + \mathcal{M}_\beta(\rho_2, \sigma_2). \tag{5.29}$$

Comparing with theorem 5.1 we can then finally conclude our main result.

Corollary 5.9 (Uniqueness of the free energy). *Let M_β a monotone of (correlated,marginal)-catalytic free transitions that continuously depends on the quantum state. Then $M_\beta(\rho, H) = C\Delta F_\beta(\rho, H)$ for some $C \geq 0$.*

Proof. By continuity in the first argument, we obtain from theorem 5.1 that $\mathcal{M}_\beta(\rho, \sigma) = CD(\rho\|\sigma)$. But then $M_\beta(\rho, H) = \mathcal{M}_\beta(\rho, \omega_\beta(H)) = C\Delta F_\beta(\rho, H)$. □

5.4 Proof of theorem 5.1

To prove the uniqueness theorem 5.1, we first have to set up a definition. Let (ρ, σ) be a pair of states on a finite-dimensional Hilbert space $\mathscr{H}$ and $\{\rho'_n\}$ be a sequence of states on the Hilbert spaces $\mathscr{H}^{\otimes n}$. Let us define a function f on pairs of quantum states to be *lower asymptotically semi-continuous* with respect to σ if

$$\lim_{n\to\infty} \left\|\rho^{\otimes n} - \rho'_n\right\|_1 = 0 \tag{5.30}$$

implies

$$\liminf_{n\to\infty} \frac{1}{n}\left(f(\rho'_n, \sigma^{\otimes n}) - f(\rho^{\otimes n}, \sigma^{\otimes n})\right) \geq 0. \tag{5.31}$$

In Ref. [155], Matsumoto proved results about the quantum relative entropy, which can be summarized as follows.

Theorem 5.10 (Matsumoto). *Let f fulfill the data-processing inequality, additivity and be lower asymptotically semi-continuous with respect to all σ. Then $f \propto S$.*

theorem 5.1 then follows if we can deduce that any continuous and additive monotone of (correlated,marginal)-catalytic transitions is lower asymptotically semi-continuous. This is shown in the subsequent Lemma, which finishes the proof.

Lemma 5.11 (Lower asymptotic continuity from super-additivity). *. Let f be a function on pairs of quantum states with the following properties,*

- *The map $\rho \mapsto f(\rho, \sigma)$ is continuous for any fixed σ.*

- *Additivity: $f(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) = f(\rho_1, \sigma_1) + f(\rho_2, \sigma_2)$.*

- *Super-additivity:*
$$f(\rho_{12}, \sigma_1 \otimes \sigma_2) \geq f(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2). \tag{5.32}$$

Then f is lower asymptotically semi-continuous with respect to any σ.

Proof. Let $\{\rho'_n\}$ be a sequence of states such that $\|\rho'_n - \rho^{\otimes n}\|_1 \to 0$. Since the trace norm fulfills the data-processing inequality, we know that $\|\rho'_{n,i} - \rho\|_1 \to 0$, where $\rho'_{n,i}$ denotes the marginal of ρ'_n on the i-th tensor-factor. Hence, the marginals converge to ρ. From the properties of f, we furthermore see that

$$\frac{1}{n}\left(f(\rho'_n, \sigma^{\otimes n}) - f(\rho^{\otimes n}, \sigma^{\otimes n})\right) \geq \frac{1}{n}\sum_i \left(f(\rho'_{n,i}, \sigma) - f(\rho, \sigma)\right)$$

$$\geq \min_i\{f(\rho'_{n,i}, \sigma)\} - f(\rho, \sigma) \overset{n \to \infty}{\longrightarrow} 0, \qquad (5.33)$$

where the limit follows from continuity and the second line from additivity and super-additivity. $\qquad\square$

5.5   Conclusions and outlook

The quantum relative entropy is an ubiquitous quantity in quantum information [159] and related fields, appearing in as diverse situations as quantum hypothesis testing [156, 157, 160], entanglement theory [161], recovery of quantum information [162] and even quantum gravity [163].

In this chapter, I presented an axiomatic way to derive the quantum relative entropy and non-equilibrium free energy from the point of view of resource theories. The essential properties which single out the quantum relative entropy is the monotonicity under (quantum) channels together with extensivity and super-additivity. We have seen how super-additivity follows once we allow catalysts to become correlated in any resource theory that allows for swapping of sub-systems. To uniquely single out the usual von Neumann free energy thus in a sense requires the forgetting of correlations between systems and simply treating them as uncorrelated. This is interesting from several point of views.

First, note that we never required that any of the systems is macroscopic in any sense. Thus, the standard von Neumann free energy, and consequently the von Neumann entropy, can be derived without referring to macroscopic systems. Rather, we have to forget information about the systems of question. This theme will be picked up in the next chapter. Nevertheless, note that the idea that correlations can be disregarded fits very well to the case of macroscopic physics: Macroscopic systems usually have a finite correlation length. Two macroscopic bodies are therefore only correlated roughly in the region where they are in contact, which is usually very small compared to their volume. Thus, for practical purposes one can usually neglect the correlations between macroscopic bodies.

Second, from an historical point of view, it is also interesting to note that disregarding correlations between particles was precisely the assumption that Boltzmann needed (in the "Stosszahlenansatz") to derive his famous H-theorem [25, 38, 39].

Third, as we have seen, the non-equilibrium free energy seems to be the only monotone of catalytic thermal operations with natural properties such as extensivity and continuity once we allow catalysts to become correlated. It is then natural to ask whether a decrease of the non-equilibrium free energy is also a *sufficient* criterion for a state transition under these operations. In the case of marginal-catalytic free transitions, this is indeed the case for quasi-classical states, as was shown in Ref. [105]. We conjectured that this was true also for correlated-catalytic transitions in Ref. [7], where we published the results described in this chapter. By now this question has been answered in the affirmative for quasi-classical states in Ref. [164]: Whenever $\Delta F_\beta(\rho) \geq \Delta F_\beta(\sigma)$ is fulfilled for two states ρ and σ that are diagonal in the energy-eigenbasis, then ρ can be mapped to σ by a correlated-catalytic operation arbitrarily well. Thus, correlated-catalytic free transitions completely characterize the non-equilibrium free energy and vice-versa. In chapter 8 we will see that this result also has important implications for the problem of equilibration in many-body physics: It precludes the existence of a "second law of Equilibration" in a well-defined sense.

6 Statistical ensembles from thermodynamic operations

IN THE PREVIOUS CHAPTERS, we have used the framework of catalytic thermal operations to derive the respective results. This approach operates on the level of quantum states of all the involved systems: all the thermodynamic state functions and thermodynamic bounds derived in this formalism are functions of the actual, microscopic quantum state. This leads to additional, more restrictive laws of thermodynamics when compared to the case of phenomenological thermodynamics. In particular, we found that if one precisely accounts also for the build-up of correlations, many different entropy functions play a significant role. Of course, such a treatment is extremely idealized. In many situations, it would be both infeasible as well as undesirable to keep track of the exact microstates of the systems at hand, but one would only have partial information about the physical systems at hand.

In this chapter [1], I introduce and discuss a resource-theoretic formulation of such a scenario under partial information. We assume that the underlying physical dynamics is given by unitary quantum mechanics as before and explore what kind of thermodynamic laws emerge when an experimentalist, which we again call Alice, only has limited knowledge about the quantum systems at hand, in the form of the expectation values of a small set of observables, and is constrained by the conservation of energy on the level of expectation values. Since the precise microstates of the systems are unknown to Alice, any thermal machine that she builds then has to work for *all* the microstates that are compatible with the partial information. In other words, such thermal machines shall be independent of the precise microstates, both for the working system as well as for any heat bath that she uses.

To make this more explicit, imagine Alice having some thermal machine that can transform a system in one particular state ρ with average energy e to a state $\rho^{(f)}$ with average energy $e^{(f)}$. To achieve this, the thermal machines uses access to a thermal environment at some temperature. But for some reason, Alice cannot know the precise quantum state of the system, but at any time only knows the average energy of the system. It can then happen that the actual state of the system is σ, still with the same average energy e, Alice inputs the system into the machine and the machine produces a state $\sigma^{(f)}$, which is completely different from $\rho^{(f)}$. In particular, the state $\sigma^{(f)}$ could have a different average energy than $e^{(f)}$. In such a case, the thermal machine would be completely useless for Alice, since she would never know what the machine would do to a system that it acts upon.

In this chapter, we therefore demand that the machine has to produce the same outcome for *all* states ρ that have the same average energy e. This introduces an additional restriction on what such machines can do. Furthermore, we also do not assume that the thermal environment is described by Gibbs states, but that the machine also has to work if only the

average energy of the systems from the environment corresponds to the energy of thermal systems at the corresponding temperature. We then derive what kind of state transformations can be implemented by Alice using such thermal machines. The main result in this chapter can be stated informally in the following way:

> *Alice can do exactly those state-transformations that she could do if the microstates of all the systems at hand were given by maximum entropy ensembles compatible with the partial information.*

From this statement then the usual phenomenological laws of thermodynamics can be derived on the basis of maximum entropy ensembles.

For example, in the case discussed above, where only the average energy is known, it would then be completely equivalent, in terms of a precise, operationally defined notion of equivalence that I will introduce below, to assign thermal states of the form

$$\omega_{\beta(e)}(H) \tag{6.1}$$

to the system, where the inverse temperature $\beta(e)$ is chosen such that $\mathrm{Tr}(\omega_{\beta(e)}(H)H) = e$. This result provides a new justification of assigning maximum entropy ensembles, like $\omega_{\beta(e)}(H)$, to systems of which one only has partial information about and with which one wants to implement thermodynamic operations. As an immediate consequence we will see that Alice can bring a system from one *macrostate* (e, H), described by the average energy e with respect to the Hamiltonian H, to a different *macrostate* (e', H) if and only if the Clausius inequality is fulfilled, where the entropy is given by the von Neumann entropy of the corresponding thermal states $\omega_{\beta(e)}(H)$ and $\omega_{\beta(e')}(H)$.

6.1   Macrostates and macrostate operations

I now formalize the notion of thermodynamic operations under partial information. For simplicity, I only discuss the case where the partial information about the system is given by the expectation value of the average energy. The formalism can straight-forwardly be generalized to situations in which a macrostate is defined by the expectation value of some commuting set of observables (for details see Ref. [11]). In this case the thermal states $\omega_\beta(H)$ are replaced by the *Generalized Gibbs ensemble* (GGE)

$$\omega_{\vec{\lambda}}(\{Q^{(j)}\}) = \frac{e^{-\sum_j \lambda_j Q^{(j)}}}{Z}, \tag{6.2}$$

where the operators $Q^{(j)}$ denote the commuting, conserved quantities (charges) and λ_j take the role of generalized inverse temperatures, which have to be chosen in such a way that the expectation values of the charges $Q^{(j)}$ match with prescribed values $q^{(j)}$ (see chapter 10 for an introduction to Generalized Gibbs ensemble and a discussion of thermodynamics under Generalized Gibbs ensembles in closed quantum systems).

In the following we consider finite-dimensional quantum systems. We assume that every system has assigned to it a Hamiltonian H. A *macrostate* is then the assignment of an energy expectation value e to the system. We can always view a macrostate as an equivalence class of microstates, namely all those microstates that have energy-expectation value e. A macrostate can hence be considered as the set

$$(e, H) := \{\rho \mid \mathrm{Tr}(\rho H) = e\}, \tag{6.3}$$

where ρ runs over normalised density matrices. We now wish to formulate a precise resource-theoretical framework of thermodynamic operations on the level of macrostates, similar to the thermal operations of the previous chapters. We therefore again assume that an experimentalist has access to arbitrary systems of her choice from the environment, i.e., can choose a Hamiltonian H_E. The average energy of this system is of course fixed since we assume that the system is in equilibrium with the rest of the environment. There is hence some function on Hamiltonians $e_\beta(H_E)$ that determines the energy of systems taken from the environment. I furthermore assume that systems that are taken from the environment

are *uncorrelated* from the working system on the level of *microstates*, i.e., on the level of their density-matrices. By iteratively taking systems from the environment, the most general initial macrostate for a thermodynamic operation is then given by

$$(e, H) \bigotimes_i (e_\beta(H_{E_i}), H_{E_i}). \tag{6.4}$$

Note that the function e_β is completely unspecified so far. In particular, we have not made the assumption that it is given by the energy of a Gibbs state at some inverse temperature β.

Given initial macrostates of the above form, we now first ask which macrostates on the system can be reached by acting with a unitary that preserves the total average energy. Importantly, since we only know the initial macrostates of the systems, *all* compatible microstates have to be mapped to the right macrostate to arbitrary precision. Formally, we can then define *macrostate-to-macrostate operations* in the following way.

Definition 6.1 (Macrostate-to-macrostate operations). *We say that a macrostate (e, H) can be mapped to a macrostate (e', H) by a* macrostate-to-macrostate operation *if for any $\epsilon > 0$ and any $\delta E > 0$ there exist $N(\epsilon, \delta E)$ environment systems with Hamiltonians H_{E_i} and a unitary U (depending on ϵ and δE) such that*

$$\left| \mathrm{Tr} \left(H\, U\, (e, H) \bigotimes_i (e_\beta(H_{E_i}), H_{E_i})\, U^\dagger \right) - e' \right| < \epsilon, \tag{6.5}$$

and the energy is conserved to arbitrary precision:

$$\left| \mathrm{Tr} \left((H + \sum_i H_{E_i}) U\, (e, H) \bigotimes_i (e_\beta(H_{E_i}), H_{E_i})\, U^\dagger \right) - (e' + \sum_i e_\beta(H_{E_i})) \right| < \delta E. \tag{6.6}$$

In the definition we used set notation, which means that the inequalities have to hold for every member of the macrostates with the same unitary U. It is in fact a priori not clear from the definition that any non-trivial state transition on the level of macrostates is possible. In fact, quite the contrary is true. It is possible to show that unless

$$e_\beta(H_E) = \mathrm{Tr}(\omega_{\beta'}(H_E) H_E) \tag{6.7}$$

for some β', *any macrostate can be reached*. This result already shows that the only valid choice of function e_β is given by choosing $e_\beta(H_E) = \mathrm{Tr}(\omega_\beta(H_E) H_E)$, since any other choice completely trivializes the resource theory. The proof-sketch of this result is given in section 14.5.3. Intuitively, the proof works by showing that one can distill from the environment two heat baths at different temperatures and use these to perform any macrostate-to-macrostate transition by running a thermal machine between these two heat baths.

We use this result to fix the energy-function $e_\beta(H) = \mathrm{Tr}(\omega_\beta(H) H)$ for the following considerations and investigate what kind of state-transitions are possible under this assumption. For concreteness, we also assume that $\beta > 0$.

To investigate in more detail what can be done by having initial knowledge only about macrostates, we now introduce a further kind of operations. These are operations that map to *micro*states by requiring that the unitary U not only maps the system always to the same macrostate, but also to the same microstate.

Definition 6.2 (Macrostate operations). *We say that a macrostate (e, H) can be mapped to the microstate $(\rho^{(f)}, H)$ by a* macrostate operation, *if for any $\epsilon > 0$ and any $\delta E > 0$, there exist $N(\epsilon, \delta E)$ environment systems with Hamiltonians H_{E_i} and a unitary U, which in general depends on ϵ and δE, such that*

$$\left\| \mathrm{Tr}_E \left(U\, (e, H) \bigotimes_i (e_\beta(H_{E_i}), H_{E_i})\, U^\dagger \right) - \rho^{(f)} \right\|_1 < \epsilon, \tag{6.8}$$

and the energy is conserved to arbitrary precision:

$$\left| \mathrm{Tr} \left((H + \sum_i H_{E_i}) U\, (e, H) \bigotimes_i (e_\beta(H_{E_i}), H_{E_i})\, U^\dagger \right) - (e + \sum_i e_\beta(H_{E_i})) \right| < \delta E. \tag{6.9}$$
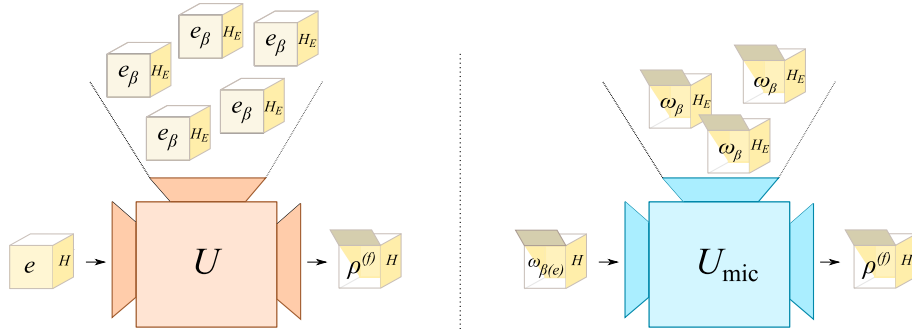
Figure 6.1: The figure illustrates the difference between macrostate operations and microstate operations. *Left:* Illustration of macrostate operations. Closed boxes represent macrostates, i.e., systems where the microstate is unknown. Macrostate of the system and the environment are processed by the energy-preserving unitary U which is required to output a fixed microstate for all possible microstates compatible with the macrostates (i.e., the possible microstates inside the boxes) to arbitrary precision. *Right:* Illustration of microstate operations. In this case one knows the exact microstate of all systems involved and systems from the environment are represented by Gibbs-states at fixed inverse temperature β. Again, we ask whether the initial microstate $(\rho^{(i)}, H)$ can be mapped to the final state $(\rho^{(f)}, H)$ by an energy-preserving unitary U_{micro}. The main result in this section is that U exists if and only if U_{micro} exists for $\rho^{(i)} = \omega_{\beta(e)}(H)$. (Figure adapted from Ref.[11].)

Figure 6.1a) provides an illustration of macrostate operations. If a macrostate (e, H) can be mapped to the microstate $(\rho^{(f)}, H)$ by a macrostate operation, I will in the following write

$$(e, H) \xrightarrow{\text{macro}} (\rho^{(f)}, H). \tag{6.10}$$

The definition of macrostate operations is stricter than that of macrostate-to-macrostate operations. That is, if $(e, H) \xrightarrow{\text{macro}} (\rho^{(f)}, H)$ is possible with $\text{Tr}(\rho^{(f)} H) = e'$, then clearly the macrostate (e, H) can be mapped to the macrostate (e', H) by macrostate-to-macrostate operations.

6.2   Microstate operations and operational equivalence

The main result in this chapter will show that macrostate operations are operationally equivalent to certain situations in which one exactly knows the microstates of the systems at hand. To state the result exactly, I now introduce a further class of operations, which operate on the level of *micro*states. They are essentially identical to thermal operations, with the only difference that we only require that the energy is preserved on average.

Definition 6.3 (Microstate operations). *We say that a microstate $(\rho^{(i)}, H)$ can be mapped to the microstate $(\rho^{(f)}, H)$ by a* microstate operation, *if for any $\epsilon > 0$ and any $\delta E > 0$, there exist $N(\epsilon, \delta E)$ environment systems with Hamiltonians H_{E_i} and a unitary U_{micro}, which in general depends on ϵ and δE, such that*

$$\left\| \text{Tr}_E \left(U_{\text{micro}} \, \rho \bigotimes_i \omega_\beta(H_{E_i}) \, U_{\text{micro}}^\dagger \right) - \rho^{(f)} \right\|_1 < \epsilon, \tag{6.11}$$

and the energy is conserved to arbitrary precision:

$$\left| \text{Tr} \left((H + \sum_i H_{E_i}) U_{\text{micro}} \, \rho^{(i)} \bigotimes_i \omega_\beta(H_{E_i}) \, U_{\text{micro}}^\dagger \right) - (e + \sum_i e_\beta(H_{E_i})) \right| < \delta E. \tag{6.12}$$

Figure 6.1 illustrates microstate operations and compares them to macrostate operations. In the following I write

$$(\rho^{(i)}, H) \xrightarrow{\text{micro}} (\rho^{(f)}, H), \tag{6.13}$$
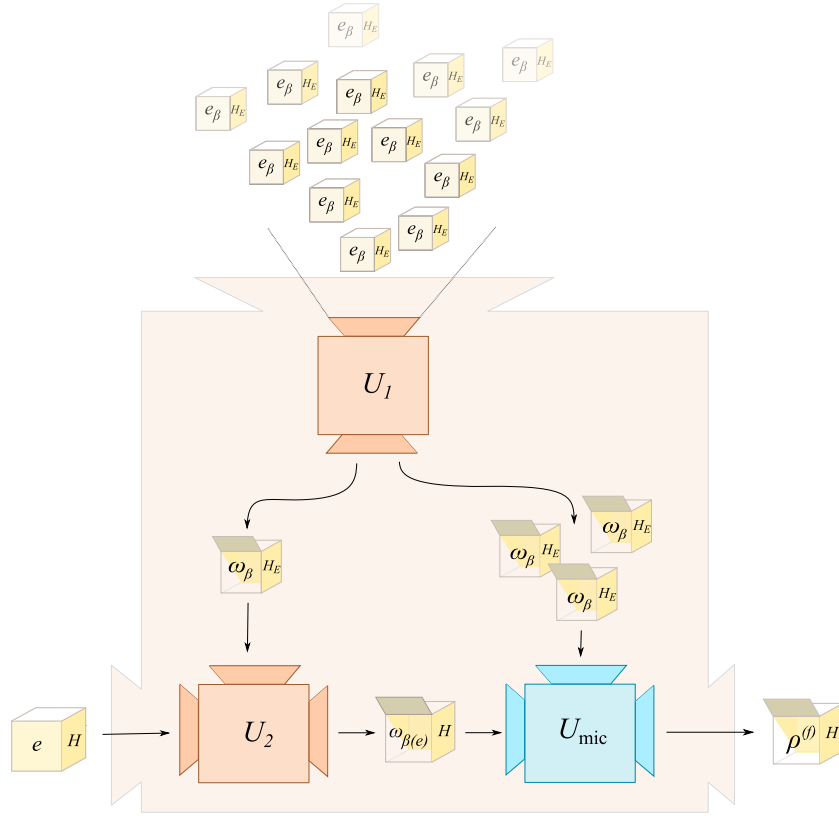
Figure 6.2: Sketch of the proof of theorem 6.5. The proof shows one way how to construct a macrostate operation from a given microstate operation (represented by the unitary U_{micro}). First, the environment is used to distill *microstates* in the maximum entropy ensemble to arbitrary precision from many more thermal *macrostates*. This can be achieved by first randomly mixing in the highly degenerate energy-eigenspaces of many non-interacting systems and then using a central limit theorem together with typicality results for such energy-eigenspaces. The latter step is essentially the usual derivation of the canonical ensemble: The maximally mixed state on a energy-eigenspace of many non-interacting systems is a microcanonical state that locally looks like a canonical state. In the second step, some of these thermal microstates from the environment are used to bring the system from the macrostate (e, H) to the microstate $\omega_{\beta(e)}(H)$ by choosing a suitable environment Hamiltonian and performing a macrostate operation. Finally, the microstate operation U_{micro} is implemented. (Figure adapted from Ref. [11].)

if the microstate $(\rho^{(i)}, H)$ can be brought to the microstate $(\rho^{(f)}, H)$ by a microstate operation. Note that microstate and macrostate operations can be composed: If we have

$$(e, H) \xrightarrow{\text{macro}} (\rho^{(i)}, H) \quad \text{and} \quad (\rho^{(i)}, H) \xrightarrow{\text{micro}} (\rho^{(f)}, H), \qquad (6.14)$$

then we also have

$$(e, H) \xrightarrow{\text{macro}} (\rho^{(f)}, H). \qquad (6.15)$$

In the following, I call a macrostate (e, H) and a microstate $(\rho^{(i)}, H)$ *operationally equivalent* if any microstate that can be reached from $(\rho^{(i)}, H)$ by microstate operations can also be reached from (e, H) from macrostate operations.

Definition 6.4 (Operational equivalence)**.** *A macrostate (e, H) is* operationally equivalent *to a microstate $(\rho^{(i)}, H)$ if for all microstates $(\rho^{(f)}, H)$*

$$(\rho^{(i)}, H) \xrightarrow{\text{micro}} (\rho^{(f)}, H) \quad \text{implies} \quad (e, H) \xrightarrow{\text{macro}} (\rho^{(f)}, H). \qquad (6.16)$$

This equivalence is denoted by $(e, H) \sim (\rho^{(f)}, H)$.

If a macrostate (e, H) is operationally equivalent to a microstate $(\rho^{(i)}, H)$, then in particular we can map the macrostate to $(\rho^{(i)}, H)$:

$$(e, H) \sim (\rho^{(i)}, H) \quad \text{implies} \quad (e, H) \xrightarrow{\text{macro}} (\rho^{(i)}, H).$$

Therefore, one can implement on (e, H) and using macrostate operations exactly the same thermodynamic transitions as on $(\rho^{(i)}, H)$ using microstate operations. Since macrostate operations have to work for all compatible microstates, we further have

$$(e, H) \xrightarrow{\text{macro}} (\rho^{(f)}, H) \quad \text{implies} \quad (\omega_{\beta(e)}(H), H) \xrightarrow{\text{micro}} (\rho^{(f)}, H).$$

Our main result shows that also the converse to the last statement is true.

Theorem 6.5 (Macrostates are operationally equivalent to maximum entropy ensembles)**.**
For any $\beta \neq 0$, macrostates are operationally equivalent to the corresponding maximum entropy ensembles:

$$(e, H) \sim \left(\omega_{\beta(e)}(H), H \right). \tag{6.17}$$

The proof-sketch of this statement is given in section 14.5.2, but figure 6.2 provides a rough illustration of how the proof works.

The theorem shows that macrostate operations are exactly equivalent to microstate operations on maximum entropy ensembles in the sense of thermodynamic transformations. Since microstate operations are simply operations that preserver the average energy and the entropy, we can now use this result to derive the usual laws of phenomenological thermodynamics. To illustrate this, we now derive two standard results: First, the fundamental bound on work extraction in terms of free energies and second the Clausius inequality.

6.3   Example application: Work extraction bounds

Let us consider the following task: We are given a system S and have access to an environment E at temperature $T = 1/\beta$. We want to use the fact that S is not in equilibrium with E to extract work from SE. In phenomenological thermodynamics, such a scenario is described by assigning a temperature T_S to S. The optimal amount of work that can be extracted is then given by

$$W = (e_{T_S} - e_T) - T(S_{T_S} - S_T), \tag{6.18}$$

[2] Here, e_T and S_T are quantities in phenomenological thermodynamics that are a priori unrelated to quantities based on a microscopic description.

where S_T and e_T are the entropy and energy of the system at temperature T, respectively [2].

We can now derive the same bound using macrostate operations. According to theorem 6.5, for this it suffices to consider the situation in terms of maximum entropy ensembles and microstate operations. We therefore assume that S and E are described by maximum entropy ensembles and look for the unitary on SE that reduces the energy on SE as much as possible. We then assume that this change of energy can be extracted as work (a more detailed construction using an explicit work-storage device is of course possible, see [80]). More explicitly, we look for the unitary V and the environment Hamiltonian H_E that maximizes the work

$$W_{\text{value}}(e, H) = \text{Tr}\left[(H + H_E) \left(\omega_{\beta(e)}(H) \otimes \omega_\beta(H_E) - V\omega_{\beta_S}(H) \otimes \omega_\beta(H_E)V^\dagger \right) \right].$$

The resulting optimal work is given by (see, e.g., Ref. [80])

$$W_{\text{value}}(e, H) = \Delta F_\beta(\omega_{\beta(e)}(H), H). \tag{6.19}$$

We can extract this amount of work using macrostate operations in the following way: First, we bring the system from the macrostate (e, H) to the microstate $(\omega_{\beta(e)}(H), H)$ and distill the optimal thermal state $(\omega_\beta(H_E), H_E)$ from the environment. This does not require any work. Then we use the optimal unitary V to extract the work $W_{\text{value}}(e, H) = \Delta F_\beta(\omega_{\beta(e)}(H), H)$. A similar procedure may, in principle, be used for any thermodynamic task that can be described as a state transformation between microstates. As mentioned before, these results can further be generalized to situations involving more conserved quantities, such as the particle number (see Ref. [11]).

6.4   Example application: Deriving the Clausius inequality

As a second illustration, let us discuss how the second law of thermodynamics and the free energy arises in macrostate operations not involving work extraction. More precisely, we will discuss what kind of state transitions are possible under macrostate operations. This can be seen as a generalization of the previous section, by including the work-storage device into the system S. We are thus wondering what is the set of achievable final states $(\rho^{(f)}, H)$ under macrostate operations, i.e., for which microstates is it true that

$$(e, H) \xrightarrow{\text{macro}} (\rho^{(f)}, H)? \qquad (6.20)$$

theorem 6.5 shows that this set of states coincides exactly with the states that are achievable from the microstate $\omega_{\beta(e)}(H)$ using microstate operations. This set of achievable states has been studied in Ref. [80], where it was shown that a transition to $(\rho^{(f)}, H)$ is possible if and only if the free energy decreases, i.e.,

$$(e, H) \xrightarrow{\text{macro}} (\rho^{(f)}, H) \Leftrightarrow \Delta F_\beta(\omega_{\beta(e)}(H), H) \geq \Delta F_\beta(\rho^{(f)}, H). \qquad (6.21)$$

This is the precise expression of the second law as expressed on macrostate-to-microstate operations. In particular, we can also use this result to derive the possible macrostate-to-macrostate transitions. Let us define the *macrostate non-equilibrium free energy* by $\mathcal{F}_\beta(e, H) := F_\beta(\omega_{\beta(e)}(H), H)$, which is a family of state functions on macrostates. Then we can define the derived *macrostate entropy* as

$$\mathcal{S}(e, H) := \beta(e)(e - \mathcal{F}_{\beta(e)}(e, H)) = S(\omega_{\beta(e)}(H)), \qquad (6.22)$$

which is also a state function on macrostates [3]. Using macrostate quantities, we can then use (6.21) to express the condition for macrostate-to-macrostate transitions using the *Clausius inequality*:

$$(e, H) \xrightarrow{\text{macro}} (e', H) \quad \Leftrightarrow \quad \mathcal{Q} \leq T\Delta\mathcal{S}, \qquad (6.23)$$

with the *heat* $\mathcal{Q} := e' - e$ and the change of macrostate entropy $\Delta\mathcal{S} = \mathcal{S}(e', H) - \mathcal{S}(e, h)$. Again, similar relations can be derived in the case of more conserved quantities [11].

[3] We can also interpret the definition of the macrostate entropy as the Legendre-transform of the *equilibrium* macrostate free energy $\mathcal{F}_{\beta(e)}(e, H)$ and hence as a function of the inverse temperature and the Hamiltonian instead of the average energy and the Hamiltonian.

6.5   Conclusions and relation to Jaynes' principle and typicality results

In this chapter, we showed that for the purpose of deriving bounds on what can be done with a quantum system using thermodynamic operations, macrostates are equivalent to maximum entropy ensembles. This provides a novel justification for the use of maximum entropy ensembles.

Importantly, our derivation does not rely on arbitrary probability measures on microstates like the postulate of equal a priori probabilities in statistical mechanics or similar measures in other approaches that derive maximum entropy ensembles from typicality arguments [26, 27, 29, 38, 39, 165–168].

On first sight, our approach seems closely connected to Jaynes' maximum entropy principle [31, 32], which very roughly asserts that the best prediction one can make about a system is by assigning the maximum entropy state compatible with the information one has about the system.

There are important differences between the two approaches, though. First, Jaynes had to first make a definition of the entropy functional that ought to be maximized. While he gave very good reasons to choose the Shannon entropy (or the von Neumann entropy in the quantum case), in principle, it is nevertheless an arbitrary choice. In the approach presented in this chapter, the entropy emerges naturally.

The key idea that makes this possible is that we do not wish to claim that *measurement statistics* of observables are well predicted by maximum entropy ensembles, but that the *laws of state transformations* are identical to situations in which one would assign maximum entropy ensembles to the respective states. This is hence a different kind of justification and can be done under a different set of assumptions.

As a second point, in the approach presented in this chapter, we never need to actually assign a concrete microscopic state to the system, since we require that macrostate operations work for all compatible microstates. The results simply show that, from a thermodynamic perspective, macrostates *could* equivalently be described by maximum entropy ensembles and that they are the only microstates with this property.

In the derivation of this chapter, we have assumed that the energy is conserved *on average* only. One might wonder, whether the results can also be recovered if one assumes that the unitary U that implements a macrostate operation commutes with the Hamiltonian of the system and environment as in the framework of thermal operations. This is impossible, because it is possible to show that if one demands exact commutation with the Hamiltonian, then the possible macrostates that can be reached from a given macrostate (e, H) are characterized by a polytope. This in turn implies that if we look for the largest energy $e^{\max}(e)$ that can be achieved from a given macrostate (e, H) using such commuting macrostate operations, then this functions is piecewise *linear*.

On the other hand, we know that for macrostate operations, which preserve energy on average, $e^{\max}(e)$ is determined by the Clausius inequality and hence by the thermal free energy, which is a non-linear function of the average energy. Therefore, demanding exact energy conservation on the level of probability distributions is too strict a constraint to derive the equivalence between macrostates and maximum entropy ensembles in this framework. For details of this argument see Ref. [11].

Nevertheless, in the case of *macroscopic* systems, we can argue that the energy fluctuations that are necessary to implement macrostate operations can in practice be very small on the relevant energy scale if we allow ourselves to make additional assumptions, which are beyond the formal framework introduced in this chapter. For large systems one may implement a macrostate operation as many independent macrostate operations on the parts of a system. For example, to extract work form a system, we may extract a small amount of work from everyone of the N subsystems independently. If we choose the subsystems to be small on a macroscopic scale, but large on a microscopic scale, we can expect them to be weakly interacting and weakly correlated on the level of microstates. Then one can use a central limit theorem to argue that the fluctuations in work will only grow like $\sqrt{N}$ instead of N and are hence vanishingly small when compared to the total extracted work.

In this chapter, and also in the previous chapters, we have always assumed the existence of thermal equilibrium in some form. In the following two chapters, let us now discuss when and how such equilibrium actually comes about from a dynamic point of view.

7 Equilibration in many-body systems

In the previous chapters, we already discussed a large variety of aspects of thermodynamics. To do so, we employed to large extend a resource theoretic point of view. In these approaches, we simply assumed the existence of thermal equilibrium. For most parts, we furthermore assumed that systems in thermal equilibrium are well described by statistical ensembles – apart from chapter 6, where we derived statistical ensembles using an argument based on thermodynamic operations.

Similarly as in phenomenological thermodynamics we therefore did not take into account any considerations about the actual time-dependent dynamics involved in any of the processes. In particular, we did not discuss how systems actually reach thermal equilibrium in a dynamical way and when we can expect that such thermal equilibrium can indeed be described by Gibbs states from a dynamical perspective.

In the following two chapters, I discuss in more detail when and how complex quantum systems can be said to reach a state of equilibrium from different point of views. Apart from being interesting in its own right from the point of view of non-equilibrium dynamics, the first chapter (in particular the first two sections), also provide the necessary background to understand chapters 10 and 11 that deal with thermodynamics in closed quantum systems using Generalized Gibbs ensembles and thermodynamics in the strong coupling regime. I first briefly review some general results about equilibration in complex many-body systems. Then, in section 7.2 I describe in much more detail how this equilibration behaviour can be understood as a mechanism of dephasing. These discussions are illustrated using numerical simulations of well-known models and I also connect the necessary assumptions with rigorous results in mathematical physics. The aim of this chapter is to provide an intuitive understanding of the mechanism of equilibration and connect it with general results in many-body theory.

In the next chapter, I will connect the problem of equilibration to a resource theoretic view (although not in the framework of thermal operations): I will provide arguments that show in a quantitative way that it is difficult to prepare a large complex many-body system in a state that does *not* equilibrate. Furthermore, I will discuss in how far such a *resilience to equilibration* can be seen as a thermodynamic resource with an appropriate corresponding "second law of equilibration".

7.1   Brief review of equilibration and thermalization in closed quantum systems

When a large many-body system is brought out of equilibrium, physically relevant observables usually quickly come to a rest again. This process, called *equilibration*, is ubiquitous in nature. A typical example in the context of quantum physics is nowadays studied in

so-called *quench-experiments* [169–174]. In such an experiment, a quantum system is prepared in a stationary state of some Hamiltonian. Then the Hamiltonian is suddenly modified, leading to a situation in which the quantum system is not in equilibrium with respect to the Hamiltonian anymore. After this sudden change of Hamiltonian, called *quench*, the relaxation dynamics back to equilibrium can be studied. For example, in optical lattices the quench could simply be a sudden change of the lattice-depth.

Despite the fact that equilibration is ubiquitous, it is somewhat surprising from a purely quantum mechanical point of view: After all, a closed system evolves unitarily and hence there are always observables which never approach a stationary value. How can it then be that physically relevant observables do indeed show equilibration? And can one understand and characterize in which kind of systems this behaviour takes place?

This question has in fact already been considered early on by some of the founders of quantum mechanics, von Neumann [29] and Schrödinger [27]. They already observed that such equilibration behaviour should be a very typical behaviour of large interacting systems. Nevertheless, general and rigorous results that prove equilibration behaviour under general assumptions mostly only emerged over the last decade [42, 167, 168, 175–191]. These results have been backed up by numerous numerical studies (see the reviews [40, 41] and references therein) and a large body of work on integrable systems [192–202] showing that equilibration is indeed a very generic feature. The revival of the problem of equilibration can be seen as caused both by new analytical tools that became available leading to rigorous proofs of equilibration behaviour as well as the new possibilities to probe such behaviour experimentally in quench experiments as described above.

In this section, I want to briefly review the simplest general results showing equilibration under general assumptions [1]. For the remainder of this and the following two sections, we consider a large quantum system described by a Hamiltonian $H = \sum_{i=1}^{d_E} E_i P_i$ on a d-dimensional Hilbert-space $\mathcal{H}$. Here, the E_i denote the energy eigenvalues and P_i the projectors onto the corresponding eigenspaces. When discussing equilibration, I usually have in mind a local, interacting Hamiltonian on a lattice of spins, but the general results in this section do not depend on that. Therefore for now I will not specify the Hamiltonian much further to keep the discussion as general as possible.

It is clear that for a system to equilibrate, there must not be any part of the system that completely decouples from the rest of the system. One way to ensure this is to demand that the Hamiltonian has *non-degenerate energy-gaps*, meaning that not only are there no degeneracies in the spectrum of the Hamiltonian, but also every difference of energy eigenvalues $E_i - E_j$ is unique. Let us denote by $\mathcal{G}(\Delta)$ the set of pairs of energies whose difference is Δ:

$$\mathcal{G}(\Delta) := \left\{ (E_i, E_j) \mid E_i - E_j = \Delta \right\}. \tag{7.1}$$

Then formally the non-degenerate energy-gaps condition means that every set $\mathcal{G}(\Delta)$ has at most one element as long as $\Delta \neq 0$. Let us, for later convenience, also define the set of non-zero gaps

$$\text{Gaps}(H) := \left\{ E_i - E_j \mid i \neq j = 1, \ldots, d_E \right\}. \tag{7.2}$$

The condition of non-degenerate energy-gaps implies that it is impossible to find a partition of the Hilbert-space $\mathcal{H}$ into tensor-factors $\mathcal{H}_1$ and $\mathcal{H}_2$ which makes the Hamiltonian non-interacting over this partition, i.e., for which

$$H = H_1 \otimes \mathbf{1}_{\mathcal{H}_2} + \mathbf{1}_{\mathcal{H}_1} \otimes H_2. \tag{7.3}$$

Indeed, if we could find such a partition, then every energy gap of the Hamiltonian H_1 would have to have a degeneracy of at least the dimension of the Hilbert-space $\mathcal{H}_2$, and vice-versa. For the results in the rest of the subsection, I assume for simplicity that the Hamiltonians in question all have non-degenerate energy gaps. While this condition is generically fulfilled (i.e., can be assured by an arbitrarily small random perturbation), it is often too strong. Indeed, much of the following results can be generalized to situations where the degeneracy of energy-gaps is not too large [42, 182]. That the condition of non-degenerate energy-gaps is not always necessary for the equilibration of many physically

[1] For a very thorough recent review see Ref. [42].

relevant observables can also be seen from the fact that local observables in non-interacting systems, described by local Hamiltonians of free fermions or free bosons, often also show equilibration behaviour [194, 196–199].

Now consider an arbitrary initial state $\rho(0)$, which might in fact be a pure state $\rho(0) = |\Psi\rangle\langle\Psi|$. The state evolves unitarily under the dynamics generated by H, evolving along a trajectory $\rho : t \mapsto \rho(t)$. Suppose we consider an arbitrary observable A. If the expectation value of the observable $\langle A \rangle_{\rho(t)} = \mathrm{Tr}(A\rho(t))$ equilibrates, then it has to equilibrate to its time-average $\mathbb{E}_t \left[\langle A \rangle_{\rho(t)} \right]$ given by

$$\mathbb{E}_t \left[\langle A \rangle_{\rho(t)} \right] = \lim_{T\to\infty} \frac{1}{T} \int_0^T \langle A \rangle_{\rho(t)} \mathrm{d}t = \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathrm{Tr}(A\rho(t))\mathrm{d}t = \langle A \rangle_{\omega(\rho)}, \quad (7.4)$$

where we have defined the time-averaged state $\omega(\rho)$ given by

$$\omega(\rho) = \mathbb{E}_t \left[\rho(t) \right] = \lim_{T\to\infty} \frac{1}{T} \int_0^T \rho(t)\mathrm{d}t = \sum_{j=1}^{d_E} P_j \rho(0) P_j. \quad (7.5)$$

Since we consider a finite system, however, perfect equilibration in the sense that $\rho(t)$ becomes stationary is impossible, due to recurrences [203]. Nevertheless, $\rho(t)$ can be very close to $\omega(\rho)$ for most times. In this case the expectation value $\langle A \rangle_{\rho(t)}$ would be very close to the time-average $\langle A \rangle_{\omega(\rho)}$ for most times, possibly with occasional larger deviations for short times. This behaviour is indeed what happens provided that the initial state has overlap with sufficiently many energy-eigenstates.

To quantify this behaviour, let us introduce the *effective dimension* as a measure for how many energy-eigenstates participate in the initial state. It is defined as [177]

$$d_{\mathrm{eff}}(\rho) := d_{\mathrm{eff}}(\rho(0), H) := \frac{1}{\sum_j \mathrm{Tr}(P_j \rho(0))^2} = \frac{1}{\mathrm{Tr}(\omega(\rho)^2)}. \quad (7.6)$$

This quantity is typically extremely big. For example, consider a micro-canonical state on a large many-body system, which is evenly distributed (coherently or as a statistical mixture) over an interval of energies $[E - \delta E, E + \delta E]$. For a fixed energy-uncertainty δE, the number of energy-eigenstates that contribute to this initial state will in general grow exponentially with the system size for local many-body systems. Hence, the effective dimension will also grow exponentially with the system size. At the same time, the uncertainty in terms of the *energy-density* will go to zero. This is even true if δE grows sub-linearly with the system-size, which would further increase the effective dimension.

For a second example, it has been shown that essentially any state with a finite correlation length that is not an eigenstate has an effective dimension that diverges with the system-size [190] (this will be discussed in more detail in section 7.2).

The following theorem then rigorously shows that equilibration indeed happens in the sense discussed above.

Theorem 7.1 (Equilibration on average [204]). *Let $\rho(t)$ evolve under an Hamiltonian H with non-degenerate energy-gaps. Then for any observable A we have*

$$\mathbb{E}_t \left[(\langle A \rangle_{\rho(t)} - \langle A \rangle_{\omega(\rho)})^2 \right] \leq \frac{\|A\|^2}{d_{\mathrm{eff}}(\rho)}. \quad (7.7)$$

For large many-body systems and "natural" initial states, we thus see that for most times, the state $\rho(t)$ is practically indistinguishable from $\omega(\rho)$ by observables with a finite norm (bounded by a constant independent of the system size).

A particular case of physically relevant observables are those that only act on a subsystem S of the total system. They are completely characterized by the reduced density matrix $\rho_S(t) = \mathrm{Tr}_B(\rho(t))$. The following theorem shows that this state is effectively indistinguishable from the corresponding reduced density matrix of the time-averaged state $\mathrm{Tr}_B(\omega(\rho))$ as measured by the trace-distance $\mathcal{D}(\cdot, \cdot)$.

Theorem 7.2 (Average equilibration of subsystems [177]). *Let $\rho(t)$ evolve under an Hamiltonian H with non-degenerate energy-gaps. Then for any subsystem S we have*

$$\mathbb{E}_t\left[\mathcal{D}\left(\rho_S(t), \mathrm{Tr}_B(\omega(\rho))\right)\right] \leq \sqrt{\frac{d_S^2}{d_{\mathrm{eff}}(\rho)}}. \tag{7.8}$$

The trace-distance $\mathcal{D}(\rho, \sigma)$ measures the theoretical statistical distinguishability of ρ and σ in a single shot, i.e., it gives the maximum probability that the two can be correctly distinguished by performing a single measurement.

Given the above comments about the effective dimension, we thus conclude that bounded and local observables in large, complex quantum systems equilibrate in a very stringent sense. Similar theorems can also be shown for more general measurements in terms of POVMs [188]. Furthermore, it can be shown that the fluctuations around equilibrium are slow [179].

As a side remark, note the effective dimension is effectively a measure of the amount of entropy in the probability distribution of energy in a quantum system. Indeed, it can also be expressed as

$$d_{\mathrm{eff}}(\rho) = e^{S_2(\omega(\rho))}, \tag{7.9}$$

where S_2 is the Rényi-2 entropy. The Rényi entropies are defined as

$$S_\alpha(\rho) := \log(d) - D_\alpha(\rho\|\mathbf{1}/d) = \frac{1}{1-\alpha}\log(\mathrm{Tr}(\rho^\alpha)), \tag{7.10}$$

where D_α are again the Rényi-divergences. From an information theoretic point of view it is then interesting to ask why it is not the von Neumann entropy of $\omega(\rho)$ that appears in the equilibration bounds, which is a more natural entropy measure. It is thus tempting to try to prove similar equilibration bounds expressed in terms of the von Neumann entropy of $\omega(\rho)$. We will see in chapter 8 that this is in fact impossible: There are states which do not equilibrate, but where the von Neumann entropy of $\omega(\rho)$ diverges with the system size. This also has interesting consequences relating to the role of correlations in the process of equilibration, which we will discuss in chapter 8.

The above results clearly establish that we should expect equilibration to be a generic phenomenon. However, it is also important to note what these results do not show. First, they don't say anything concrete about the equilibrium state $\omega(\rho)$. In particular, they do not imply that this equilibrium state is well described by a thermal state of H. Additional assumptions are necessary to formally prove such behaviour [190, 205]. Even in this case, usually referred to as *thermalization*, it is important to observe that equilibration to a thermal state only implies that local subsystems equilibrate to the reduced state of a thermal state,

$$\rho_S(t) \approx \mathrm{Tr}_B\left(\frac{e^{-\beta H}}{Z_\beta}\right), \tag{7.11}$$

but *not* to the thermal state of the Hamiltonian H_S of the subsystem S. Indeed, for strongly interacting systems it is not even clear how one would unambiguously define such a local system Hamiltonian H_S (see, however, Ref. [206]). In chapter 11, we will explore the consequences of this for thermal machines in the strong coupling regime.

Second, the above theorems do not rule out the possibility of having initial states for which relevant observables do *not* equilibrate. All one needs is an initial state with a small effective dimension that is not diagonal in the energy-eigenbasis. However, as the examples for the effective dimension given above already suggest, it seems difficult to prepare such states, since it would seem to require extremely high control over a large, complex quantum system. In chapter 8, I will provide quantitative arguments which show that it is indeed very difficult to prepare quantum states on complex quantum systems that do not equilibrate.

Third, the above equilibration bounds do not say anything about *how long* it actually takes to reach equilibrium. While general bounds on such an *equilibration time* can be derived [186], the general bounds diverge exponentially with the size of the *total system*.

Moreover, one can construct counter-examples which essentially saturate these bounds [183]. Therefore, it is impossible to prove, without making more concrete physical assumptions, that bounded observables equilibrate in finite time also when one takes the thermodynamic limit. In the next sections, we will therefore discuss this problem of equilibration times in more detail. In particular, I will argue for a simple mechanism that suggests that generic systems in fact equilibrate very rapidly, in a time independent of the total system size, whereas the precision with which they equilibrates increases with the total system size.

7.2   Equilibration as dephasing and the problem of timescales

In remainder of this chapter[2], I discuss the problem of equilibration from a more concrete point of view. The aim of this section is to give an intuitive understanding of equilibration and to make plausible that in many relevant cases the expectation value of an observable indeed equilibrates in a time that is essentially independent of the system size.

I will not present any rigorous theorem about equilibration times, but I will connect the intuitive discussion to rigorous results in mathematical physics and illustrate the arguments using numerical examples.

In this section, I focus on the equilibration of expectation values of some observable A, given an initial state ρ and a Hamiltonian H. The figure of merit that we will be considering to discuss equilibration is given by the instantaneous deviation from the equilibrium value, measured by the quantity

$$\Delta A_\rho(t) := \langle A \rangle_{\rho(t)} - \langle A \rangle_{\omega(\rho)}. \tag{7.12}$$

Roughly speaking, we can say that the system equilibrates if $\Delta A_\rho(t)$ becomes very small after some time and remains to be small afterwards. Of course, by varying the Hamiltonian H, the initial state ρ and the observable A we can in principle describe a vast range of different physical situations. We should therefore first specify more clearly in what kind of phenomena we are interested in.

In the following sections, we will be concerned mostly with the equilibration of expectation values of locla observables in a many-body system that is initially not at equilibrium. After such a local equilibration, which we expect to happen rather quickly, different parts of the system might in general still be out of equilibrium on macroscopic length and time scales. The remaining equilibration on these scales, which can be expected to be effectively described by a semi-classical, hydrodynamic approach [207, 208], is not in focus in this section. We therefore simply assume that the system is initially homogeneous and evolves under a roughly translational invariant Hamiltonian.

We thereby exclude all physical phenomena that happen on macroscopic length- and time-scales, including transport effects on such scales, such as heat conduction or electrical currents. This might seem like an overly strong restriction, but even this restricted scenario is not yet completely understood from a theoretical point of view. In particular, there exist no rigorous proofs that show that such an equilibration happens after a time that does not diverge with the system size under a general, but reasonable set of assumptions.

Indeed, such a proof can be expected to be extremely difficult, if not impossible, due to the fact that there are very reasonable interacting Hamiltonians, with reasonable initial states and simple observables, which seem to stubbornly refuse to equilibrate. We will see such an example later.

Nevertheless, in the following I will argue that the basic mechanism of equilibration is in fact quite simple and this mechanism suggests that such a refusal to equilibrate should only happen in very peculiar circumstances, while local equilibration generically happens quickly.

It is important to emphasize that this restriction to somewhat homogeneous systems is not due to the fact that it is expected that systems which are not translational invariant would not equilibrate. The following simple argument suggests that the basic picture of local equilibration also holds for systems which are not homogeneous on larger scales. If we know that a local observable equilibrates in time t_{eq} in a translational invariant system, then *Lieb-Robinson Bounds* (see section 7.5.4) show that the observable (in the Heisenberg-picture)
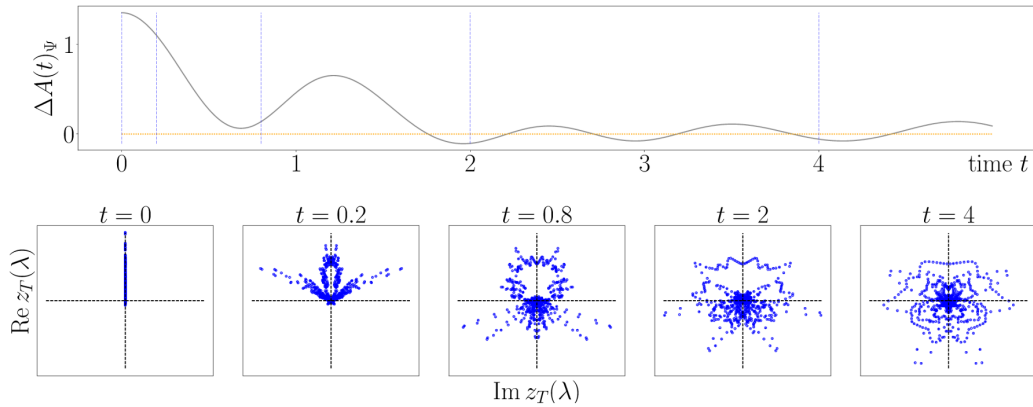
Figure 7.1: Time evolution of an equilibrating system. The model is the transverse field Ising model with Hamiltonian $H_{\text{Ising}} = \sum_{j=1}^{L} (J\sigma_j^x \sigma_{j+1}^x + h_x \sigma_j^x + h_z \sigma_j^z)$ with periodic boundary conditions. Here, $L = 15$ sites with parameters $J = 4, h_x = 1, h_z = -2.1$. The initial state Ψ is a random product state with a spin-up state in the middle of the chain and the observable is a σ^z in the middle of the chain as used in Ref. [211]. The lower panel shows the evolution of the individual terms contributing to the Fourier transform of the distribution z_T in the complex plane. For numerical reasons $z_T(\lambda)$ (here $T \approx 33$) is evaluated at 5000 points λ_i which linearly interpolate the minimal and maximal gap occurring for the respective system size. The discretization of z_T is hereby ensured to approximate the regularized function well and we verified, for instance, that $\sum_i z_T(\lambda_i)$ agrees up a relative error of 10^{-8} with $\Delta A(0)_\rho$. Therefore no significant weight of the distribution is lost. The lower figures show the values of $z_T(\lambda_i)e^{i\lambda_i t}$ at the times marked in the evolution in the upper panel. For the computation of $z_T(\lambda)$ all values z_Δ with $|\Delta| < 10^{-13}$ are discarded in order to account for the subtraction of the infinite time averaged expectation value. While initially strongly localized and anisotropic, the time evolved $z_T(\lambda_i)e^{i\lambda_i t}$ quickly relaxes into a roughly isotropic distribution that is, up to small fluctuations, constant in time. The plot of the distribution z_T over λ shown in Fig. 7.3 (top) accordingly shows a large number of different gaps (and therefore angular velocities) that carry about equal weights of the distribution z_T. This is directly reflected in the time evolution of $\Delta A(t)_\rho$ which decreases in time as the distribution spreads. (Figure from Ref. [10].)

has essentially only "visited" a sub-system of size $v_{\text{LR}} t_{\text{eq}}$, where v_{LR} is the maximum group velocity in the system implied by the Lieb-Robinson bounds. Thus, if the system is inhomogeneous on scales larger than $v_{\text{LR}} t_{\text{eq}}$, we can expect that rapid, local equilibration still holds, possibly followed by a much slower equilibration/homogenization on larger scales.

For example, many-body localized systems equilibrate (but not thermalize), although they break translational invariance [209, 210].

7.3   The basic mechanism

We now discuss equilibration from the point of view of dephasing (see Ref. [213] for a similar exposition that appeared simultaneously with our manuscript [10]). To understand this basic mechanism behind equilibration, it is useful to re-express the instantaneous deviation from equilibrium $\Delta A_\rho(t)$ in the energy-eigenbasis as

$$\Delta A_\rho(t) = \sum_{E_i \neq E_j} \text{Tr}\left(P_i A P_j \rho\right) e^{i(E_j - E_i)t} = \sum_{0 \neq \Delta \in \text{Gaps}(H)} z_\Delta e^{i\Delta t}, \qquad (7.13)$$

where we have introduced the complex numbers

$$z_\Delta = \sum_{(E_i, E_j) \in \mathcal{G}(\Delta)} \text{Tr}\left(P_i A P_j \rho\right). \qquad (7.14)$$

Since A is hermitian and the time-evolution is unitary, the numbers z_Δ fulfill the relation $z_{-\Delta} = \overline{z_\Delta}$, where the overline denotes complex conjugation.

With this expression at hand, we can now develop a basic understanding of how equilibration comes about. In a large quantum system, the number of different gaps Δ grows exponentially with the system size. For generic observables and states, we can therefore expect that the number of terms in the sum in $\Delta A_\rho(t)$ is huge. We can visualize this, by drawing a point for each z_Δ in the complex plane. We then obtain a cloud of points in the complex plane, each rotating with a different angular velocity Δ on a circle of radius $|z_\Delta|$
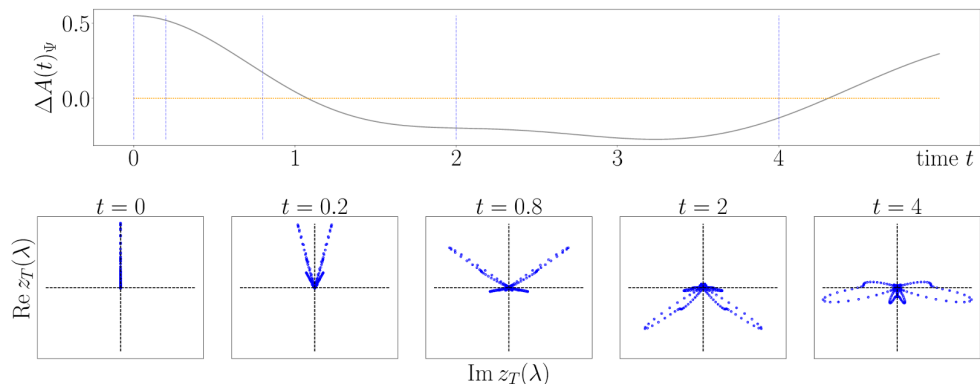
Figure 7.2: Time evolution in a system failing to equilibrate. We show the exact time evolution of $\Delta A(t)_\rho$. The model is the transverse field Ising defined as in Fig. 7.1 on $L = 15$ sites, but with parameters $J = 1, h_x = 0.5, h_z = -1.05$. The initial state Ψ is a product state composed of only spin-up states and the observable is a σ^z operator in the middle of the chain as used in Ref. [212]. As in Fig. 7.1, the lower plot shows the evolution of the contributions to the Fourier transform of the smoothed distribution z_T in the complex plane at the times marked in the evolution. We apply the same scheme as described in the caption of Fig. 7.1, again with 5000 interpolation points, $T \approx 33$ and treating gaps $|\Delta| < 10^{-13}$ as zero. Initially, the smoothed z_T is strongly localized and anisotropic. When evolved in time, we find two distinct and large contributions that revolve around the zero without canceling out one another. These contributions stay approximately in phase and do not disperse strongly as their parts revolve with roughly the same angular velocity. This agrees well with the result shown in Fig. 7.3, which displays the distribution z_T in dependence of λ and shows two distinct peaks concentrating most of the weight of the distribution. As a result, the deviation from the steady-state expectation value shows strong oscillations, which only weakly decay. (Figure from Ref. [10].)

(see Fig. 7.1). If the system is initially out of equilibrium, $\Delta A(0) \gg 0$, this cloud of points is not distributed isotropically. Since each point moves with a different angular velocity, the cloud of points necessarily starts to spread and distribute more isotropically. But an isotropically distributed cloud of points in the complex plane has a small total absolute value, hence $|\Delta A(t)| \ll 1$ and the system equilibrated.

Once the points are spread out approximately isotropically, we can expect that they remain so for a long time: There are vastly more configurations which remain approximately isotropic for some time under the following time-evolution than configurations which yield a sudden anisotropic distribution again. Nevertheless, there will be rare fluctuations and, importantly, a recurrence time like in any finite system [203], which sometimes bring the system out of equilibrium again. As the system size, and hence number of points, increases, however, we can expect that the fluctuations will become increasingly rare and small and the recurrence time will diverge with the system size, explaining why in large systems, equilibration is essentially perfect.

In finite systems with local interactions, besides the recurrence time there is an additional time-scale which yields a recurrence-like behaviour and which dominates in any finite-system numerics: information can be transported ballistically and when it reaches the boundaries of the system it is backscattered. This can be seen in $\Delta A(t)$ as a perturbation on time-scales which increases linearly with the system size.

On the other hand, the time t_{eq} it takes until the system has equilibrated to precision ϵ, i.e., $|\Delta A_\rho(t_{eq})| \leq \epsilon$, essentially only depends on the distribution of Δ and the shape of the distribution of the z_Δ as a function of Δ. If these distributions become roughly independent of the system size as it increases, also the equilibration time t_{eq} should become roughly independent of the system size as the system size increases. Similarly, we can then expect that for a fixed time larger than t_{eq}, the size ϵ of the remaining oscillations of $\Delta A_\rho(t)$ should decrease with the system size, while the recurrence-time increases with the system size.

To make these claims clearer, it is useful to consider a very simple toy-model for the mechanism described above. The assumptions that we will make for this toy-model can be motivated from the general results in many-body theory that we will review in the later sections. Let us choose a large number N of gaps Δ_i uniformly at random from some interval $[-\Delta_{max}, \Delta_{max}]$ (this choice will be motivated in sections 7.5.1 and 7.5.3) and for simplicity assume that the corresponding z_{Δ_i} are real and follow a Gaussian distribution

$\mathcal{G}_{1/\tau}(\Delta)$ with variance $1/\tau \ll \Delta_{\max}$. In the following we assume that we chose a typical distribution of the gaps Δ_i. We then normalize the z_{Δ_i} to ensure $\sum_i z_{\Delta_i} =: \Delta\tilde{A}(0)_\rho$ is fixed and of order one (independent of N)[3]. Here, we have invented an imaginary observable $\tilde{A}$ and a state ρ that realize the distribution of z_{Δ_i}. Of course this is only to keep the notation in analogy to the previous and the subsequent sections.

As we increase N, the time-dependent expectation $\Delta\tilde{A}(t)$, given by

$$\Delta\tilde{A}(t)_\rho = \sum_i z_{\Delta_i} e^{i\Delta_i t}, \tag{7.15}$$

then approximates to higher and higher precision the Fourier-transform of the Gaussian $\mathcal{G}_{1/\tau}(\Delta)$, since we are essentially sampling the corresponding integral. Thus

$$\Delta\tilde{A}(t)_\rho \approx \int \mathcal{G}_{1/\tau}(\Delta) e^{i\Delta t} d\Delta. \tag{7.16}$$

This approximation will have some error, and for any given error ϵ there will be a maximum time $T_N(\epsilon)$ such that this approximation remains valid. For small times $t < T_N(\epsilon)$ we then get

$$\Delta\tilde{A}(t)_\rho \approx_\epsilon \Delta\tilde{A}(0) e^{-(t/\tau)^2}, \quad t < T_N(\epsilon). \tag{7.17}$$

The time $T_N(\epsilon)$ increases with increasing N for any fixed ϵ. We can hence identify the equilibration time-scale as τ and define the equilibration time t_{eq} as $t_{eq} = C\tau$ for any constant $C > 0$ which controls the precision of equilibration that we demand.

In more realistic set-ups, the distribution of z_Δ also has features on smaller scales than $1/\tau$, resulting in correspondingly longer equilibration times. However, many of these features can be expected to have small weight, and hence only yield small and slow oscillations before they equilibrate. Importantly, while the equilibration times for these features is much longer, they are still largely independent of N as long as the distribution z_Δ is essentially independent of N. In particular, they should not diverge with the system size. The actual problem of explaining equilibration is therefore to explain when and why the distribution of z_Δ becomes essentially independent of the system size for large systems.

In the previous discussion, we used a very simplistic model with various implicit assumptions. With this intuition we can now start to discuss the problem of equilibration in more detail and develop some formal machinery. We then discuss the various assumptions that go into the analysis and connect them with known rigorous results in many-body theory.

7.4   Refining the notion of equilibration

As we have seen before, to sensibly discuss equilibration we need to consider different time-scales and always have to consider a finite precision of equilibration: The system will only equilibrate to some finite precision and only stay equilibrated for a finite but long time (which we expect to diverge with the system size). In the following, we therefore only ask whether equilibration occurs to precision ϵ before some chosen cut-off time T. Both quantities in general depend on the system size and we expect that for a fixed ϵ, the cut-off time T diverges with the system size.

We can formalize this procedure by introducing regularized quantities. The crucial quantity that governs the dynamics of the chose observable is the distribution of the z_Δ, which is a discrete distribution for any finite system. We now regularize this quantity into a smooth distribution by convoluting it with a Gaussian of variance $1/T$, where T is the cut-off time. We hence define the function

$$z_T(\lambda) := \sum_{0 \neq \Delta \in \text{Gaps}(H)} z_\Delta \mathcal{G}_{1/T}(\lambda - \Delta). \tag{7.18}$$

Note that we can recover the non-regularized distribution as

$$z_\infty(\lambda) = \sum_{0 \neq \Delta \in \text{Gaps}(H)} z_\Delta \delta(\lambda - \Delta). \tag{7.19}$$

[3] With this construction, $\sum_i z_{\Delta_i}$ can be a complex number. But its imaginary part will be extremely small for large N, we hence ignore this problem.

Figure 7.3: System size scaling of the regularized z_T distribution. At all system sizes the distribution z_T is again evaluated at 5000 points interpolating the extremal gaps linearly for $T \approx 33$. *Top:* Equilibrating model discussed in Fig. 7.1. With growing system size, the resulting distribution z_T spreads its weights more and more evenly leading to the equilibrating behavior of the considered system as shown in Fig. 7.1. *Bottom:* non-equilibrating model discussed in Fig. 7.2. The resulting distribution z_T concentrates most of the weight in two localized peaks which yield the non-equilibrating behavior shown in Fig. 7.2. (Figure adapted from Ref. [10].)
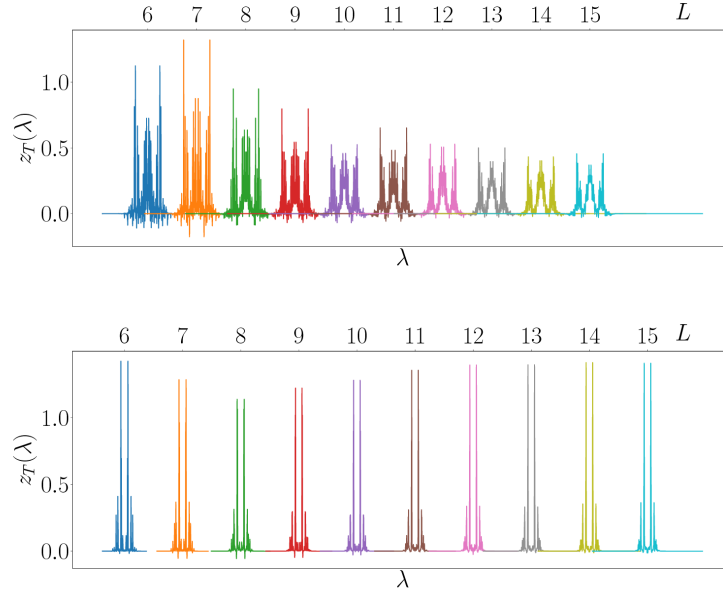
With this notation we have $z_T(\lambda) = [z_\infty * \mathcal{G}_{1/T}](\lambda)$, where $*$ denotes the convolution operator. The deviation from equilibrium $\Delta A(t)_\rho$ can then be bounded, using the convolution theorem of Fourier analysis, as

$$
\begin{aligned}
|\Delta A(t)_\rho| &\leq \left| \int z_T(\lambda) \mathrm{e}^{\mathrm{i}\lambda t} \mathrm{d}\lambda \right| + \left| \int \left(z_\infty(\lambda) - [z_\infty * \mathcal{G}_{1/T}](\lambda) \right) \mathrm{e}^{\mathrm{i}\lambda t} \mathrm{d}\lambda \right| \\
&\leq \left| \int z_T(\lambda) \mathrm{e}^{\mathrm{i}\lambda t} \mathrm{d}\lambda \right| + 2 \|A\| \left| 1 - \mathrm{e}^{-(t/T)^2} \right| \\
&\approx |\Delta A_T(t)_\rho| + 2 \|A\| (t/T)^2, \quad t \ll T,
\end{aligned}
\tag{7.20}
$$

with the regularized deviation from equilibrium

$$
\Delta A_T(t)_\rho := \int z_T(\lambda) \mathrm{e}^{\mathrm{i}\lambda t} \mathrm{d}\lambda.
\tag{7.21}
$$

The regularized quantity $\Delta A_T(t)_\rho$ always decays to zero on the time-scale of the cut-off time T. It is therefore essential that we only consider times much smaller than T once we use this regularized quantity.

Once we have introduced regularized quantities, we can meaningfully compare different system sizes. In particular we can formalize what it means that the distribution of z_Δ becomes independent of the system size N by saying that there exists a bounded function $\lambda \mapsto z(\lambda)$ such that

$$
\lim_{T \to \infty} \lim_{N \to \infty} \int |z_T(\lambda) - z(\lambda)| \mathrm{d}\lambda = 0.
\tag{7.22}
$$

We will later see (in section 7.4.1) that if such a function $z(\lambda)$ exists for a local observable and a local Hamiltonian, then equilibration (in a time independent of the system size) follows.

To see an example of such behaviour, see Figs. 7.1 and 7.3 (top). Fig 7.1 shows the time-evolution of a local observable in an equilibrating system and the corresponding dynamics of the z_{Δ_i}. Fig 7.3 (top) shows the behaviour of $z_T(\lambda)$ in the same system for different system sizes, indicating the emergence of a well-defined function $z(\lambda)$.

Conversely, we can also argue that equilibration is expected to fail if $\lambda \mapsto z(\lambda)$ is not bounded. We now argue that in this case there will be remaining oscillations with finite amplitude for all times. This can be seen in the following way. For any cut-off time T and

for any system size we have

$$\Delta A(0)_\rho = \sum_{0 \neq \Delta \in \text{Gaps}(H)} z_\Delta = \int z_T(\lambda) d_\lambda. \tag{7.23}$$

Hence, $z(\lambda)$ needs to have a finite integral, but is unbounded by assumption. The function will therefore have a finite contribution to the integral from some singularities. Such singularities concentrate a finite weight into an arbitrarily small region. As a consequence, they lead to a non-dispersing evolution of the z_Δ. In the simplest example, assume that these contributions originate from a finite number of δ-distributions. Since $z(\lambda) = z(-\lambda)$, they need to come in pairs of gaps $\{-\Delta_i, \Delta_i\}$. In the time-evolution this leads to a contribution of the form

$$\Delta A(t)_\rho \overset{t \to \infty}{\longrightarrow} \sum_i r_i \cos(\Delta_i t), \tag{7.24}$$

with r_i some real constants.

An example of a system that seems to show such a behaviour was found in Ref. [212] and its dynamics is illustrated in Figs. 7.2 and Fig. 7.3 (bottom). This example is given by a non-integrable, translational invariant, and local Hamiltonian with a local observable and a pure product-state as initial state. One would hence usually expect that the system equilibrates. In the numerics presented here and in Ref. [212] this does not seem to be the case. However, there is some debate about whether the oscillations indeed persist forever in the infinite system or whether they can be understood to originate from quasi-particles with very long, but finite lifetime [214, 215].

7.4.1   A simple argument from harmonic analysis

Let us now argue that equilibration of a local observable in a time independent of the system size is inevitable if the dynamics is generated by a local Hamiltonian and $z(\lambda)$ exists as a bounded function. To do this, we make use of recent rigorous results in many-body theory, which show that local observables can only connect energy-eigenstates corresponding to energy-eigenvalues that differ by a small amount. If a local Hamiltonian has eigenstates $|E_i\rangle$, then [216]

$$\left| \langle E_i | A | E_j \rangle \right| \leq \|A\| \, e^{-\alpha(|E_i - E_j| - 2R)}. \tag{7.25}$$

Here, $\alpha > 0$ and $R > 0$ are constants independent of the system size and R is proportional to the size of the region on which A acts.

In the case of a generic, interacting Hamiltonian, neither the gaps Δ nor the eigenvalues E_i are expected to be highly degenerate (potentially after restricting to a super-selection or symmetry sector). We then see from (7.14) and (7.25) that for large Δ, all z_Δ have to fall off exponentially, independent of the system size. Hence, also $z(\lambda)$ has to fall off exponentially for large $|\lambda|$. The function $z(\lambda)$ therefore has the following properties:

1. It is bounded (by assumption),

2. it has a finite integral,

3. and it falls off exponentially with $|\lambda|$.

These three properties together imply that the function is absolutely integrable, $\int |z(\lambda)| d\lambda < \infty$. The rest of the argument is then essentially given by the Riemann-Lebesgue Lemma [217]. An absolutely integrable function can always be approximated, to an arbitrary error $\delta > 0$, by a smooth function of compact support g_δ:

$$\int |g_\delta(\lambda) - z(\lambda)| d\lambda < \delta. \tag{7.26}$$

Then the Fourier transform of $z(\lambda)$, which gives $\Delta A(t)_\rho$ in the thermodynamic limit, can be approximated by the Fourier-transform $\hat{g}_\delta$ of g_δ:

$$|\Delta A(t)_\rho| \leq \left| \int g_\delta(\lambda) e^{i\lambda t} d\lambda \right| + \delta = |\hat{g}_\delta(t)| + \delta. \tag{7.27}$$
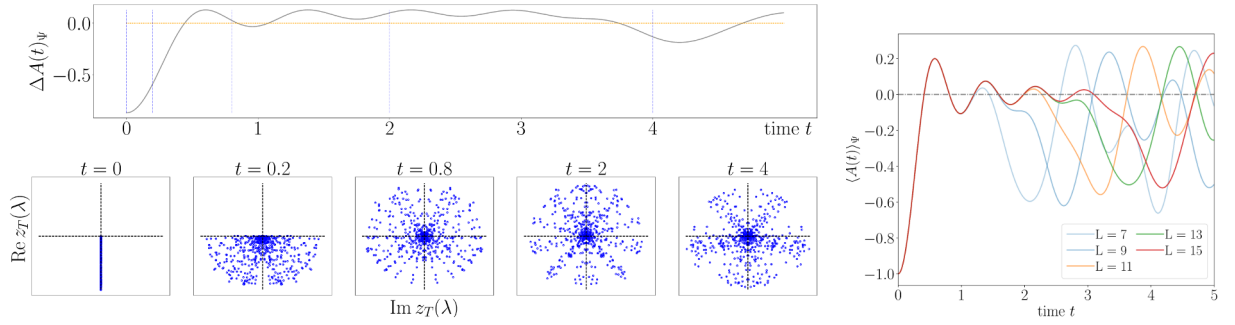
Figure 7.4: Illustration of finite-size effects: Equilibration in the XX chain with Hamiltonian $H_{XX} = \sum_{j=1}^{L} (\sigma_j^x \sigma_{j+1}^x + \sigma_j^y \sigma_j^y)$ on $L = 15$ sites with periodic boundary conditions. The top-left plot shows the evolution of the deviation of the instantaneous expectation value from the infinite time average of a σ^z operator acting on the first site. The initial state is a charge density wave state, i.e., $|\Psi\rangle = |1, 0, 1, 0, \ldots, 0, 1\rangle$, with $|0\rangle$ and $|1\rangle$ denoting the spin up and down state respectively. Up to time $t \approx 3$, the system seems to equilibrate to an expectation value such that $\Delta A(t)_\Psi \approx 0.1$. Thus, the physically relevant equilibrium value in the thermodynamic limit does not coincide with the time-average in the finite system. The right plot shows a finite-size scaling of the actual expectation value $\langle A(t)\rangle_\Psi$, indicating that finite-size effects indeed seem to become relevant at about $t \approx 3$. The bottom-left panel again shows the contribution to the Fourier transform of z_T by plotting the evolution of $z_T(\lambda_i)$ in the complex plane at the times marked in the evolution, where $T \approx 33$ and λ_i interpolate the between the larges and smallest gap in 5000 steps. Again gaps of the size $|\Delta| < 10^{-13}$ are considered to be zero and discarded in order to account for the subtraction of the steady-state value. (Figure adapted from Ref. [10].)

But the Fourier-transform of a compactly supported smooth function decays faster than any power. Therefore, for any fixed precision of equilibration δ, $t \mapsto \Delta A(t)_\rho$ equilibrates to this precision faster than any power. This means that for any $\delta > 0$ and $k \in \mathbb{N}$, there exists a constant $C_k(\delta)$ such that

$$|\Delta A(t)_\rho| \leq \min\left\{\|A\|, \frac{C_k(\delta)}{t^k}\right\} + \delta. \tag{7.28}$$

The constants $C_k(\delta)$ can, however, be very large and in fact diverge for $\delta \to 0$. This explains why this bound is not in conflict with equilibration in terms of a slow power-law like $t^{-1/2}$, as seen, for example, in integrable models (see, for example, [189, 196, 199, 200]).

7.4.2 A brief comment on finite-size effects

The above results concern the dynamics in the thermodynamic limit. In numerical investigations, these dynamics are in principle not accessible exactly, but one has to restrict oneself to finite systems. It is then important to keep in mind how the two settings are related. Suppose an observable equilibrates in a time t_{eq} (w.r.t. some suitable state) to the value A_{eq} in the thermodynamic limit. Now suppose that one runs a numerical situation on a large but finite system (with the corresponding reduced state as initial state). It then follows from Lieb-Robinson bounds that the observable should also evolve to the (approximately) the same value A_{eq} in time t_{eq}. However, in the finite system, this value A_{eq} does *not* necessarily coincide with the equilibrium value in the finite system, which is given by the time-average of $\langle A(t)\rangle_\rho$ on the finite system. The difference between these two values will become arbitrarily small when one considers increasingly large systems, but can be observably large in system sizes that are amenable to numerical simulations.

So, while it is true that if a finite system equilibrates, it equilibrates to the time-average (as discussed in section 7.1), this equilibrium value might not be the physically relevant one, if one is interested in what would happen in arbitrarily large systems. For an example of such behaviour, see Fig. 7.4.

7.5 Connecting the discussion to general results

In the previous sections, we have discussed the equilibration of local observables in complex quantum systems and argued that rapid, local equilibration is a generic feature. To do

this, we made several implicit and explicit assumptions. In this section, we discuss these assumptions in more detail and connect them with known rigorous results and conjectures in quantum many-body theory.

From the definition of z_Δ we find that essentially three properties are important for how a system equilibrates: The distribution of energy gaps and how local observables and physically relevant initial states look in the energy-eigenbasis. I now discuss these properties in more detail. After that I further discuss the role of the finite group velocity implied by Lieb-Robinson bounds for equilibration.

7.5.1   The energy spectrum

When discussing local many-body systems, it is important to keep in mind that the number of energy levels in a system of N degrees of freedom and local dimension d is given by d^N. However, the energy-range of a local Hamiltonian grows only linearly with the system size, $\|H\| \propto N$. This implies that typical energy-differences between consecutive energy levels becomes exponentially small in the system size. In the thermodynamic limit, this essentially leads to a continuous spectrum. The same is true for the gaps Δ: The number of gaps increases exponentially, while the largest gap only increases linearly with the system size, leading to an essentially continuous distribution for large system sizes.

For a systems with local interactions, the distribution of energy levels follow roughly a Gaussian distributions with a standard deviation that diverges as $\sqrt{N}$ with the system size. Intuitively, this can be seen in the following way. The distribution of energy-eigenstates of the local Hamiltonian $H = \sum_x h_x$ can be seen as the measurement statistics of the observable H in the maximally mixed state $\mathbf{1}/d^N$. Here, the Hamiltonian terms h_x only act on subsystems of some finite size and only overlap with finitely (uniformly bounded) many other terms. Hence the measurement statistics can be understood as that of a large number N of identical and almost independent random variables, yielding a Gaussian distribution with standard deviation of order $\sqrt{N}$.

This argument can be made precise using the following theorem [147]. A similar statement can be proven in one-dimensional systems more directly, see Ref. [218].

Theorem 7.3 (Berry-Esseen theorem [147]). *Let H be a k-local Hamiltonian in $\Lambda = [L]^D$ with $N = L^D$ particles and ρ a state with correlation length $\xi > 0$. Let*

$$\mu = \mathrm{Tr}(\rho H), \quad \sigma = \mathrm{Tr}(\rho(H-\mu)^2)^{1/2}, \quad s = \frac{\sigma}{\sqrt{N}k^{D/2}}. \tag{7.29}$$

Then

$$\sup_y |F(y) - G(y)| \leq \Gamma \frac{\log^{2D}(N)}{\sqrt{N}}, \tag{7.30}$$

where

$$F(y) := \sum_{k:E_k \leq y} \langle E_k | \rho | E_k \rangle, \tag{7.31}$$

and

$$G(y) := \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{y} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \, \mathrm{d}z \tag{7.32}$$

is the Gaussian cumulative distribution with mean μ and variance σ. The quantity Γ is given by

$$\Gamma = C_D \frac{(\max\{k, \xi\})^{2D}}{\sigma/\sqrt{N}} \max\left\{ \frac{1}{\max\{k, \xi\} \ln(N)}, \frac{1}{\sigma^2/N} \right\}, \tag{7.33}$$

where C_D only depends on the dimension of the lattice.

Using this theorem, the above intuition can be made precise in the following way. Since

$$\Gamma \to C_D \frac{k^{2D}}{\sigma/\sqrt{N}} \max\left\{\frac{1}{k\ln(N)}, \frac{1}{\sigma^2/N}\right\} \tag{7.34}$$

as $\xi \to 0$, it is sufficient to consider a sequence of states $\rho^{(m)}$ that converges to the maximally mixed state to obtain that the distribution of energy levels (the density of states) converges to a Gaussian (in distribution).

As a consequence of the distribution of energy levels being roughly Gaussian, the distribution of all energy gaps $\Delta = E_i - E_j$ also has to follow a roughly Gaussian distribution with standard deviation of order $\sqrt{N}$. In particular, we can expect that the distribution of gaps smaller in absolute value than some $\Delta_{\max}$ becomes more and more uniform as the system size increases.

7.5.2 The state

The second important ingredient in the distribution of z_Δ is the initial state ρ. As can be seen from the expression for z_Δ, only the off-diagonal elements of its density matrix in the energy eigenbasis play a role for the equilibration behaviour. To obtain a smooth distribution of z_Δ for large systems, it is necessary that there are sufficiently many and densely distributed off-diagonal elements in the density matrix that contribute to the z_Δ. In particular there should not be very few and large off-diagonal elements but many small ones.

The off-diagonal elements are not independent of the diagonal elements, due to positivity and normalization of the density matrix. In particular, we can bound the off-diagonal elements by their corresponding diagonal elements:

$$|\langle E_i|\rho|E_j\rangle|^2 \leq \langle E_i|\rho|E_i\rangle \langle E_j|\rho|E_j\rangle. \tag{7.35}$$

If the state is sufficiently non-diagonal in the energy-eigenbasis, we can therefore expect that there are many non-zero entries $\langle E_i|\rho|E_i\rangle$, namely a number that diverges as the system size increases. Since they have to add up to unity, most of them will be very small. In turn, the off-diagonal elements will also be small.

In the case of product states and more general states with a finite correlation length, theorem 7.3 shows that the energy-distribution $\langle E_i|\rho|E_i\rangle$ falls off (sub-)exponentially for energies that differ macroscopically from the mean value, i.e., for energies that differ by more than a sub-linear function in the system size from the mean. More precisely, the distribution of the energy-density is highly peaked around its mean in the sense that the *total probability* of energies whose density differs from the mean energy-density decays subexponentially with the system-size. This is made more explicit by the following theorem.

Theorem 7.4 (Finite correlation length implies peaked energy distribution [219])**.** *Let ρ be a quantum state with correlation length $\xi > 0$ in a D-dimensional system and $\langle H \rangle_\rho = \mathrm{tr}(H\rho)$ be the average energy of ρ. For any $a \geq (2^{O(D)}/n\xi)^{1/2}$, it holds that*

$$\mathrm{tr}(\rho\Pi_{[\langle H\rangle_\rho + na, \infty)}) \leq O(\xi)\exp\left(-\frac{(na^2\xi)^{1/(D+1)}}{O(1)D\xi}\right) \tag{7.36}$$

and

$$\mathrm{tr}(\rho\Pi_{[0, \langle H\rangle_\rho - na]}) \leq O(\xi)\exp\left(-\frac{(na^2\xi)^{1/(D+1)}}{O(1)D\xi}\right). \tag{7.37}$$

While at first this theorem and the above Berry-Esseen theorem might be interpreted as saying that only few energy levels contribute to the state ρ, this is not true: remember that the total number of energy levels in any interval of energies of non-decreasing width increases exponentially with the system size. Thus the total number of energy levels that contribute to a state with a finite correlation length can be expected to diverge exponentially with the system size.

As explained in section 7.1, a particularly useful measure for how many energy levels participate in a given state in the context of equilibration is given by the effective dimension: A high effective dimension ensures that a system equilibrates to high precision in infinite time, which is clearly a pre-requisite for equilibration in finite time.

Recently, it was indeed proven that states with a finite correlation length have an effective dimension that diverges with the system size. The results in Ref. [190] show that in such a case

$$\frac{1}{d_{\text{eff}}(\rho)} \leq C \frac{\ln^{2d}(N)}{s^3 \sqrt{N}}, \qquad (7.38)$$

where $C > 0$ is a constant independent of the system size and $s = \sigma/\sqrt{N}$. Here, σ is the standard-deviation of the energy of ρ. The quantity s is always upper bounded independent of the system size, but since σ can be expected to be of order $\sqrt{N}$, it is usually also lower bounded by a constant (as long as we are not dealing with an energy eigenstate). We thus see that Eq. (7.38) shows that an ever-increasing number of energy levels participate in the initial state. We can conclude that the probability distribution of energies is spread out over many energy levels, each of which only carrying a small amount of probability.

Together with (7.35), these arguments make it plausible that in natural initial states the off-diagonal elements of the density matrix are sufficiently small. This implies that if the initial deviation from equilibrium $\Delta A(t)_\rho$ is large, then there have to be many off-diagonal elements with small absolute value each. On the other hand, if $\Delta A(t)_\rho$ is small, then the system can be expected to already be equilibrated. To add further plausibility to rapid equilibration, let us now discuss the role of the observable.

7.5.3   The observable

Let us finally discuss the shape of a local observable in the energy-eigenbasis. We have seen before already that a local observable only has significant off-diagonal entries for gaps of the order of the size of the support of the observable, which is independent of the system size. Therefore only gaps of the size of the order of the support of the observable are relevant for the distribution of z_Δ. We also know from the previous section that the distribution of energy gaps has a standard deviation that grows sublinearly with the system size. Combining this fact with the assumption that the distribution of gaps becomes essentially smooth in large systems we conclude that the distribution of gaps should become essentially *uniform* on the relevant scale, the size of the support of the observable. This justifies the assumption of a uniform distribution of gaps within some interval $[-\Delta_{\max}, \Delta_{\max}]$ in the toy-model discussed in section 7.3.

While there is little known rigorously about the shape of local observables in the energy eigenbasis, in the study of *thermalization* of closed quantum systems a strong conjecture (backed up by numerical evidence) about the shape of local observables in the energy eigenbasis has taken a prominent role over the last years: this conjecture is called the *eigenstate thermalization hypothesis* (ETH) [41, 220–222]. The ETH states that when we consider expectation values of local observables, individual eigenstates of H already give predictions very similar to those of Gibbs states, due to the fact that eigenstates of interacting many-body systems are highly entangled (at least in the bulk of the spectrum).

One requirement for such a behaviour is that the diagonal matrix-elements $\langle E | A | E \rangle$ depend smoothly on the energy E since thermal expectation values depend smoothly on the temperature. In addition, it is often argued that also the off-diagonal elements of A obey a smooth distribution up to small fluctuations. There are several technical formulations of the ETH. A particularly strong formulation is the following one from the review in Ref. [41].

Assumption 7.5 (Eigenstate equilibration hypothesis [41])**.**

$$\langle E_i | A | E_j \rangle = g(\bar{E})\delta_{i,j} + e^{-S(\bar{E})/2} f_A(\bar{E}, \omega) R_{i,j} \qquad (7.39)$$

where $\bar{E} := (E_i + E_j)/2$, $\omega = E_i - E_j$, *and* $E \mapsto S(E)$ *is the micro-canonical entropy associated to energy E. g and* f_A *are both assumed to be smooth functions of their argu-*

ments, $g(\bar{E})$ being the expectation value of the microcanonical ensemble at energy $\bar{E}$, and $R_{i,j}$ being a random (real or complex) variable with zero mean and unit variance.

Of course, it is only expected that the ETH is fulfilled by generic, sufficiently interacting systems, the random variables $R_{i,j}$ being fixed by the particular choice of local observable and Hamiltonian. If we accept the ETH, we can therefore expect that the local observable, and hence the z_Δ, approach a smooth distribution for large systems and initial states with energy densities in the bulk of the spectrum. This is true despite of the fluctuations $R_{i,j}$: In the limit of large systems they should average to smooth distributions after we regularize with any Gaussian of finite width (as is always required, compare with section 7.4).

Summarizing the above discussions, we can conclude that we should generically expect that in a situation where we consider an initial state with a finite correlation length and a local observable in a generic interacting many-body system, one of the following two alternatives is true: i) the system is already at equilibrium or ii) the system equilibrates in a time independent of the system size.

7.5.4   *Implications of the finite group velocity*

Let us now discuss the implication of the finite group velocity in many-body systems in the context of equilibration. As mentioned several times in this chapter already, in every local quantum many-body system with finite-dimensional local Hilbert-spaces, there is a finite velocity v_{LR} with which information and excitations can propagate through the system. This is similar to a light-cone in relativistic physics. However, in the case of many-body systems, this "light-cone" is not strict, but processes which violate the "light-cone" condition are exponentially suppressed. This result is known as *Lieb-Robinson bounds* (LR-bounds) and the corresponding "light-cone" is usually called the *Lieb-Robinson cone* (LR-cone). Mathematically, one way to express this result is [223]

$$\| [A(t), B] \| \le c \, \|A\| \, \|B\| \, \mathrm{e}^{-a(\mathrm{d}(A,B) - v_{\mathrm{LR}}|t|)}, \tag{7.40}$$

where A, B are local observables, $a, c > 0$ are constants and $\mathrm{d}(A, B)$ is the distance (measured by the lattice-distance) between the observables A and B. Due to the exponential surpression of the tails, one can neglect the tails of the LR-cone in exchange for an arbitrarily small error in many applications. The group velocity v_{LR} is essentially determined by the interaction strength. Since the bound is expressed as a bound on the norm of the commutator of observables in the Heisenberg-picture, it is independent of the initial state of the system.

Since information can only travel with a finite velocity, a local observable A only sees a small part of the full system in a given time t. If it is necessary to sense a length-scale l for the local system to equilibrate, the finite group velocity therefore puts a *lower bound* on the equilibration scale as $t_{\mathrm{eq}} \ge l/v_{\mathrm{LR}}$.

In many cases we therefore do not need simulate arbitrarily large system sizes when we do numerical checks of equilibration times. The LR-bounds imply that if we see rapid equilibration for small system sizes, we know that it will also happen for larger system sizes. Similarly, if we are interested in effects that should be independent of the boundary conditions then if we can simulate a system up to linear system size L, we should only consider the time-evolution up to a time of about $t = L/v_{\mathrm{LR}}$.

If we would let the simulation run for a time longer than L/v_{LR}, information could travel across the system and come back to the local subsystem we were interested in. This leads to a similar effect as a recurrence time. Thus, only simulations up to times of the order of L/v_{LR} should be considered. If we do not see equilibration in such a simulation, we have to increase the system size. Lieb-Robinson bounds hence allow us to confirm but not to falsify equilibration on small systems.

In this chapter, we were only interested in systems that are translationally invariant to some degree. As emphasized before, however, much of the arguments given should also hold for systems that deviate from this assumptions. As mentioned before, disordered, but strongly interacting systems, so-called *many-body localized* systems, indeed also equilibrate. These systems fulfill stronger bounds than the LR-bounds, in which the light-cone is deformed away from a linear cone into a logarithmic cone. There is still a spread of

information (but not of quasi-particles) throughout the system and dephasing in the sense as we have explained it can happen [224–226]. Contrarily, in non-interacting disordered systems, like Anderson insulators [227], not even information can propagate across length-scales larger than the localization length [228, 229]. As a consequence, local excitations effectively live in small, finite systems and no local equilibration is observed.

7.5.5 Typicality results

In this chapter, I have emphasized a "mechanistic" notion of equilibration, avoiding any notion of probabilities and putting in focus assumptions which can be expected to be fulfilled due to the locality of the Hamiltonians and observables and finite correlation lengths in initial states. This is in contrast to a notion of equilibration in terms of "typicality" [188, 230]. In this approach, inspired from random matrix theory, the focus is on the unitary transformation U that transforms the eigenbasis of the observable A into that of H. As in random matrix theory, it is argued that one should expect that the overwhelming majority of randomly sampled U (according to the Haar measure) should typically also apply to the particular, non-random actual system of interest. Assuming this, one can indeed derive rigorous bounds on equilibration times, yielding suprisingly accurate predictions for some systems. On the other hand, it is not easy to argue why any given fixed system should indeed be typical. The discussion in this chapter should thus be seen as complementary to that in Ref. [230]. Both discussions support the view that rapid local equilibration is indeed a generic feature of complex quantum systems.

7.6 Conclusions and Outlook

How statistical mechanics follows from unitary dynamics in quantum mechanics is one of the big questions in the sudy of complex quantum systems. For a long time, statistical mechanics was motivated through the notion of ensembles, as introduced by Gibbs [26]. Today, we know that thermal equilibrium can be explained from the fact that interacting quantum systems entangle their microscopic parts over time and that this entanglement naturally leads to a notion of entropy and statistical ensembles on parts of the whole system – this is the point of view advocated in the field of pure state statistical mechanics [231]. A necessity for this is that complex quantum systems indeed equilibrate. In this chapter, I have reviewed general results that show that generic, complex quantum systems are expected to equilibrate in a well defined sense. An important figure of merit that measures how well a given quantum state equilibrates relative to a given Hamiltonian is the *effective dimension*. While progress has been made on showing that natural states have a large effective dimension, it remains to be an important open problem to find strong lower bounds on the effective dimension from well-motivated physical assumptions.

Going beyond equilibration on average, in this chapter I have put forward arguments that help to explain why complex quantum systems do not only equilibrate, but that they also equilibrate in a time that is not astronomically large. These arguments are supposed to provide intuition on how this happens and what kind of figures of merit are relevant for the problem; clearly they are not rigorous arguments proving equilibration in the mathematical sense.

The real work for the future lies in connecting the figures of merit laied out in this chapter with concrete physical properties. It is an intriguing open problem to formulate precise, physical conditions that allow us to prove equilibration in finite time from first principles.

8 The cost of avoiding equilibration

In the previous chapter, we have discussed in great detail the mechanism of equilibration and have argued that for many "natural" initial states we can expect that a sufficiently complex many-body system equilibrates in a reasonable time. We can therefore expect that it is not easy to prepare a large interacting system in a state that does *not* equilibrate. Nevertheless, we have not given any quantitative argument that really proves that such a preparation is difficult in some well-defined measure. In this chapter[1], I will provide a quantitative way to establish that it is indeed extremely difficult to prepare a large system in a state that does not equilibrate. To achieve this, we will combine reasoning of resource theories (see chapter 2) with the general results about equilibration from the previous chapter.

The problem that we will consider is the following: Suppose Alice has a large system Q that is initially in equilibrium, i.e., its initial state σ_Q commutes with the Hamiltonian and is therefore stationary. Now she wants to bring the system to a state that does not equilibrate. This requires to bring it to a state that has small effective dimension, due to theorems 7.1 and 7.2. To achieve this, Bob gives to Alice a second system R in state σ_R, which is also stationary. Alice then acts with a unitary operation on the joint-system QR with the aim to produce a state on Q with small effective dimension. For example, she introduces a time-dependent interaction between the system, or implements quenches on the systems. In fact, we can be more generous and assume that Alice can implement any quantum channel Λ that is unital, i.e., leaves the maximally mixed state invariant. She thus has arbitrary high control over the microscopic degrees of freedom of QR and is not even subjected to energy-conservation or entropy-conservation. In particular Alice can first bring the system R into a non-stationary state using a unitary transformation and if the initial effective dimension is small, this non-stationary state can be chosen so that the system does not equilibrate. Thus, the assumption that R is initially at stationary is not as strong as it may at first seem. Nevertheless, we will relax this condition later on.

In the following sections, I will show that Alice can only bring the system Q to a state with small effective dimension if R itself had low effective dimension from the beginning. Thus, to bring a state of a large system into a state that has the potential to withstand equilibration, it is necessary to have another large system that already has this property.

As we will see, the essential property that is responsible for this mechanism is that given a quantum state with a certain effective dimension, there are in general many more state with higher effective dimension than with lower effective dimension. In particular, if the state is stationary, *all states* that can be reached by a unital channel have a higher effective dimension. For example, if a local many-body system is initially thermalized and its global state is a Gibbs state, then its effective dimension is very large. It is then impossible to bring the system to a state that does not equilibrate to high precision by acting with any unitary transformation on this system.

The main result in this chapter shows that a (large) system R that has the potential to withstand equilibration is necessary to bring another (large) system out of equilibrium. We

will also see that this mechanism is, in principle, catalytic: Provided that we assume that R and Q are initially uncorrelated but can become correlated in the process of bringing Q out of equilibrium, we will see that R can be re-used to bring further systems out of equilibrium. In this sense R is not being "used up". This is in close analogy to the discussions in chapters 4 and 5.

8.1   Formalization of the problem and main result

To formalize our approach, let us define the *resilience (to equilibration)* $\mathcal{R}$ of a system initially in state ρ and evolving under the Hamiltonian H as

$$\mathcal{R}(\rho, H) := \log \left(\frac{d}{d_{\text{eff}}(\rho, H)} \right), \tag{8.1}$$

where d is the Hilbert-space dimension of the system. Recall that the effective dimension is defined as

$$d_{\text{eff}}(\rho, H) = \frac{1}{\text{Tr}(\omega_H(\rho)^2)}, \tag{8.2}$$

where $\omega_H(\rho) = \sum_k P_k \rho P_k$ and P_k are the energy-projectors of H. Here, we have implicitly defined ω_H as the dephasing-channel that maps an initial state to its time-average.

The resilience $\mathcal{R}$ measures the potential to withstand equilibration. For example, if ρ is the maximally mixed state, which is always perfectly equilibrated, we have $\mathcal{R}(\mathbf{1}/d, H) = 0$. On the other hand, ρ is a super-position of very few energy-eigenstates, which in general does not equilibrate, then $\mathcal{R}(\rho) \approx \log(d) = n$, where n is the system size.

Thus, a large resilience is a necessary condition to withstand equilibration, but it is not a sufficient condition to withstand equilibration. For example, energy-eigenstates have a very large resilience, but are perfectly stationary. This will, however, not be a problem in the following, because we want to show that all states that do not equilibrate are difficult to prepare. It is then not problematic that it is also difficult to prepare certain states that do equilibrate.

As stated in the introduction, we consider the situation where we have initially a system Q with Hamiltonian H_Q and a system R with Hamiltonian H_R. We will assume that the systems Q and R are fully interacting systems and the Hamiltonians H_Q and H_R have non-degenerate energy-gaps (see chapter 7). Let us denote the total Hamiltonian by $H_{QR} = H_Q \otimes \mathbf{1} + \mathbf{1} \otimes H_R$. The initial state of RQ is given by

$$\sigma_{QR} = \sigma_Q \otimes \sigma_R. \tag{8.3}$$

For now, let us assume that the state σ_{QR} is stationary, i.e., we have

$$\sigma_Q(t) = \mathcal{T}_t(\sigma_Q) := e^{-iH_{QR}t} \sigma_{QR} e^{iH_{QR}t} = \sigma_{QR}. \tag{8.4}$$

Then, we allow to change the Hamiltonian on Q and R arbitrarily over time for some period of time T. The only condition that we impose is that the final Hamiltonian $\tilde{H}_{QR}$ at time T again does not couple Q and R and that the final local Hamiltonians $\tilde{H}_Q$ and $\tilde{H}_R$ are fully interacting Hamiltonians with non-degenerate energy gaps again. Such a time-dependent trajectory of Hamiltonians results in some overall unitary transformation U on QR.

Going beyond that, we can imagine that we could randomly implement different trajectories of Hamiltonians, which all end with the final Hamiltonian $\tilde{H}_{QR}$. If the i-th trajectory occurs with probability i and leads to the unitary U_i, then the resulting quantum channel on QR is simply given by

$$\Lambda(\cdot) = \sum_i p_i U_i \cdot U_i^\dagger. \tag{8.5}$$

Note that any such transformation is unital, i.e., it leaves the maximally mixed state $\mathbf{1}/d_{QR}$ invariant.

To summarize, the process of bringing the system out of equilibrium can be described abstractly as

$$(\sigma_{QR}, H_{QR}) \mapsto (\Lambda(\sigma_{QR}), \tilde{H}_{QR}) =: (\rho_{QR}, \tilde{H}_{QR}), \qquad (8.6)$$

with $\Lambda(\mathbf{1}) = \mathbf{1}$. The systems Q and R continue to evolve under their respective Hamiltonians $\tilde{H}_Q$ and $\tilde{H}_R$ after this process, leading to trajectories of states $\rho_Q(t)$ and $\rho_R(t)$. Then the question is whether the trajectory $\rho_Q(t)$ can be made to be non-equilibrating by a suitable transformation Λ and Hamiltonian $\tilde{H}_{QR}$. The first result is the following theorem.

Theorem 8.1 (Large change of resilience requires large resilience). *Let two systems Q and R with Hamiltonians H_Q and H_R in a stationary state $\sigma_Q \otimes \sigma_R$ be given. Then every trajectory $\rho_Q(t)$ generated according to the above procedure fulfills*

$$\Delta \mathcal{R}_Q := \mathcal{R}(\rho_Q, \tilde{H}_Q) - \mathcal{R}(\sigma_Q, H_Q) \leq \mathcal{R}(\sigma_R, H_R) \qquad (8.7)$$

for every final Hamiltonian $\tilde{H}_Q \otimes \mathbf{1} + \mathbf{1} \otimes \tilde{H}_R$.

The theorem expresses that to increase the resilience on Q by some amount, this amount of resilience needs to have been present in the system R. Hence, the system R should be viewed as a *resource* to prepare states that do not equilibrate. To get a better understanding of this result, let us consider an example. Suppose the initial state on Q is a microcanonical state, which has equal overlap with $K(n)$ energy-eigenstates. For a usual micro-canonical state we expect that $K(n) \approx e^{\alpha n}$ for some constant $\alpha > 0$. This leads to a finite entropy-density and for thermal states the constant α would control the temperature of the initial state. Suppose that the system R is a m-partite system with local Hilbert-space dimension d_R. Then the theorem tells us that

$$d_{\text{eff}}(\rho_Q, \tilde{H}_Q) \geq \frac{K(n)}{d_R^m} \approx \frac{e^{\alpha n}}{d_R^m}. \qquad (8.8)$$

Therefore, as the system size of Q increases, the size of the system R also has to increase at least linearly with the size of Q to have any hope of bringing the system Q to a state that does not equilibrate very well.

We can also use theorem 8.1 to show that local quenches cannot bring a system far out of equilibrium. Imagine a system Q that is initially in a stationary state σ_Q of the Hamiltonian H_Q. Then quench the system locally by applying an *arbitrary* quantum channel Φ to a subsytem $X \subset Q$ with Hilbert-space dimension d_X. Then theorem 8.1 implies that the change of resilience is bounded as

$$\Delta \mathcal{R}_Q = \mathcal{R}(\Phi(\sigma_Q), \tilde{H}_Q) - \mathcal{R}(\sigma_Q, H_Q) \leq \log(d_X). \qquad (8.9)$$

Thus, a local quench, for example a local spin-flip, cannot significantly change the resilience on stationary initial states. Importantly, there is no restriction on the channel Φ apart from acting only locally on X. This result follows from the fact that every quantum channel can be dilated to a unitary [62]: If Φ acts on a system of dimension d_X, there exists a state σ_R of dimension at most d_X and a unitary U_Φ such that

$$\Phi(\sigma_Q) = \text{Tr}_R(U_\Phi \sigma_Q \otimes \sigma_R U_\Phi^\dagger). \qquad (8.10)$$

Since we can choose the Hamiltonian on R freely, we can assume that σ_R is stationary and the result follows directly from theorem 8.1.

Proof. Let us now prove the theorem. The proof follows similar ideas as the proofs in the resource theory of thermodynamics (see chapters 2–5) and is surprisingly simple. To prove the theorem, we first express the resilience in terms of Rényi-divergences (see chapter 2):

$$\mathcal{R}(\sigma_{QR}, H_{QR}) = D_2\left(\omega_{H_{QR}}(\sigma_{QR}) \| \mathbf{1}/d_{QR}\right). \qquad (8.11)$$

Now note that the map $\omega_{\tilde{H}_{QR}}$ is unital. We can now use the data-processing inequality for the unital map $\omega_{\tilde{H}_{QR}} \circ \Lambda$:

$$\mathcal{R}(\sigma_{QR}, H_{QR}) = D_2(\sigma_{QR} \| \mathbf{1}/d_{QR}) \tag{8.12}$$

$$\geq D_2(\omega_{\tilde{H}_{QR}} \circ \Lambda(\sigma_{QR}) \| \mathbf{1}/d_{QR}) \tag{8.13}$$

$$= \mathcal{R}(\Lambda(\sigma_{QR}), \tilde{H}_{QR}) = \mathcal{R}(\rho_{QR}, \tilde{H}_{QR}). \tag{8.14}$$

Furthermore, Tr_R is also a unital map and we have $\mathrm{Tr}_R \circ \omega_{\tilde{H}_{QR}} = \omega_{\tilde{H}_Q} \circ \mathrm{Tr}_R$. We thus obtain

$$\mathcal{R}(\rho_{QR}, \tilde{H}_{QR}) \geq \mathcal{R}(\rho_Q, \tilde{H}_Q). \tag{8.15}$$

We now make use of the fact that the resilience is additive if the initial state on QR is stationary:

$$\mathcal{R}(\sigma_{QR}, H_{QR}) = \mathcal{R}(\sigma_Q, H_Q) + \mathcal{R}(\sigma_R, H_R) \tag{8.16}$$

This follows directly from the additivity of Rényi-divergences under tensor products. Putting together Eqs. (8.12),(8.15) and (8.16) we obtain the desired result. □

8.2   Non-stationary initial states

In the previous section, we assumed that the initial states on Q and R are stationary. This is a natural assumption, since it asks whether we can create non-equilibrium states from perfectly equilibrated states, and we see that this can only be possible if at least one of the systems R or Q already has small effective dimension. Let us nevertheless now also discuss a way of relaxing this assumption by replacing it with an assumption on the possible operations. To motivate this assumption, consider for now a single system, say Q, which is initially in a pure state of the form

$$|\psi\rangle_Q := \frac{1}{\sqrt{d_Q}} \sum_i |E_i\rangle, \tag{8.17}$$

where $|E_i\rangle$ denote the energy eigenstates of the Hamiltonian H_Q on Q. Provided that the Hamiltonian H_Q has non-degenerate spectrum, this state has zero resilience to equilibrate, since its time-average is simply the maximally mixed state.

Nevertheless, it is clear that we can bring it to any pure state with arbitrary large resilience by a single unitary operation. Hence it can be brought to a state that does not equilibrate, for example an equal superposition of two energy eigenstates. In this case it is therefore, in principle, perfectly possible to bring the system to a state that does not equilibrate, even without using any resource.

Since the state $|\psi\rangle_Q$ fluctuates rapidly, however, the final state after the operation has been applied will in general depend very sensitively to the time t_{init} after which the operation is applied:

$$\Lambda \left(\mathcal{T}_{t_{\text{init}}}^Q(|\psi\rangle\langle\psi|_Q) \right) \neq \Lambda \left(|\psi\rangle\langle\psi|_Q \right) = \rho_Q. \tag{8.18}$$

Thus, in an experiment, one would have to have very precise control over the timing to consistently prepare the same trajectory $\rho_Q(t)$ to gather statistics and verify that the operation indeed produces the non-equilibrating trajectory $\rho_Q(t)$. Due to the time-scale of the fluctuations, such a degree of control is in many situations unrealistic. We can incorporate this formally into our set-up by assuming that for every time t there exists a (potentially different) time t' such that

$$\Lambda \circ \mathcal{T}_t(\sigma_{QR}) = \tilde{\mathcal{T}}_{t'} \circ \Lambda(\sigma_{QR}), \tag{8.19}$$

where $\tilde{\mathcal{T}}_t$ denotes the time-evolution under the post-quench Hamiltonian $\tilde{H}_{QR}$. In particular, this contains the special case that σ_{QR} is already stationary, but also different well-known classes of maps. Clearly, in the case $H_{QR} = \tilde{H}_{QR}$ the condition holds for *covariant*

maps, which fulfill $\Lambda \circ \mathcal{T}_t = \mathcal{T}_t \circ \Lambda$ [107], which do not require any timing precision. It also holds for *phase-invariant* maps, which fulfill $\Lambda \circ \mathcal{T}_t = \Lambda$. In general, however, condition (8.19) is much weaker than any of the two, since it only has to hold for the particular initial state σ_{QR} and t' can be *any* function of t. However, we leave a more detailed investigation of this condition for future work.

The second assumption that we will make is that the dephasing-map with respect to H_{QR} does not create any correlations between Q and R:

$$\omega_{H_{QR}} = \omega_{H_Q} \otimes \omega_{H_R}. \tag{8.20}$$

This condition is fulfilled as long as there are no distinct energies $E_Q \neq E_Q'$ and $E_R \neq E_R'$ such that $E_R + E_Q = E_R' + E_Q'$. In particular it is fulfilled if H_{QR} is non-degenerate. Note that for the equilibration bounds, we anyway assume that H_Q and H_R are non-degenerate. It is then generically true that also H_{QR} is non-degenerate. However, the condition (8.20) can also be true if H_Q and H_R have degeneracies, as long as no new degeneracies are created by adding them. It is, however, not fulfilled if $H_Q = H_R$.

Under these assumptions, we can now proof the following generalization of theorem 8.1.

Theorem 8.2. *Let two systems Q and R with Hamiltonians H_Q and H_R in the state $\sigma_Q \otimes \sigma_R$ be given. Assume that $\omega_{H_{QR}} = \omega_{H_Q} \otimes \omega_{H_R}$. Then for every state ρ_Q generated using a unital map Λ that fulfills* (8.19) *with respect to $\tilde{H}_{QR} = \tilde{H}_Q \otimes \mathbf{1} + \mathbf{1} \otimes \tilde{H}_R$ we have*

$$\Delta\mathcal{R}_Q \coloneqq \mathcal{R}(\rho_Q, \tilde{H}_Q) - \mathcal{R}(\sigma_Q, H_Q) \leq \mathcal{R}(\sigma_R, H_R). \tag{8.21}$$

Proof. The proof of the theorem is the same as that of Theorem 8.1 once we have shown monotonicity and additivity of the resilience. Additivity follows straightforwardly from the assumption $\omega_{H_{QR}} = \omega_{H_Q} \otimes \omega_{H_R}$. For monotonicity we use assumption (8.19). First we compute

$$\omega_{\tilde{H}_{QR}} \circ \Lambda \circ \omega_{H_{QR}}(\sigma_{QR}) = \lim_{T \to \infty} \frac{1}{T} \int_0^T \omega_{\tilde{H}_{QR}} \circ \Lambda \circ \mathcal{T}_t(\sigma_{QR})$$

$$= \lim_{T \to \infty} \frac{1}{T} \int_0^T \omega_{\tilde{H}_{QR}} \circ \tilde{\mathcal{T}}_{t'} \circ \Lambda(\sigma_{QR})$$

$$= \lim_{T \to \infty} \frac{1}{T} \int_0^T \omega_{\tilde{H}_{QR}} \circ \Lambda(\sigma_{QR})$$

$$= \omega_{\tilde{H}_{QR}} \circ \Lambda(\sigma_{QR}), \tag{8.22}$$

where we have used $\omega_{\tilde{H}_{QR}} \circ \tilde{\mathcal{T}}_{t'} = \omega_{\tilde{H}_{QR}}$ for all t'. We then have

$$\mathcal{R}(\Lambda(\sigma_{QR}), \tilde{H}_{QR}) = D_2\left(\omega_{\tilde{H}_{QR}} \circ \Lambda(\sigma_{QR}) \| \mathbf{1}/d_{QR}\right)$$

$$= D_2\left(\omega_{\tilde{H}_{QR}} \circ \Lambda(\omega_{H_{QR}}(\sigma_{QR})) \| \mathbf{1}/d_{QR}\right)$$

$$\leq D_2\left(\omega_{H_{QR}}(\sigma_{QR}) \| \mathbf{1}/d_{QR}\right) = \mathcal{R}(\sigma_{QR}, H_{QR}), \tag{8.23}$$

which shows monotonicity. $\qquad\square$

8.3   *Correlations and the second law of equilibration*

The result in the previous section shows that to bring a system into a non-equilibrating state, it is necessary to have at our disposal a system which can withstand equilibration to a sufficient degree. It is natural to ask, whether in the process of bringing the system Q out of equilibrium, the system R is automatically brought *into* equilibrium, i.e., whether the resilience on R is spent in the process. We can thus ask whether a relation of the form

$$\Delta\mathcal{R}_Q \leq -\Delta\mathcal{R}_R \tag{8.24}$$

holds true. In this case, we would obtain a "second law of equilibration", since the resilience to equilibrate could never be created but only re-shuffled. In general, this re-shuffling would

diminish the resilience, so that after sufficiently long time, all systems would have small resilience and equilibrate. Let us first discuss a case in which this relation holds. Suppose that after the application of the map Λ the systems Q and R are uncorrelated, i.e.,

$$\rho_{QR} = \rho_Q \otimes \rho_R \tag{8.25}$$

and that $\omega_{\tilde{H}_{QR}} = \omega_{\tilde{H}_Q} \otimes \omega_{\tilde{H}_R}$. We can now use the following property of the resilience.

Lemma 8.3 (Super-additivity on product-states)**.** *For any bipartite, non-interacting system we have*

$$\mathcal{R}(\rho_Q \otimes \rho_R, \tilde{H}_{QR}) \geq \mathcal{R}(\rho_Q, \tilde{H}_Q) + \mathcal{R}(\rho_R, \tilde{H}_R). \tag{8.26}$$

Proof. A direct calculation using the formulation of $\omega_{\tilde{H}_{QR}}$ as time-averages yields

$$\omega_{\tilde{H}_{QR}} \circ \omega_{\tilde{H}_Q} \otimes \omega_{\tilde{H}_R} = \omega_{\tilde{H}_Q} \otimes \omega_{\tilde{H}_R} \circ \omega_{H_{QR}}. \tag{8.27}$$

But clearly, $\omega_{\tilde{H}_{QR}} \circ \omega_{\tilde{H}_Q} \otimes \omega_{\tilde{H}_R} = \omega_{\tilde{H}_Q} \otimes \omega_{\tilde{H}_R}$. Using the data-processing inequality and the fact that maximally mixed states are product-states, we then get

$$
\begin{aligned}
\mathcal{R}(\rho_Q \otimes \rho_R, \tilde{H}_{QR}) &= D_2\left(\omega_{\tilde{H}_{QR}}(\rho_Q \otimes \rho_R)\|\mathbb{1}/d_{QR}\right) \\
&\geq D_2\left(\omega_{\tilde{H}_Q} \otimes \omega_{\tilde{H}_R} \circ \omega_{\tilde{H}_{QR}}(\rho_Q \otimes \rho_R)\|\mathbb{1}/d_{QR}\right) \\
&= D_2\left(\omega_{\tilde{H}_{QR}} \circ \omega_{\tilde{H}_Q} \otimes \omega_{\tilde{H}_R}(\rho_Q \otimes \rho_R)\|\mathbb{1}/d_{QR}\right) \\
&= D_2\left(\omega_{\tilde{H}_Q} \otimes \omega_{\tilde{H}_R}(\rho_Q \otimes \rho_R)\|\mathbb{1}/d_{QR}\right) \\
&= \mathcal{R}(\rho_Q, \tilde{H}_Q) + \mathcal{R}(\rho_R, \tilde{H}_R).
\end{aligned}
$$

$\square$

Using this Lemma we now get

$$
\begin{aligned}
\mathcal{R}(\sigma_Q, H_Q) + \mathcal{R}(\sigma_R, H_R) &= \mathcal{R}(\sigma_{QR}, H_{QR}) \\
&\geq \mathcal{R}(\rho_Q \otimes \rho_R, \tilde{H}_{QR}) \\
&\geq \mathcal{R}(\rho_Q, \tilde{H}_Q) + \mathcal{R}(\rho_R, \tilde{H}_R). \tag{8.28}
\end{aligned}
$$

Hence, (8.24) is fulfilled. However, the assumption that ρ_{QR} is uncorrelated is in general unjustified. Indeed, we will now see that (8.24) can be violated in an extreme way even if the correlations are arbitrarily small in terms of the mutual information.

Theorem 8.4 (No second law of equilibration)**.** *Consider a family of n-partite many-body systems with increasing n. For large enough n, there are stationary states $\sigma(n)_Q$ and initial and final non-interacting Hamiltonians $H_{QR} = \tilde{H}_{QR}$ such that for any $\epsilon > 0$ there exists a corresponding stationary resource state $\sigma_R^{(n)}$ and a mixture of unitaries $\Lambda^{(n)}$, such that*

(1) The resulting trajectory $\rho_Q(t)$ on Q does not equilibrate.

(2) The state of the resource remains unchanged upon application of the channel:
$\rho_R^{(n)} := \mathrm{Tr}_Q(\Lambda^{(n)}(\sigma_Q^{(n)} \otimes \sigma_R^{(n)})) = \sigma_R^{(n)}.$

(3) The change in resilience of the systems $\Delta\mathcal{R}_Q^{(n)}$ diverges with $n \to \infty$.

(4) The correlation between R and Q as measured by the mutual information remain arbitrarily small:

$$I(Q:R) = D_1(\Lambda^{(n)}(\sigma_Q^{(n)} \otimes \sigma_R^{(n)})\|\rho_Q^{(n)} \otimes \rho_R^{(n)}) \leq \epsilon. \tag{8.29}$$

Let us discuss the implication of this theorem. Put into words, it states that we can find arbitrary large many-body systems Q and R and operations on RQ with the following properties: i) Q is initially stationary, ii) Q finally does not equilibrate, iii) the state on R *does not change* in the process, i.e., in the words of chapter 2 it is a catalyst, iv) arbitrary little correlations between Q and R are established (as measured by the mutual information). We thus conclude that it is, in principle, not necessary to *spend* the resilience in R in order to bring Q to a state that does not equilibrate.

Now suppose that Alice initially has M copies Q_j of Q which are initially uncorrelated. Then she can repeat this procedure: According to the theorem, she can use *a single system R* to bring *all* of the systems Q_j out of equilibrium by first applying the channel R to R and Q_1, then to R and Q_2 etc. In other words, it is, in principle, possible to bring arbitrary many systems out of equilibrium without spending any resilience to equilibrate.

How is this result compatible with theorem 8.1? After all it seems that the total change of resilience of all the M systems Q_j is arbitrarily large. However, this is not true. In the process of bringing all the systems Q_j out of equilibrium, these systems become correlated with each other. Furthermore due to the fact that the systems Q_j are copies of the same system, the total Hamiltonian $H_{Q_1 \cdots Q_M}$ is non-interacting and has huge degeneracies in the spectrum. This results in the fact that time-averaging all the copies $Q_1 \ldots Q_M$ creates a large amount of correlations and we have

$$\omega_{H_{Q_1 \cdots Q_M}} \neq \omega_{H_{Q_1}} \otimes \omega_{H_{Q_2}} \otimes \cdots \otimes \omega_{H_{Q_M}}. \tag{8.30}$$

This in turn implies that

$$\Delta \mathcal{R}_{Q_1 \cdots Q_M} \neq \sum_j \Delta \mathcal{R}_{Q_j} = M \Delta \mathcal{R}_{Q_1}. \tag{8.31}$$

Put in different words: It makes a huge difference if one tries to bring out of equilibrium many non-interacting systems, considered as one large system, which is easy in the sense of theorem 8.4, or to bring out of equilibrium one large, interacting system, which is extremely difficult as shown by theorem 8.1.

It is important to note that theorem 8.4 is an "in principle" result, since we allowed for arbitrary channels Λ that can be applied with arbitrary precision. Considering the results in chapter 7, it seems plausible that essentially any small perturbance or imperfection in the implementation of the operation Λ would result in a state on Q (or the Q_j) that equilibrates. Similarly, if there would be any, even arbitrarily small, interaction between the systems Q_j it would not be possible to bring any of the M systems out of equilibrium with a single copy of R. Nevertheless, it shows that a "second law of equilibration" in the sense of (8.24) cannot be proven in general. Importantly, theorem 8.4 is independent of the measure that is used to quantify the resilience to equilibration. Thus, the violation of the "second law of equilibration" is not an artefact of the choice of equilibration measure.

Proof of theorem 8.4. The proof of theorem 8.4 rests on a result by Markus P. Müller, which states that a state σ can be brought to a state ρ with the help of a catalyst that can become correlated to the system if and only if the von Neumann entropy increases: $S(\rho) \geq S(\sigma)$ (see chapter 5 for a thorough discussion of catalysts that can be come correlated with a system). In the form that is relevant for the purpose here, we can state it in the following way.

Theorem 8.5 (von Neumann entropy characterizes correlated catalytic transitions [164])**.**
For any two finite-dimensional states σ_Q and ρ_Q of same dimension and with $S(\sigma_Q) \leq S(\rho_Q)$, any $\delta I > 0$ and any $\epsilon > 0$, there exists a finite-dimensional state σ_R and a mixture of unitaries Λ such that

(a) *The channel produces the state ρ_Q on Q to accuracy ϵ:*
    $\left\| \rho_Q - \mathrm{Tr}_R(\Lambda(\sigma_Q \otimes \sigma_R)) \right\|_1 < \epsilon,$

(b) *The state on R after Λ coincides with the state in which it was originally given:*
    $\sigma_R = \mathrm{Tr}_Q(\Lambda(\sigma_Q \otimes \sigma_R)) =: \rho_R,$

(c) *The mutual information between R and Q after Λ has acted is upper bounded by δI:*
    $D_1(\Lambda(\sigma_Q \otimes \sigma_R) \| \rho_Q \otimes \rho_R) \leq \delta I.$

The Hilbert-space dimension of R may in general depend on both ϵ and δI.

Note that in the theorem we can choose the eigenbasis of σ_R as we wish. In particular, if we pick some arbitrary Hamiltonian $H_R^{(n)}$ on R, we can always choose the state σ_R to be diagonal in the energy basis. In the following construction, the initial and final Hamiltonians coincide, but they can also be chosen differently.

Consider a sequence of n-partite systems with Hamiltonians $H_Q^{(n)}$ of dimension d^n and the sequence of states

$$\rho_Q^{(n)} := a \, |\Psi\rangle\langle\Psi| \, \Psi + (1-a)\frac{\Pi}{d^n - 2}. \tag{8.32}$$

Here, $\Psi = \frac{1}{\sqrt{2}}\left(|E_1\rangle + |E_2\rangle\right)$ with $|E_1\rangle$ and $|E_2\rangle$ two arbitrary energy eigenstates of $H_Q^{(n)}$ and Π the projector onto the subspace orthogonal to $|E_1\rangle$ and $|E_2\rangle$. These states clearly do not equilibrate, since there are persistent Rabi-oscillations with amplitude a and frequency $E_2 - E_1$. Furthermore, the entropy of these states is given by

$$\begin{aligned}
S\left(\rho_Q^{(n)}\right) &= aS(|\Psi\rangle\langle\Psi|) + (1-a)S\left(\frac{\Pi}{d-2}\right) + H_2(a) \\
&= (1-a)\,n\,\log(d) + (1-a)\log(1 - 2d^{-n}) + H_2(a) \\
&\approx (1-a)\,n\,\log(d) + H_2(a),
\end{aligned} \tag{8.33}$$

where $H_2(a) = -a\log(a) - (1-a)\log(1-a)$ denotes the binary entropy and the last equation holds with an error exponentially small in the system size. Thus the entropy diverges with the system size. The effective dimension approaches a constant, on the other hand:

$$d_{\text{eff}}\left(\rho_Q^{(n)}, H_Q^{(n)}\right) = \frac{1}{a^2 + \frac{(1-a)^2}{d^n - 2}} \leq \frac{1}{a^2}. \tag{8.34}$$

The states $\rho_Q^{(n)}$ will be the final states on the system Q. Let us now construct a sensible class of initial states. From theorem 8.5, we see that it is sufficient to have any states $\sigma_Q^{(n)}$ that are stationary and fulfill

- Sufficiently small entropy: $S(\sigma_Q^{(n)}) < S(\rho_Q^{(n)})$,

- Diverging effective dimension: $d_{\text{eff}}(\sigma_Q^{(n)}, H_Q^{(n)})$ diverges with the system size.

To achieve this, we can simply consider any micro-canonical state with a microcanonical window of dimension $d^{\gamma n}$ for some $\gamma < a$. Such a state has both entropy and effective dimension given by $\gamma n \log(d)$. The constant γ depends on the effective temperature of the micro-canonical state, but as long as $\gamma < a$, both conditions are fulfilled. Another possible choice is given by stationary states with a finite correlation length with entropy-density smaller than $a \log(d)$, since it has been proven that states with finite correlation length have an effective dimension that diverges with the system-size (see section 7.5.2). □

To complete this section, let us also discuss the behaviour of the resilience of R as a function of the correlations. We will see that the resilience of R has to diverge as $\epsilon \to 0$. To understand this, suppose that this was not true. Then there would exist constants $C^{(n)}$ such that

$$\mathcal{R}(\sigma_R^{(n)}, H_R^{(n)}) \leq C^{(n)} \quad \text{for all} \quad \epsilon \geq 0. \tag{8.35}$$

Let us now fix a system-size and drop the n-labels everywhere for notational simplicity. Then we can use one system σ_R to bring m identical systems Q_j to a final state $\rho_{Q_1 \cdots Q_m}$ that does not equilibrate. Using theorem 8.1 we then have

$$C \geq \mathcal{R}(\rho_{Q_1 \cdots Q_m}, H_{Q_1 \cdots Q_m}) - m\mathcal{R}(\sigma_{Q_1}, H_{Q_1}). \tag{8.36}$$

Now, all the systems Q_j have pair-wise correlations that are bounded by ϵ in terms of mutual information, since the correlations have been established through R and R is correlated only by an amount ϵ with each of the systems. Thus $I(Q_i : Q_j) \leq \epsilon$. We can now take the limit $\epsilon \to 0$. Since the mutual information vanishes only for product-states we then obtain

$$\lim_{\epsilon \to 0} \rho_{Q_1 \cdots Q_m} = \rho_{Q_1} \otimes \cdots \otimes \rho_{Q_m}. \tag{8.37}$$

We can now use the super-additivity of the resilience on product-states to obtain

$$C \geq m \left(\mathcal{R}(\rho_{Q_1}, H_{Q_1}) - \mathcal{R}(\sigma_{Q_1}, H_{Q_1}) \right). \tag{8.38}$$

Since m is arbitrary we obtain a contradiction.

8.4 Impossibility equilibration bounds in terms of von Neumann entropy

In the previous chapter, I briefly mentioned that the results in this chapter show that we cannot find similar equilibration bounds in terms of the von Neumann entropy instead of the Rényi-2 entropy of the time-averaged state $\omega_H(\rho)$. Let us now discuss this in some more detail.

Imagine that we could find an equilibration bound akin to (7.1) in terms of the von Neumann entropy. In this case it would be sensible to define a corresponding resilience

$$\tilde{\mathcal{R}}(\rho, H) := D \left(\omega_H(\rho) \| \mathbf{1}/d \right), \tag{8.39}$$

where $D = D_1$ again notes the quantum relative entropy. Due to the super-additivity of the relative entropy (see chapter 5), we would then find that the resilience $\tilde{\mathcal{R}}$ is super-additive even for correlated states:

$$\tilde{\mathcal{R}}(\rho_{12}, H_1 \otimes \mathbf{1} + \mathbf{1} \otimes H_2) \geq \tilde{\mathcal{R}}(\rho_1, H_1) + \tilde{\mathcal{R}}(\rho_2, H_2). \tag{8.40}$$

This in turn implies that the corresponding "second law of equilibration" would *always* be fulfilled:

$$\Delta \tilde{\mathcal{R}}_Q \leq -\Delta \tilde{\mathcal{R}}_R. \tag{8.41}$$

But this is in conflict with theorem 8.4. Indeed, we can also see directly from the proof of theorem 8.4 that the states $\rho_Q^{(n)}$ constructed in the proof have the property that they do not equilibrate but the von Neumann entropy of the corresponding time-averaged states diverges linearly with the system size.

In fact, theorem 8.4 shows that there cannot be any meaningful measure of resilience to equilibrate, which is also super-additive.

8.5 Conclusions and Outlook

Equilibration is an ubiquitous feature of complex of physical systems. In the previous chapter, we saw that this fact can be explained if natural quantum states have large overlap with many energy eigenstates. Heuristically, this follows from the fact that the density of states grows exponentially with the system size for reasonable energy densities and generic many body systems. Any, ever so small, uncertainty in the energy density is then compatible with an exponentially large effective dimension. Nevertheless, this argument is neither really quantitative, nor is it formulated in an operational manner. In this chapter we showed that one can indeed derive quantitative bounds from an operational perspective that establish that it is extremely difficult to prepare large quantum systems in states that do not equilibrate. We also elucidated the role of correlations that can be established with resource systems. As a side result, we found that it is impossible to formulate equilibration bound in terms of the Shannon entropy of the energy distribution and showed that this fact is intimately connected to the role of correlations in a "second law of equilibration".

From a formal point of view, we used techniques from the resource theory of coherence and purity. The results in this chapter show that the quantities studied in these resource

theories [56, 57, 107–111, 113–115] are indeed useful to study fundamental problems in the foundations of statistical mechanics by providing a link between the resource theoretical formulation of quantum thermodynamics and the problem of emergence of equilibrium statistical mechanics in complex quantum systems.

This finishes the first part of this thesis, dealing with fundamental problems. We will now turn to more practically relevant problems by studying how limited control abilities and strong interactions with heat baths influence fundamental thermodynamic bounds on work extraction and efficiencies of thermal machines.

9 Leaving the ideal world: Thermal machines under control restrictions

IN THE PREVIOUS CHAPTERS, I discussed thermodynamics from a fundamental perspective in the framework of resource theories. Fundamental here means that the focus of the discussion was not given by concrete physical models or concrete physical effects in certain situations, but by the basic conceptual and mathematical structure of thermodynamic laws that can be derived from minimal assumptions. The framework of resource theories allowed us to take an operational stance while discussing such abstract questions. And indeed we have seen that we can answer a great variety of questions: how the second law of thermodynamics arises and can be quantified; how the third law of thermodynamics arises and can be quantified; how we can define work from an operational point of view in settings that do not adhere to the traditional assumptions; how the non-equilibrium free energy emerges as a unique figure of merit and how this is related to neglecting correlations; and finally, how one can rigorously derive statistical ensembles and the associated laws of phenomenological thermodynamics by forgetting about the detailed microscopic description of systems.

While such general results are very satisfying from a theoretical and conceptual perspective, they also have the downside that they are indeed "only" fundamental laws of thermodynamics, assuming perfect experimental capabilities (within certain limits). In concrete experimental situations, experimentalist can never implement arbitrary energy-preserving unitary operations, as assumed in the framework of thermal operations. Instead, they will usually be able to control only a few parameters and face additional physical restrictions. For example, an experimenter usually cannot modify the inherent interaction between particles in a many-body system directly and arbitrarily, but only apply external fields. But this might not be enough to implement optimal protocols and saturate thermodynamic bounds.

It is thus an interesting question to study how thermodynamic bounds are modified if such additional restrictions are taken into account. This will be the topic of this and the following chapters[1].

Unfortunately, the resource-theoretical framework of thermal operations is not particularly suited for such a study. This is due to the fact that it requires to always incorporate all physical systems into the description. But if we consider an experiment in a, say, quantum optics laboratory we clearly do not want to (and cannot) describe every laser in the room and the power plant (work-storage device) which supplies the necessary electrical power explicitly as quantum mechanical systems.

It is thus desirable to resort to a more effective model, which only explicitly accounts for

[1] This chapter is based on work together with Rodrigo Gallego, Jens Eisert and Jaqueline Lekscha, published in Refs. [1, 6].

the degrees of freedom under study and can hence be analyzed in concrete situations. We will thus first be faced with setting up a basic framework in which we can study such questions. Our aim will be to be as concrete as possible while still being as model-independent as possible. That means that we want to end up with a framework that can be applied to a multitude of concrete physical situations, but is broad enough to allow to derive general results.

I will then discuss within this simplified model, which is however still quite general, how limited control capabilities influence basic thermodynamic tasks. The focus will be on how such limitations influence how much work can be extracted from a non-equilibrium system – either in the case of a single run using a single heat bath, which ends with the system that was initially in thermal equilibrium reaching thermal equilibrium, or in the case of a thermal machine operating cyclically between two heat baths. In this chapter, we will essentially assume that the contact of systems to heat baths is "weak" in a certain sense. In the later chapter 11, I will discuss in detail how strong coupling to heat baths and the possibility of non-thermalizing heat baths influence thermodynamics.

9.1   Simplifying the model

Let us now begin to develop an effective model in which we can describe the task of work extraction. Just as before, we will describe this model in terms of the operations that an experimenter is allowed to perform. To obtain non-trivial thermodynamics and work-extraction we need to be able to couple the system to a heat bath and we need to be able to perform operations which in some way allow for performing work on the system. In the following, we will distinguish these two kind of operations as

1. Thermal contacts: bringing the system at hand in contact with a heat bath for some amount of time,

2. Adiabatic evolution: changing the Hamiltonian of the system in a time-dependent way while it is not coupled to a heat bath.

Since the above operations include changing the Hamiltonian over time, we need to keep track of both the Hamiltonian and the quantum state of the system over time. The basic objects that we will be working with will thus again be pairs of quantum states of Hamiltonians, which I will write as (ρ_t, H_t). I will now describe each of these classes of operations in more detail. For the moment, this will still be on a very general level and to simplify the description, I will for now only consider a single heat bath at inverse temperature β. The generalization to several heat baths or additional particle reservoirs etc. is straight-forward.

9.1.1   Thermal contacts

The first kind of operation that we consider is that of bringing the system in contact with the heat bath for some amount of time. From an abstract point of view the effect of such a coupling is described by some quantum channel G acting on the quantum state of the system, but leaving the Hamiltonian unchanged:

$$(\rho_t, H_t) \mapsto (G(\rho_t), H_t). \tag{9.1}$$

The precise form of the channel G depends on how exactly, and for how long the system is coupled to the heat bath. At any time, there will thus in general be a whole set $\mathcal{G}_t$ of possible channels G that could correspond to a valid thermal contact. The defining feature of any such thermal contact is that if the system is already at thermal equilibrium, its state cannot change. We thus require that all the channels in $\mathcal{G}_t$ preserve the Gibbs-state:

$$G(\omega_\beta(H_t)) = \omega_\beta(H_t), \quad \forall G \in \mathcal{G}_t. \tag{9.2}$$

We can distinguish different levels of control that an experimenter has over the heat bath and the coupling between the heat bath and the system by restricting the set $\mathcal{G}_t$. For example, we could assume that the experimenter has extremely little control, in which case it

would be reasonable to assume that the only channel she can implement is the *thermalizing map T_t*:

$$T_t(\rho_t) = \omega_\beta(H_t) \quad \forall \rho_t. \tag{9.3}$$

Such thermalization happens if the experimenter couples the system weakly to a heat bath for a sufficiently long, random time. We could also assume that the experimenter can weakly couple the system to the heat bath for a well-defined time and that during this time the dynamics of the system is described by a Markovian master equation. Then the system would evolve towards a thermal state but maybe not thermalize completely. We could also consider the extreme situation where the experimenter has arbitrary quantum control over system and bath, in which case she could implement arbitrary Gibbs-preserving channels. For example, she could implement the anomalous heat flow, described in chapter 2. We will later see when and how such different levels of control influence how much work can be extracted from a system out of equilibrium in certain situations.

In the above description, I have always labelled states, Hamiltonians and sets of channels with the letter t, suggesting that they are time-dependent quantities. It is important to recognize that the sets $\mathcal{G}_t$ depend on t, but only implicitly through the Hamiltonian H_t, $\mathcal{G}_t = \mathcal{G}(H_t)$. This is due to the fact that the condition (9.2) depends on the Hamiltonian and therefore is time-dependent.

In this chapter, I will not account for the time it takes to couple the system to the heat bath. That is, we treat the thermalizing maps $G \in \mathcal{G}_t$ as happening instantaneously, even though in reality they take a finite time to be implemented. This is not problematic, since in this chapter we are not considering the power of thermal machines.

Coupling the system to the heat bath with channel G will in general lead to a change of energy on the system, which we interpret as *heat* since no other systems than a heat bath and the working system are involved. We will interpret energy absorbed from the heat bath as positive heat Q_t:

$$Q_t = \mathrm{Tr}(H_t(G_t(\rho_t) - \rho_t)), \tag{9.4}$$

where G_t is the channel that describes the thermal contact at time t. This completes the general description of thermal contacts. Let us now turn to adiabatic evolutions.

9.1.2   *Adiabatic evolution*

The second kind of operations that I will consider adiabatic evolutions, in which the experimenter modifies the Hamiltonian of the system in a time-dependent way, for example, by altering external fields. The effective evolution then acts as

$$(\rho_t, H_t) \mapsto (\rho_{t+\tau}, H_{t+\tau}) = (U_{t \to t+\tau} \rho_t U^\dagger_{t \to t+\tau}, H_{t+\tau}), \tag{9.5}$$

where

$$U_{t \to t+\tau} = \mathcal{T} \exp\left(-\mathrm{i} \int_t^{t+\tau} H_{t'} \, \mathrm{d}t'\right). \tag{9.6}$$

As discussed above, in general an experimenter cannot modify the Hamiltonian H_t arbitrarily. Similarly as the set of of possible thermal contacts, there will therefore also be a set $\mathcal{H}$ of possible Hamiltonians that can be implemented. We will thus assume that $H_t \in \mathcal{H}$ in the following and assume that every to Hamiltonians in $\mathcal{H}$ can be connected by a smooth curve in $\mathcal{H}$. The set $\mathcal{H}$, reflecting the control capabilities of the experimenter during adiabatic evolution, will have important consequences on the possible thermodynamic protocols that can be implemented by the experimenter. Furthermore, since the set of possible thermal contacts depend on the current Hamiltonian, the set $\mathcal{H}$ also indirectly influences the possible thermal contacts. This will turn out to be crucial in certain situations.

As in the case of thermal contacts, adiabatic evolutions in general do not preserve the energy of the system. But since in this case no coupling to heat baths is involved, we consider the change of average energy as *work $W(t \to t + \tau)$* in this case:

$$W(t \to t + \tau) := \mathrm{Tr}\left(H_t \rho_t - H_{t+\tau} \rho_{t+\tau}\right). \tag{9.7}$$

This definition is subject to the condition that no thermal contact has occurred between times t and $t + \tau$. I will say that work is *extracted* if $W(t \to t + \tau) \geq 0$. These definitions ensure the first law of thermodynamics in terms of average energy; denoting the total change of average energy between t and t' by $\Delta E(t \to t')$ and the total absorbed heat as $Q(t \to t')$, we then have

$$\Delta E(t \to t') := \mathrm{Tr}(H_t \rho_t) - \mathrm{Tr}(H_{t'} \rho_{t'}) = W(t \to t') - Q(t \to t'). \qquad (9.8)$$

As discussed in detail in chapter 4, when we say "work", in principle we have to specify in detail what we mean by this, that is, what kind of physical work-storage devices we have in mind and what kind of transformations of them we allow for. How is this compatible with the above ad-hoc definition of work?

It turns out that it is always possible to use an auxiliary system, which models an idealization of a lifted weight, and apply an energy-conserving unitary on the working system together with this idealized weight such that the dynamics on the working system is given to arbitrary precision by a prescribed unitary $U_{t \to t+\tau}$ (also see section 2.5). By energy-conservation, the average change of energy on the weight is then given by the work $W(t \to t + \tau)$. Furthermore, this change of average energy does not depend on the initial average[2] "height" of this weight and its entropy remains approximately constant. The change of free energy on this weight then therefore coincides with its change of average energy and $W(t \to t + \tau)$ can be seen as measured by this idealized weight as a work-storage device. The construction of such an idealized weight and the corresponding energy-preserving unitaries are, for example, discussed in detail in Refs. [54, 60, 80, 232]. A further way to motivate this definition is from classical mechanics, where the mechanical work associated to a trajectory $(q(t), p(t))$ undergoing evolution under a time-dependent Hamiltonian with control parameter λ can be written as

$$W = \int \frac{\partial H}{\partial \lambda}(q(t), p(t)) \frac{\partial \lambda}{\partial t} \mathrm{d}t. \qquad (9.9)$$

Eq. (9.7) is simply the verbatim translation of this quantity to the quantum setting when a probability distribution over trajectories is included in the description.

9.1.3 Thermodynamic protocols and isothermal processes

After having discussed the possible operations, we can combine them to construct a *thermodynamic protocol* $\mathcal{P}$. This is simply a prescription of an arbitrary sequence of alternating adiabatic evolutions and thermal contacts to one (or more) heat baths. Let us adopt the convention that every such protocol begins with an adiabatic evolution and ends with a thermal contact [3]. To every protocol there correspond initial and final Hamiltonians $H^{(i)}(\mathcal{P})$ and $H^{(f)}(\mathcal{P})$, respectively.

Let t_i denote the time right before the i-th thermal contact. Then we can label the channels representing thermal contacts by $G_i = G_{t_i}$ and the unitary time-evolution that results from the driving between t_{i-1} and t_i as $U_i = U_{t_{i-1} \to t_i}$. Any thermodynamic protocol $\mathcal{P}$ can then be specified as a list of triples (t_i, U_i, G_i) together with the initial Hamiltonian $H^{(i)}(\mathcal{P})$. I will call a protocol *cyclic* if $H^{(i)}(\mathcal{P}) = H^{(f)}(\mathcal{P})$.

A thermodynamic protocol $\mathcal{P}$ maps any initial state ρ_{t_0} to some final state $\mathcal{P}(\rho_{t_0})$. Associated to any protocol and initial state ρ_{t_0} are the total work $W(\mathcal{P}, \rho_{t_0}, H_{t_0})$ and heat $Q(\mathcal{P}, \rho_{t_0}, H_{t_0})$. In the case of multiple heat baths at different temperatures we can furthermore split up the heat into the contributions from the different heat baths. To prevent further notational overload, I will not introduce further special notation for such cases for now.

Let us now discuss a specific example of a thermodynamic protocol, which will turn out to be of great importance. Suppose our system is initially in thermal equilibrium with respect to the initial Hamiltonian $H^{(i)} = H_{t_0}$. Then we want to extract as much work as possible using a protocol that changes the initial Hamiltonian $H^{(i)}$ to some final Hamiltonian $H^{(f)} = H_{t_N}$ in N steps with the state of the system ending as the equilibrium state of the final Hamiltonian $\rho_{t_n} = \omega_\beta(H^{(f)})$. To achieve this we mimic an isothermal quasistatic process from phenomenological thermodynamics. Pick any sufficiently smooth curve

[2] In the sense of the expectation value of the position operator in the initial state.

[3] This is without loss of generality if we assume that the sets $\mathcal{G}_t$ always include the identity channel, i.e., the operation of doing nothing.

$H_t : [0, 1] \mapsto \mathcal{H}$ such that $H_0 = H^{(i)}$ and $H_1 = H^{(f)}$. Since we assumed $\mathcal{H}$ to be path-connected this is always possible. Now discretize the curve into N steps $t_i = i/N$. In each step, we first quickly change the Hamiltonian from $H_{t_{i-1}}$ to H_{t_i}. Then we let the system thermalize completely. In the i-th step, we hence extract the amount of work

$$W(t_{i-1} \to t_i) = \text{Tr}\left(\omega_\beta(H_{t_{i-1}})H_{t_{i-1}}\right) - \text{Tr}\left(\rho_{t_i}H_{t_i}\right), \tag{9.10}$$

with $\rho_{t_i} = U_{t_i}\omega_\beta(H_{t_{i-1}})U_{t_i}^\dagger$. Since entropy is conserved in this process, we can rewrite this work-value using non-equilibrium free energies as

$$W(t_{i-1} \to t_i) = F_\beta(\omega_\beta(H_{t_{i-1}}), H_{t_{i-1}}) - F_\beta(\rho_{t_i}, H_{t_i}). \tag{9.11}$$

Summing over i and re-arranging we can then write the total extracted work as

$$W(0 \to 1) = \sum_{i=1}^{N} F_\beta(\omega_\beta(H_{t_{i-1}}), H_{t_{i-1}}) - F_\beta(\rho_{t_i}, H_{t_i})$$

$$= F_\beta(\omega_\beta(H_0), H_0) - F_\beta(\omega_\beta(H_1), H_1) - \sum_{i=1}^{N-1} F_\beta(\rho_{t_i}, H_{t_i}) - F_\beta(\omega_\beta(H_{t_i}), H_{t_i})$$

$$= F_\beta(\omega_\beta(H_0), H_0) - F_\beta(\omega_\beta(H_1), H_1) - \sum_{i=1}^{N-1} \Delta F_\beta(\rho_{t_i}, H_{t_i}) \tag{9.12}$$

$$\leq F_\beta(\omega_\beta(H_0), H_0) - F_\beta(\omega_\beta(H_1), H_1), \tag{9.13}$$

where we have used that ΔF_β is a positive semi-definite quantity. In the limit $N \to \infty$, the protocol changes the Hamiltonians only in an infinitesimal way in every step, keeping the system essentially at equilibrium while changing the Hamiltonian. It is then not surprising that it is possible to show that the correction term can be made arbitrarily small in this limit:

$$\lim_{N \to \infty} \sum_{i=1}^{N-1} \Delta F_\beta(\rho_{t_i}, H_{t_i}) = 0. \tag{9.14}$$

The proof of this is given in section 14.6.1. The optimal protocol that achieves this bound is the one simply *quenching* the Hamiltonians between the thermalization processes, so that $U_{t_i} = \mathbf{1}$ for all t_i. In the following, I will refer to this protocol between two equilibrium states as an *isothermal proccess*.

It is important to note that the calculation for the upper bound (9.13) on the work did not depend on the particular protocol connecting the initial and final equilibrium states. Furthermore, we have only used fully thermalizing contacts to the heat bath.

In phenomenological equilibrium thermodynamics, it is well known that the optimal amount of work is achieved by a *reversible* protocol. That is any protocol that returns the system to its initial condition when run backward on the final condition of the system after the system was subjected to the protocol. In this case the extracted work is only a function of the initial and final states and Hamiltonians.

The isothermal process and its analysis shows that (arbitrarily) reversible protocols between equilibrium states also exist in this framework and that the optimal work that can be extracted by a protocol connecting two equilibrium states is given by the value obtained by the isothermal protocol.

This finishes the basic introduction of the framework. We will from now on also consider non-equilibrium states. In the rest of the chapter, we will discuss questions of the following kinds: Given some arbitrary initial state, sets $\mathcal{G}_t$ and $\mathcal{H}$, how much work can be extracted using cyclic protocols using a single heat bath? Given access to two heat baths, what is the maximum efficiency that can be achieved by cyclic thermal machines? How do these quantities depend on $\mathcal{H}$ and $\mathcal{G}_t$? Are the restrictions on $\mathcal{H}$ and $\mathcal{G}_t$ independent?

9.2   The fundamental bound for a single heat bath

Let us now begin to analyze the fundamental thermodynamic bounds that arise in this framework. First, we consider the case of a single heat bath and ask for the largest value of

total work $W(\mathcal{P}, \rho, H)$ that can be extracted from (ρ, H) using only cyclic protocols, i.e., those whose final Hamiltonian coincides with H. In general, this value is a function of the sets $\mathcal{H}$ and $\mathcal{G}_t$ as well as ρ and H. In analogy to chapter 4, let us denote this optimal work by

$$W_{\text{value}}(\rho, H | \mathcal{H}, \mathcal{G}) := \sup_{\mathcal{P} | H, \mathcal{H}, \mathcal{G}} W(\mathcal{P}, \rho, H), \qquad (9.15)$$

where $\sup_{\mathcal{P} | H, \mathcal{H}, \mathcal{G}}$ denotes optimization over all cyclic protocols starting with Hamiltonian H and subject to the constraint of being compatible with the given $\mathcal{H}$ and $\mathcal{G}_t$.

Before considering non-trivial restrictions on the set of allowed Hamiltonians, let us first investigate $W_{\text{value}}(\rho, H | \mathcal{H}, \mathcal{G})$, when $\mathcal{H}$ is fully unrestricted. This corresponds to the second law of thermodynamics in terms of work extraction and the bound that we will obtain is well-known, although under different sets of assumptions (see, for example, [54, 233–235]). To state the corresponding result, recall that T_t denotes the thermalization map with respect to Hamiltonian H_t.

Theorem 9.1 (second law). *If $\mathcal{H}$ is unrestricted and $T_t \in \mathcal{G}_t$ for all t, we have*

$$W_{\text{value}}(\rho, H | \mathcal{H}, \mathcal{G}) = \Delta F_\beta(\rho, H). \qquad (9.16)$$

Proof. Here, I will only show how the above value can be achieved. For optimality of the protocol, I refer to section 14.6.2. Any full-rank state can be thought of as a thermal state of the modular Hamiltonian H_ρ, introduced in chapter 5. If ρ does not have full rank, it can be arbitrarily well approximated by such a thermal state. Then construct the following protocol: First change ("quench") the Hamiltonian H to H_ρ very quickly, such that the first unitary coincides with the identity: $U_{t_1} = \mathbf{1}$. This can be done to arbitrary accuracy and has the result that the system is in equilibrium. The step has work-cost $\text{Tr}(\rho H - \rho H_\rho) = F_\beta(\rho, H) - F_\beta(\rho, H_\rho)$. Then perform an isothermal process back to Hamiltonian H, with work-cost $F_\beta(\rho, H_\rho) - F_\beta(\omega_\beta(H), H)$, again to arbitrary accuracy. The total extracted work is hence given by $\Delta F_\beta(\rho, H)$. $\qquad \square$

Note that the optimal protocol constructed in the proof is in fact reversible. This result has three important implications. First, no protocol can ever extract more work than the non-equilibrium free energy. In particular, no work can be extracted from a state at thermal equilibrium. The framework we have constructed thus indeed allows us to derive fundamental thermodynamic bounds.

Second, if the Hamiltonians are unrestricted this optimal value can already be reached if an experimenter has very poor experimental control on the coupling to the heat bath. Indeed it is completely sufficient to be able to let the system thermalize.

Third, it shows that letting the system fully thermalize is *universal* for work-extraction if Hamiltonian control on the working system is unrestricted: For example, even if an experimenter were capable of controlling in detail the microscopic interactions between all the heat bath degrees of freedom and the system's degrees of freedom with arbitrary energy-preserving unitaries (that is, she could implement any thermal operation), this would not yield any advantage from the point of view of work-extraction.

9.3   Thermalizing contacts are not universal

Let us now investigate how the bound (9.16) changes when we restrict the set of Hamiltonians $\mathcal{H}$. To do that it is useful to first restrict thermal contacts to only allow for full thermalizations. Let us denote the corresponding sets of channels by $\mathcal{T}_t := \{\mathbf{1}, T_t\}$. In this case, we can derive a general and tight bound as a function of $\mathcal{H}$. To state it, let $\mathcal{U}_\mathcal{H}$ denote the unitary group generated by all time-dependent Hamiltonians within $\mathcal{H}$ and let $\mathcal{U}_\mathcal{H}[\rho]$ be the set of states that can be reached from ρ by all such unitaries. Then we have the following result.

Lemma 9.2 (Restricted second law).

$$W_{\text{value}}(\rho, H | \mathcal{H}, \mathcal{T}) = \Delta F_\beta(\rho, H) - \inf_{\substack{\sigma \in \mathcal{U}_\mathcal{H}[\rho], \\ H_t \in \mathcal{H}}} \Delta F_\beta(\sigma, H_t). \qquad (9.17)$$
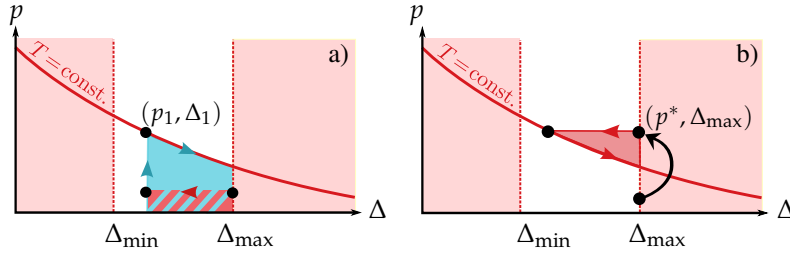
Figure 9.1: An example exhibiting a constraint on field-strengths. p is the excitation probability of the two-level system and Δ is the energy in the excited state. The light red region corresponds to field-strength that cannot be experimentally achieved. a) Example of a protocol using thermalizing contacts only. The system starts at maximum energy gap $\Delta_{\max}$. The first transition $(p_0, \Delta_{\max}) \rightarrow (p_0, \Delta_1)$ gives a positive amount of work equal to the red dashed area. Then, the system is put in contact with the heat bath, leaving the system in state (p_1, Δ_1). The last step is to come back to the original Hamiltonian (in this example using an isothermal reversible process). The minimum work that is spent in this last step corresponds to the area of the blue solid region. Clearly, the total work is negative and this is true for any protocol that only uses thermalizing maps and does not cross the light red region. b) The black arrow denotes an anomalous heat flow (see 2) as initial thermal contact. Then the transition $(p^*, \Delta_{\max}) \rightarrow (p^*, \Delta_1)$ takes place, followed by an isothermal reversible process back to the initial Hamiltonian with gap $\Delta_{\max}$. The total extracted work is positive and corresponds to the red solid area. (Figure adapted from Ref. [1].)

Proof. Again, here I only proof achieveability, but not optimality (see section 14.6.3). Consider the following protocol. First evolve the system to some Hamiltonian H_t using a unitary U, resulting in the state $\sigma = U\rho U^\dagger$ with work-cost $tr(\rho H) - \mathrm{Tr}(\sigma H_t) = F_\beta(\rho, H) - F_\beta(\sigma, H_t)$. Then thermalize the system, which will result in some dissipation, and come back to H using an isothermal process. The total work is given by:

$$W(\mathcal{P}, \rho, H) = F_\beta(\rho, H) - F_\beta(\sigma, H_t) + F_\beta(\omega_\beta(H_t), H_t) - F_\beta(\omega_\beta(H), H)$$
$$= \Delta F_\beta(\rho, H) - \Delta F_\beta(\sigma, H_t). \tag{9.18}$$

Since $\Delta F_\beta(\sigma, H_t) \geq 0$, minimizing over all U and H_t yields the desired result. $\square$

It is now natural to ask, whether the fully thermalizing thermal contacts $\mathcal{T}_t$ are also universal in the case of restricted Hamiltonians. I will now illustrate that this is not the case. To do that I will discuss two examples in which $W_{\text{value}}(\rho, H | \mathcal{H}, \mathcal{T}) = 0$ but $W_{\text{value}}(\rho, H | \mathcal{H}, \mathcal{G}_{\text{TO}}) > 0$, where $\mathcal{G}_{\text{TO}}$ denotes all thermal operations.

9.3.1 The general argument for non-universality

Before discussing the specific examples, let me explain the general argument that underlies both of them. Again assume that the initial condition is given by (ρ, H). To understand the argument, recall that the optimal protocol in the case of fully thermalizing maps $\mathcal{T}_t$ consists of two steps: First, we quench the Hamiltonian to the modular Hamiltonian H_ρ. Then we implement an isothermal. The first step is not possible if $H_\rho \notin \mathcal{H}$ and results in the penalty term in (9.17). However, if we have access to more general thermal contacts, we might be able to first apply some thermal contact G with the effect that the final states modular Hamiltonian is indeed inside $\mathcal{H}$, i.e., $H_{G(\rho)} \in \mathcal{H}$. This would have the advantage that we could then implement the optimal protocol to extract an amount of work given by $\Delta F_\beta(G(\rho), H)$. On the other hand, this first thermal contact will cost some amount of heat compared to the fundamentally optimal value $\Delta F_\beta(\rho, H)$, given by $F(\rho, H) - F(G(\rho), H)$. Only if this dissipated heat is smaller than the correction term in (9.17) will this procedure outperform the protocol developed in the proof of lemma 9.2.

9.3.2 Example: Non-universality from restrictions on field strengths

Let us now illustrate the previous argument in a simple way. The first example explores the setting where an experimenter only has limited field strengths available. To keep the problem as simple as possible, consider a two-level system and consider as set of allowed Hamiltonians the set

$$\mathcal{H}_{\text{field}} = \left\{ \Delta \, |e\rangle\langle e| \, | \, \Delta_{\min} \leq \Delta \leq \Delta_{\max}, \, |e\rangle \in \mathbb{C}^2, \langle e|e\rangle = 1 \right\}. \tag{9.19}$$

This set is unitarily invariant, and the set of unitaries it generates is the full special unitary group $SU(2)$. Each such Hamiltonian can be thought of as the Hamiltonian of a single spin-1/2 in some arbitrarily oriented magnetic field, where the field strength is restricted to lie between $\Delta_{\min}$ and $\Delta_{\max}$.

Now assume that the density matrix of the initial state ρ_0 commutes with the initial Hamiltonian $H_0 = \Delta |e\rangle\langle e|$. Thus $\rho_0 = p_0 |e\rangle\langle e| + (1 - p) |g\rangle\langle g|$. For concreteness assume that the initial Hamiltonian H_0 has $\Delta = \Delta_{\max}$ and that the excitation probability p_0 is smaller than the thermal one of H_0, given by $\exp(-\beta\Delta_{\max})/Z_\beta(\Delta_{\max})$. The initial condition then corresponds to a non-equilibrium thermal state that is colder than the heat bath.

We will now first show that no work can be extracted if only fully thermalizing maps $\mathcal{T}_t$ are available. The bound in lemma 9.2 in this setting is always achieved for a Hamiltonian and density matrix after the first adiabatic evolution which are still diagonal in the same basis and where the state remains unchanged. It therefore suffices to show that no work can be extracted if we further restrict the Hamiltonians to only those of the form $\Delta |e\rangle\langle e|$, where the state $|e\rangle$ is now fixed. Then the problem is fully classical and we can describe the system and the Hamiltonian right before the first thermal contact by the pair (p_0, Δ_1). The work-cost of this first step is given by $W_1 = p_0(\Delta_{\max} - \Delta_1)$ with $\Delta_1 \leq \Delta_{\max}$. The second part of the optimal protocol implements an isothermal process back to the initial Hamiltonian, with work-cost $W_2 = F_\beta(\omega_\beta(\Delta_1 |e\rangle\langle e|), \Delta_1 |e\rangle\langle e|) - F_\beta(\omega_\beta(\Delta_{\max} |e\rangle\langle e|), \Delta_{\max} |e\rangle\langle e|)$. It is easy to see in a picture (see Fig. 9.1) that the total work extracted in this process is negative for any $\Delta_1 < \Delta_{\max}$. Hence we have

$$W_{\text{value}}(\rho_0, H_0 | \mathcal{H}_{\text{field}}, \mathcal{T}) = 0. \tag{9.20}$$

The specific initial condition that we chose has the property that no work can be extracted from it, despite being out of equilibrium. Such initial conditions are commonly called *passive* with respect to the given set of operations [236–238].

Having shown passivity of the initial condition with respect to fully thermalizing maps, I will now show that the initial condition looses its passivity when we allow for thermal operations as thermal contacts. To see this, we make use of the *anomalous heat flow*, that I introduced in section 2.2. Using anomalous heat flow, we can first map the initial state to a state $G(\rho_0)$ that is hotter than the heat bath (see Fig. 9.1). After having done this, we can now find a modular Hamiltonian $H_{G(\rho)} \in \mathcal{H}$, because the modular Hamiltonian of $G(\rho)$ now has a *smaller* gap than $\Delta_{\max}$. We can then use the usual optimal protocol to extract a positive amount of work.

9.3.3   Example: Non-universality from restrictions on locality and interactions

For the second example, consider two spin-1/2 particles which interact through an interaction of the form $H = \sigma_z \otimes \sigma_z$, where σ_z is the Pauli-matrix in z-direction. Suppose an experimenter can only apply external magnetic fields, but she is capable to do so independently on the two spins and also with arbitrary field strength. She can then reach any Hamiltonian of the form

$$\mathcal{H}_{\text{local}} = \left\{ \sigma_z \otimes \sigma_z + \vec{B}_1 \cdot \vec{\sigma} \otimes \mathbf{1} + \mathbf{1} \otimes \vec{B}_2 \cdot \vec{\sigma} \right\}, \tag{9.21}$$

with $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ and $\vec{B}_i \in \mathbb{R}^3$. I will now show that also in this case there are *passive* states for thermalizing contacts, i.e., initial conditions such that no work can be extracted. After showing passivity, I will then also show that those initial conditions are not passive if we consider arbitrary thermal operations as initial conditions.

To show the existence of passive states, we make use of the *Peierls-Bogoliubov inequality* [34], which can be reformulated as in the following Lemma.

Lemma 9.3 (Upper bound to free energies). *For any two Hamiltonians H_1 and H_B, we have*

$$F_\beta(\omega_\beta(H_1 + H_2), H_1 + H_2) \leq F_\beta(\omega_\beta(H_1), H_1) + \text{Tr}(\omega_\beta(H_1), H_2). \tag{9.22}$$

Proof. A short proof can be found in section 14.6.4. $\square$

Now imagine that there exists a Hamiltonian H_0 in $\mathcal{H}_{\text{local}}$ with the property

$$F_\beta(\omega_\beta(H_0), H_0) \geq F_\beta(\omega_\beta(H_t), H_t) \quad \text{for all } H_t \in \mathcal{H}_{\text{local}}. \qquad (9.23)$$

Since the maximally mixed state $\mathbf{1}/4$ is invariant under unitary operations, Lemma 9.2 then implies that no work can be extracted from $(\mathbf{1}/4, H_0)$. I will now prove that $H_0 = \sigma_z \otimes \sigma_z$ indeed has this property. First, since Z is traceless it follows that $\text{Tr}(H_0 \vec{B} \cdot \vec{\sigma} \otimes \mathbf{1}) = 0$ and the same is true for the other spin. Second, since $\sigma_z^2 = \mathbf{1}$, the Gibbs-state of H_0 takes the form $C(\mathbf{1} + \tanh(\beta)H_0)$ for some constant C. Because all Pauli-matrices are traceless, we therefore also have

$$\text{Tr}\left(\omega_\beta(H_0)\left(\vec{B}_1 \cdot \vec{\sigma} \otimes \mathbf{1} + \mathbf{1} \otimes \vec{B}_2 \cdot \vec{\sigma}\right)\right) = 0 \qquad (9.24)$$

for any fields $\vec{B}_1, \vec{B}_2$. Using, Lemma 9.3 we obtain $F_\beta(\omega_\beta(H_0), H_0) \geq F_\beta(\omega_\beta(H_t), H_t)$ for all $H_t \in \mathcal{H}_{\text{local}}$, which proves the claim.

From the arguments given, it is clear that a similar argument can be made in larger multi-partite systems. Clearly, the maximally mixed state $\mathbf{1}/4$ is out of equilibrium for every $\beta \neq 0$. Thus, work can be extracted from it if arbitrary Hamiltonian control can be achieved. What is left to show, is that also under restricted Hamiltonian control, but with thermal operations as maps for thermal contact, some work can be extracted.

Using a technique called *thermo-majorization* [46, 47, 239], it is possible to show that for $\beta = 1$ and $s \leq s_c := \tanh^{-1}(e^2 - 1/(2e^2)) \simeq 0.46$ it is possible to find a thermal operation that maps the maximally mixed state $\mathbf{1}/4$ to $\omega_\beta(H_s)$ with $H_s = H_0 + s\mathbf{1} \otimes \sigma_z$. Since this technique is somewhat technical, I refer to section 14.6.5 for more details.

Knowing that we can reach the state $\omega_\beta(H_s)$ and that $H_s \in \mathcal{H}_{\text{local}}$, we can now use the optimal protocol to extract an amount of work given by

$$F_\beta(\omega_\beta(H_s), H_s) - F_\beta(\omega_\beta(H_0), H_0) = s \tanh(s) - \log(\cosh(s)) > 0.$$

This shows again that improved control over thermal contacts is useful if Hamiltonian control is restricted.

This finishes the discussion of work-extraction from non-equilibrium states with respect to a single heat bath. We will now move on to discuss thermal machines that cyclically extract work from the temperature difference between two heat baths.

9.4   Thermal machines and efficiencies

So far, we have treated only thermal contacts to a single heat bath. From now on will consider the possibility to put the system in contact with heat baths at two different inverse temperatures β_c and β_h with $\beta_c > \beta_h$ and study machines which extract work by alternatingly combining adiabatic evolution and thermal contact to one of the heat baths. Furthermore, we will consider cyclic protocols – where now a cyclic protocol means that after a given number of steps of the protocol the system returns to both *its initial state and Hamiltonian*, whereas previously only the Hamiltonian returned to its initial condition. Note that in such a setting, a protocol $\mathcal{P}$ is independent of initial conditions: any pair of state and Hamiltonian (ρ_t, H_t) during the protocol returns to its initial state after one cycle of the protocol.

Let us first make some general observations about any such protocol. In particular, we will be interested in the efficiency of the process, which we define in the traditional way as

$$\eta(\mathcal{P}) := \frac{W(\mathcal{P})}{Q_h(\mathcal{P})}, \qquad (9.25)$$

where $W(\mathcal{P})$ is the total work extracted during one cycle and $Q_h(\mathcal{P})$ is the total heat absorbed from the hot reservoir. Unsurprisingly, we obtain Carnot's bound for the efficiency for any cyclic protocol:

Theorem 9.4 (Carnot efficiency). *For any cyclic protocol $\mathcal{P}$ between two heat baths at inverse temperatures $\beta_c > \beta_h > 0$ the efficiency fulfills*

$$\eta(\mathcal{P}) \leq \eta_c := 1 - \frac{\beta_h}{\beta_c}. \qquad (9.26)$$

Furthermore, if the set of Hamiltonians $\mathcal{H}$ is unrestricted, the bound can be saturated.

To prove this theorem, we need to first introduce the notion of a Carnot-like protocol. Let us assume that the protocol starts with the system in the thermal state of Hamiltonian H_1 at the temperature of the hot heat bath. Then such a protocol consists of four parts:

1. An isothermal in contact with the hot bath from Hamiltonian H_1 to Hamiltonian H_2,

2. an adiabatic evolution from Hamiltonian H_2 to Hamiltonian H_3, implementing a unitary transformation U,

3. an isothermal in contact with the cold bath from Hamiltonian H_3 to Hamiltonian H_4,

4. an adiabatic evolution back from H_4 to Hamiltonian H_1, implementing a unitary transformation V.

Importantly, the only dissipation in such a protocol can occur when the system is put in thermal contact with one of the heat baths after the adiabatic evolution. The condition for this dissipation to vanish is that the final state of the adiabatic evolution matches the thermal state at the beginning of the isothermal. Thus, there is no dissipation if and only if

$$U\omega_{\beta_h}(H_2)U^\dagger = \omega_{\beta_c}(H_3) \text{ and } V\omega_{\beta_c}(H_4)V^\dagger = \omega_{\beta_h}(H_1). \tag{9.27}$$

Due to the close correspondence between Hamiltonians and thermal states, we can express this condition as

$$UH_2U^\dagger = \frac{\beta_c}{\beta_h}H_3 \text{ and } VH_4V^\dagger = \frac{\beta_h}{\beta_c}H_1. \tag{9.28}$$

The heat exchanged with the bath in the first isothermal is given by

$$Q_h(\mathcal{P}) = T_h\Delta S_h^{1\to 2} := T_h(S(\omega_{\beta_h}(H_1)) - S(\omega_{\beta_h}(H_2))). \tag{9.29}$$

Similarly, the heat exchanged during the isothermal in contact with the cold bath is given by

$$Q_c(\mathcal{P}) = T_c\Delta S_c^{3\to 4} = -T_c\Delta S_h^{1\to 2} = -\frac{T_c}{T_h}Q_h(\mathcal{P}), \tag{9.30}$$

where we have used that the whole protocol is cyclic and the total change of entropy vanishes along one cycle. Using the first law of thermodynamics we then obtain the efficiency of the protocol as

$$\eta(\mathcal{P}) = \frac{W(\mathcal{P})}{Q_h(\mathcal{P})} = \frac{Q_h(\mathcal{P}) + Q_c(\mathcal{P})}{Q_h(\mathcal{P})} = 1 - \frac{T_c}{T_h}. \tag{9.31}$$

To see that no protocol can exceed Carnot efficiency, observe that any protocol can be subdivided into segments in which only thermal contacts to the hot bath occur and segments in which only thermal contacts with the cold bath occur. If these segments are not isothermal processes, the extracted work will be decreased and the heat will be increased. At the same time, the matching conditions (9.28) will not be matched, with a similar effect. Therefore the efficiency will be reduced in comparison to the Carnot efficiency.

Given the above result, it is a natural question to ask whether the Carnot-bound can also be achieved with only limited control over the Hamiltonians. For example, we might consider as working medium an interacting spin system where the experimenter can only control external magnetic fields. It is then interesting to know whether Carnot efficiency can in principle be achieved, and whether this depends on further properties of the model. We will find in section 9.6 that the achievable indeed strongly depends on the sign and strength of the interaction in such an example.

Before we come to that example, let us see from a more abstract point of view what could go wrong. From the above discussion, we see that we need to be able to implement isothermals and meet the matching conditions (9.28) to achieve Carnot efficiency. Since we always assume that the set of allowed Hamiltonians $\mathcal{H}$ is well-behaved, in particular any

two Hamiltonians can be connected by a smooth curve, isothermal processes are always possible. The matching condition, however, cannot always be met, in the same way as for a single heat bath in Lemma 9.2. The following theorem then gives a similar bound as Lemma 9.2 for the case of efficiencies of thermal machines.

Theorem 9.5 (Efficiency under control limitations). *Let $\mathcal{P}$ be any cyclic protocol between two heat baths at inverse temperatures $\beta_c > \beta_h > 0$ employing Hamiltonians from the set $\mathcal{H}$ and fully thermalizing maps $\mathcal{T}_t$ as thermal contacts. Then there exist four Hamiltonians $H_1, \ldots, H_4 \in \mathcal{H}$ such that the efficiency is bounded as*

$$\eta(\mathcal{P}) \leq 1 - \frac{\beta_h}{\beta_c} \frac{\Delta S + \min_U D(U \omega_{\beta_h}(H_2) U^\dagger \| \omega_{\beta_c}(H_3))}{\Delta S - \min_V D(V \omega_{\beta_c}(H_4) V^\dagger \| \omega_{\beta_h}(H_1))}, \qquad (9.32)$$

where U, V are unitaries that can be generated by time-dependent Hamiltonians in $\mathcal{H}$, $\Delta S = S(\omega_{\beta_h}(H_2)) - S(\omega_{\beta_c}(H_4))$ and S the von Neumann entropy. Furthermore, for any choice of $H_1, \ldots, H_4 \in \mathcal{H}$ there exists a protocol that saturates the bound.

The formal proof of the theorem is given in section 14.6.6. It roughly works in the same way as theorem 9.4, but taking care of the dissipation terms. The theorem gives a tight, but very abstract bound. In particular, at first sight the optimization over all the unitaries that can be implemented using time-dependent Hamiltonians form $\mathcal{H}$ poses a serious challenge to evaluating it. In the next section, we will see that in generic situations this optimization can in fact be eliminated.

9.5   Local control: Implications from the theory of quantum control

The tight bound in theorem 9.5 involves an optimization over all unitary transformations that can be generated from time-dependent Hamiltonians in $\mathcal{H}$. At first sight this seems like a completely unmanageable object. However, it turns out that quite the opposite is true. The question of what kind of unitary transformations can be implemented (and how) given a fixed set of Hamiltonians is the subject of the field of *quantum control*. One of the most important insights in that field was that essentially any non-trivial set of Hamiltonians allows to generate to arbitrary precision *all possible* unitaries [240].

In particular, this is true in a very strong sense in interacting many-body systems. Suppose for example that we have a finite Heisenberg spin-chain and can control only the magnetic field on the last qubit. Then by ingenuineously tuning the magnetic field over time, it is possible to implement any unitary transformation *on the whole chain* [241]. The caveat is of course, that the protocol of how to tune the magnetic field is in general extremely difficult to find and implement.

But as long as we do not care about such "practical problems", as we do when we consider optimal thermodynamic efficiencies, it turns out that as long as the spins in a many-body system interact with a generic nearest-neighbor interaction, any unitary transformation can in principle be implemented by tuning the position-dependent magnetic field in the right way over time.

Therefore, in many applications, we can replace the optimization over unitaries in theorem 9.5 with an optimization over all unitaries from the unitary group. In this case, we can in fact find the optimal unitaries. To do this, consider one of the correction terms in the Theorem. We can re-write it in terms of non-equilibrium free energies as

$$\begin{aligned}
D(U \omega_{\beta_h}(H_2) U^\dagger \| \omega_{\beta_c}(H_3)) &= \Delta F_{\beta_c}(U \omega_{\beta_h}(H_2) U^\dagger, H_3) \\
&= \mathrm{Tr}\left(H_3 \left(U \omega_{\beta_h}(H_2) U^\dagger - \omega_{\beta_c}(H_3) \right) \right) \\
&\quad - T_c (S(\omega_{\beta_h}(H_2)) - S(\omega_{\beta_c}(H_3))).
\end{aligned}$$

We see that the unitary U only affects the energy-expectation value in the first term. We thus have to optimize U in such a way that the energy-expectation in the Hamiltonian H_3 is minimized. But this is easy. We simply diagonalize $\omega_{\beta_h}(H_2)$ and order its eigenvalues in the energy-eigenbasis of H_3 in such a way that they are non-increasing with increasing energy. This is the procedure to transform the state into a *passive state* [236, 237].

To state the resulting bound, let us define the quantity

$$D^{\downarrow}(\rho\|\sigma) := D(p^{\downarrow}\|q^{\downarrow}) = \sum_i p_i(\log(p_i) - \log(q_i)), \qquad (9.33)$$

where $p^{\downarrow}$ and $q^{\downarrow}$ are the vectors of eigenvalues of ρ and σ, respectively, ordered in non-increasing fashion: $p_i^{\downarrow} \geq p_{i+1}^{\downarrow}$ and similarly for $q^{\downarrow}$. The above calculation then shows that

$$D(U\omega_{\beta_h}(H_2)U^{\dagger}\|\omega_{\beta_c}(H_3)) \leq D^{\downarrow}(\omega_{\beta_h}(H_2)\|\omega_{\beta_c}(H_3)) \qquad (9.34)$$

and that equality is achieved if we optimize over all unitaries U. The bound on the efficiency then simplifies to

$$\eta(\mathcal{P}) \leq 1 - \frac{\beta_h}{\beta_c} \frac{\Delta S + D^{\downarrow}(\omega_{\beta_h}(H_2)\|\omega_{\beta_c}(H_3))}{\Delta S - D^{\downarrow}(\omega_{\beta_c}(H_4)\|\omega_{\beta_h}(H_1))}. \qquad (9.35)$$

To summarize, we have seen that if the working system is an interacting spin system with a generic interaction all unitaries can in principle be implemented and the bound simplifies a lot. Nevertheless, the bound is in principle very difficult to evaluate. I will now discuss one example in the next section, where indeed it can be evaluated and which already shows a surprisingly rich behaviour.

9.6   *Illustration: The classical one-dimensional Ising model*

I will now illustrate the implications of the general bound in a concrete example. To simplify the problem, this will be a *classical* example, i.e., one where all Hamiltonians in the set $\mathcal{H}_t$ commute – namely, the case of the one-dimensional classical Ising model with an external field. Furthermore, I will assume that we can only control the external field in a global, translational invariant way. This is indeed a very relevant control restriction, since usually the external magnetic field cannot be controlled with single-spin resolution.

If we consider a system of N spins, the set of Hamiltonians is then given by all Hamiltonians of the form

$$H_t = -h(t) \sum_{j=1}^N \sigma_j^z - J \sum_{j=1}^N \sigma_j^z \sigma_{j+1}^z, \qquad (9.36)$$

where we assume periodic boundary conditions. Since all the Hamiltonians H_t commute, the unitaries that can be generated from these Hamiltonians are all trivial.

Note that there are three different energy scales in this problem: the two temperatures from the heat baths and the interaction strength J. I will now discuss in detail how the efficiency at maximum work density behaves in this model as a function of the interaction strength J, once we have fixed the temperatures. The qualitative results are independent of the temperatures of the heat bath and are summarized in Figure 9.2, where we have taken the thermodynamic limit $N \to \infty$ and plot the efficiency at maximum work per particle. There are essentially four important features of the resulting efficiency.

1. Carnot efficiency is reached for a vanishing interaction. This has to be the case, since in the non-interacting case, the field strength simply control the effective temperature and the matching condition (9.28) can always be fulfilled.

2. In the strongly *ferromagnetic* regime, the efficiency at maximum work density as well as the maximum work density vanishes.

3. In the strongly *anti-ferromagnetic* regime, Carnot efficiency can be achieved to arbitrary accuracy with finite work per particle.

4. There is a sudden change of behaviour at the critical interaction strength J^*, where the optimal protocol changes in a non-analytic way.
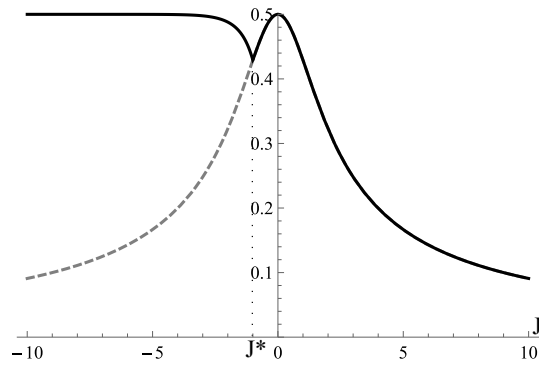
Figure 9.2: Efficiency at maximum work density in the Ising model in the thermodynamic limit when only the global magnetic field can be controlled ($T_c/T_h = 1/2$). There are three important features to observe: i) In the limit of arbitrarily strong, ferromagnetic interactions $J \to \infty$, the efficiency vanishes. ii) In the limit of arbitrarily strong, anti-ferromagnetic interactions $J \to -\infty$, the Carnot efficiency 0.5 can be achieved arbitrarily well. iii) There is a sudden change of behaviour at $J = J^*$. For $0 > J > J^*$ the efficiency monotonically decreases with increasing interaction strengths, while for $J < J^*$ the efficiency increases with increasing interaction strength. This can be attributed to a sudden change of the optimal protocol. The detailed explanation of points i) – iii) is the content of the rest of this chapter. (Figure from Ref. [6].)

In the rest of this chapter, I will provide analytical arguments for all of the above be-haviour. While I will not be working on the level of mathematical physics, all of the arguments can be made completely rigorous.

As we will see, a particularly subtle behaviour is present in the ferromagnetic case. I will show that for any finite system it is in fact in principle possible to achieve Carnot efficiency to arbitrary accuracy (although with vanishing work density), but it is impossible to do so in the thermodynamic limit. The reasons is that to achieve Carnot efficiency, it is necessary to control the external magnetic field to a precision which scales like $1/N$ as the system size is increased. Any finite error in the magnetic field yields vanishing efficiency in the thermodynamic limit.

Before discussing the ferromagnetic and anti-ferromagnetic cases, let us make some general observations. Since we are interested in optimal efficiencies, we will in the follow-ing only consider Carnot-like protocols, which saturate the bound in theorem 9.5. For any fixed temperatures and interaction strength J, a protocol is then defined by four choices of magnetic field strengths $h_1, \ldots, h_4$. We therefore have to optimize over these four values of the magnetic field strength to extract the maximum work density per cycle.

The proof of theorem 9.5 shows that the total extracted work per cycle is given by

$$W(\mathcal{P}) = \Delta T \Delta S - T_h D(\omega_{\beta_c}(H_4) \| \omega_{\beta_h}(H_1)) - T_c D(\omega_{\beta_h}(H_2) \| \omega_{\beta_c}(H_3)), \quad (9.37)$$

with $\Delta T = T_h - T_c$. To maximize the extracted work we therefore have to maximize the difference in entropy ΔS while we minimize the dissipation terms. In particular, we will be interested in the limit $|J| \to \infty$ for fixed temperatures of the heat baths. In this limit, the entropy of a thermal state is given by ground state entropy of the Hamiltonian involving the external fields. Depending on whether the external fields also diverge with $|J|$, they can be treated perturbatively or not. This will be important later.

9.6.1   *Ferromagnetic regime: A tale of instability*

Let us begin with examining the ferromagnetic regime. We are interested in the limit of very strong interactions, $J \to \infty$. As noted above, to obtain a finite work density in this limit, the density in the change of entropy $\Delta S/N$ needs to remain finite and larger than the dissipation terms. It is easy to understand that the optimal protocol (at least for large J) consists of choosing $J \ll h_1 = h_4 \to \infty$ and $h_2 = h_3 = 0$. This is due to the fact that the thermal entropy in the ferromagnetic case is maximized by a vanishing field and minimized by an arbitrarily strong field, which essentially projects the thermal state into one of the two ground states (all spins pointing up or all spins pointing down).

If we choose this protocol we have for any system size

$$\omega_{\beta_{c/h}}(H_1) \approx \omega_{\beta_{c/h}}(H_4) \approx |\downarrow,\dots,\downarrow\rangle\langle\downarrow,\dots,\downarrow|, \tag{9.38}$$

$$\lim_{J\to\infty}\omega_{\beta_{c/h}}(H_2) = \lim_{J\to\infty}\omega_{\beta_{c/h}}(H_3) = \frac{1}{2}\left(|\downarrow,\dots,\downarrow\rangle\langle\downarrow,\dots,\downarrow| + |\uparrow,\dots,\uparrow\rangle\langle\uparrow,\dots,\uparrow|\right). \tag{9.39}$$

As a consequence, we have $\lim_J \Delta S = \log(2)$. Therefore the total work-*density* vanishes (the above relations also show that the density of the correction terms vanish). This can also be verified by a direct calculation using the exact solution of the Ising model in the thermodynamic limit (see section 9.6.3).

While the above calculation shows that the work-density vanishes, the total work of the machine remains finite. I will now show, however, that also the efficiency of the machine vanishes in the limit of large systems and arbitrary strong coupling. This will be done using a calculation on a *finite system*.

For this discussion, it will be useful to not consider the actual optimal protocol above, but to allow for some arbitrary small error $\epsilon > 0$ in the field strengths $h_2 = h_3$, i.e., to allow them to deviate from 0 by a small amount. This would be the case in any real experiment. For simplicity, however, we assume that they deviate from zero by the same amount. They are therefore always equal. It is also useful to keep in mind that the temperature difference ΔT is fixed. Therefore, the ratio $\Delta T/J \to 0$ and the parameter ΔT can be seen as a small parameter when compared to J. Since $h_1 = h_4$ is larger than any other scale in the problem (by assumption) in the optimal protocol, (9.38) implies that the correction term involving the Hamiltonians H_1 and H_4 vanishes. Similarly, the entropy in the thermal state of H_4 vanishes.

Then the efficiency fulfills

$$\lim_{J\to\infty}\eta(J) = \lim_{J\to\infty}\left(1 - \frac{T_c}{T_h}\frac{S_h^{(2)} + D(\omega_{\beta_h}(H_2)\|\omega_{\beta_c}(H_2))}{S_h^{(2)}}\right), \tag{9.40}$$

with $S_{h/c}^{(j)} := S(\omega_{\beta_{h/c}}(H_j))$. To prove that the efficiency vanishes, we have to prove that the second term in the parenthesis vanishes. To see this, first write the relative entropy in terms of non-equilibrium free energies and cancel entropy terms to get

$$\frac{T_c}{T_h}\frac{S_h^{(2)} + D(\omega_{\beta_h}(H_2)\|\omega_{\beta_c}(H_2))}{S_h^{(2)}} = \frac{T_c}{T_h}\frac{S_h^{(2)} + \frac{1}{T_c}\left(E_h^{(2)} - T_c S_h^{(2)} - F_c^{(2)}\right)}{S_h^{(2)}}$$

$$= \frac{1}{T_h}\frac{E_h^{(2)} - T_c S_h^{(2)} - F_c^{(2)}}{S_h^{(2)}} = 1 + \frac{1}{T_h}\frac{F_h^{(2)} - F_c^{(2)}}{S_h^{(2)}}.$$

Here, $E_{h/c}^{(j)}$ denotes the internal energy of Hamiltonian H_j at inverse temperature $\beta_{h/c}$ and $F_{h/c}^{(j)}$ denotes the corresponding *equilibrium* free energy. It is useful to write the free energies as

$$F_{h/c}^{(2)} = Jf(T_{h/c}/J), \quad f(T) := -T\log Z_T(1,\epsilon/J), \tag{9.41}$$

where $Z_T(J,h)$ is the partition function of the Ising model at temperature T, interaction strength J and field strength h. In the limit $J \to \infty$, it is then useful to expand the free energy $F_{h/c}^{(2)}$ in terms of the small parameter $\Delta T/J$ around T_c/J. This yields

$$F_h^{(2)} - F_c^{(2)} = J\left(\left.\frac{\partial f(x)}{\partial x}\right|_{x=T_c/J}\frac{\Delta T}{J} + O(1/J^2)\right) \tag{9.42}$$
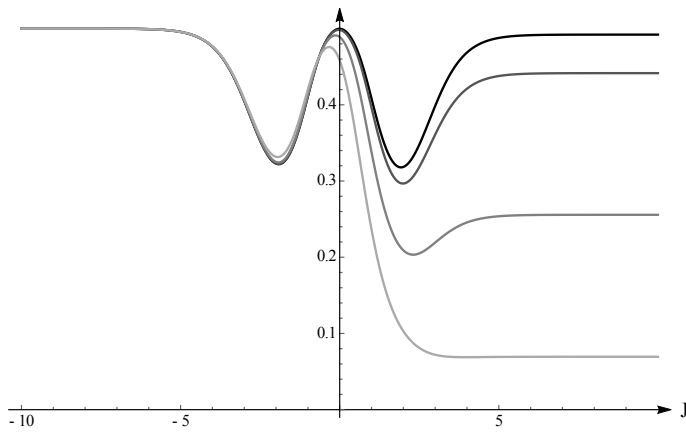
$$= -\Delta T S_c^{(2)} + O(1/J). \tag{9.43}$$

Plugging this back into the efficiency, we obtain

$$\lim_{J \to \infty} \eta(J) = \lim_{J \to \infty} \left(\frac{\Delta T}{T_c} \frac{S_c^{(2)}}{S_h^{(2)}} + O(1/J) \right). \tag{9.44}$$

We therefore have to understand how the ratio of entropies $S_c^{(2)}/S_h^{(2)}$ behaves as J diverges. In section 14.6.7, we calculate this limit for arbitrary ferromagnetic Hamiltonians with a gap, which include the Ising model as a special case. The result is

$$\lim_{J \to \infty} \frac{S_c^{(2)}}{S_h^{(2)}} \begin{cases} \leq O(e^{-(\beta_c - \beta_h)\epsilon N}), & \text{if } \epsilon \neq 0 \\ = 1, & \text{if } \epsilon = 0. \end{cases} \tag{9.45}$$

We thus see that if ϵ vanishes identically, we get Carnot efficiency. However, for any $\epsilon > 0$, the efficiency at $J \to \infty$ goes to zero exponentially as $N \to \infty$ (also see Figure 9.3). This behaviour is also what is seen if one calculates the efficiency directly in the thermodynamic limit, as was done in Figure 9.2.

9.6.2 Anti-ferromagnetic regime

Let us now discuss the anti-ferromagnetic regime. As explained above, finite work-density is only possible if finite entropy density can be reached. Suppose that the external field is of the form $h = k(J)J$. Then in the limit $|J| \gg 0$, the thermal state at any finite temperature approximates the ground state of the effective Hamiltonian

$$\tilde{H}(J) = -k(J) \sum_{j=1}^{N} \sigma_j^z - \text{sign}(J) \sum_{j=1}^{N} \sigma_j^z \sigma_{j+1}^z. \tag{9.46}$$

It is therefore necessary to understand in some detail the ground state degeneracy of the model as a function of the external magnetic fields. In the case of a ferromagnetic interaction this was very easy, since the degeneracy is of order one for any external field.

Let us now discuss the ground state degeneracy in the anti-ferromagnetic case $J \ll 0$. Clearly, for $k(J)$ extremely large, the ground-state of the effective Hamiltonian $\tilde{H}$ is unique. Similarly, for $k(J) = 0$, the ground-state is unique for an even number of spins and there is a ground-state degeneracy that scales linearly with the number of spins for an odd number of spins. In both cases, the entropy density vanishes since an exponentially large ground-state degeneracy is necessary to obtain a finite entropy density.

This raises the question whether it is possible to actually achieve a finite entropy density. I will now show that this is indeed possible. To see this, let us set $k(J)$ to be constant and equal to the number of nearest neighbors of a spin in the lattice. We are thus slightly generalizing from the one-dimensional case. For example, in a d-dimensional square lattice, we set $k(J) = 2d$. Despite the fact that the interaction is anti-ferromagnetic, one of the ground

state of the effective Hamiltonian $\tilde{H}(J)$ is then given by the state $|\uparrow\rangle^{\otimes N}$, where $|\uparrow\rangle$ is the local spin-state pointing in the direction of the external field. The state has energy

$$-(2d)N + N = -(2d - 1)N. \tag{9.47}$$

Now suppose we flip a single spin in this state. The energy contribution from the external field increases by an amount $2(2d)$. On the other hand, there are $2d$ interaction terms connected to this spin, and each of them decreases its energy by an amount 2. Hence the total energy is unchanged and we have found a new ground state. We cannot flip a neighbour of this spin, since this would increase the energy. However, if we flip a next-nearest neighbour, the energy remains unchanged and we have produced yet another ground-state. We can thus decide for every second spin, whether we want to flip it or not without changing the energy. In fact we could create even more ground-state configurations by sometimes moving an odd number of sites before flipping. This shows that the ground-state degeneracy is *at least* $2^{N/2}$. Hence the entropy at zero temperature of the effective Hamiltonian $\tilde{H}(2d)$ is lower bounded by $N/2 \log(2)$. We thus see that in our one-dimensional example we can obtain a finite entropy density by setting $h = 2J$ for any J.

With this in mind, let us first consider the protocol where we set $h_2 = h_3 = 2J$ and $h_1 = h_4$ arbitrarily large. We will later see that this protocol is optimal for very large $|J|$ but not optimal for small $|J|$.

Then, as $J \to -\infty$, the states $\omega_{\beta_c}(H_3)$ and $\omega_{\beta_h}(H_2)$ both converge to the ground state of $H_2 = H_3$ and their relative entropy density vanishes. Similarly, $\omega_{\beta_h}(H_1)$ and $\omega_{\beta_c}(H_4)$ both approach the pure state with all spins pointing in the direction of the magnetic field and their relative entropy also vanishes. Furthermore, also the entropy $S_c^{(4)}$ converges to zero. We thus see that the total work-density in the protocol fulfills

$$\lim_{J \to -\infty} \frac{W(\mathcal{P})}{N} \geq \Delta T \frac{1}{2} \log(2). \tag{9.48}$$

This shows that finite work density can be achieved in the anti-ferromagnetic regime, in strong contrast to the ferromagnetic case.

The above arguments also show that in the limit of $J \to -\infty$ the efficiency fulfills

$$\lim_{J \to -\infty} \eta(J) = 1 - \lim_{J \to \infty} \frac{T_c}{T_h} \frac{\Delta S}{\Delta S} = 1 - \frac{T_c}{T_h}. \tag{9.49}$$

Therefore the protocol achieves Carnot efficiency with finite work per particle as $|J|$ diverges.

9.6.3   The kink and the optimal protocol

Let us finally also explain how the kink in Figure 9.2 arises. To understand this, we work directly in the thermodynamic limit and make use of the well-known exact solution of the Ising model in this case. The free energy density in the thermodynamic limit takes the form

$$f(\beta, J, h) = -\frac{1}{\beta} \log \left(e^{\beta J} \cosh(\beta h) + (e^{2\beta J} \sinh(\beta h)^2 + e^{-2\beta J})^{1/2} \right).$$

From the above discussions, it is clear that to obtain the protocol we have to maximize the entropy density by varying the magnetic field for a given temperature and interaction strength. The entropy density can be calculated from the free energy density as $s = -\partial_T f$, however giving a complicated expression that is not very insightful. To find the optimal magnetic field, we compute the derivative $\partial_h s$ and find its roots. This derivative is then given by

$$\frac{\partial s(\beta, J, h)}{\partial h} = -\beta^2 \frac{e^{\beta J} (h \cosh(\beta h) + 2J \sinh(\beta h))}{(e^{-2\beta J} + e^{2\beta J} \sinh(\beta h)^2)^{1/2} (1 + e^{4\beta J} \sinh(\beta h)^2)}. \tag{9.50}$$

One way to let this expression vanish is to take $h \to \infty$, which results in a diverging denumerator. However, in this case the entropy vanishes since it corresponds to a thermal state
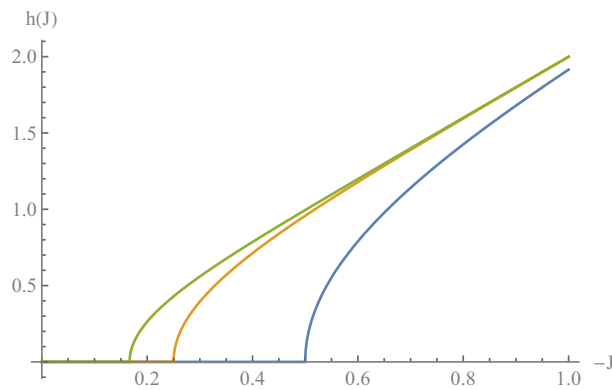
Figure 9.4: The magnetic field $h_2(J)$ that maximizes the entropy as a function of J for inverse temperatures $\beta = 1$ (blue), $\beta = 2$ (orange) and $\beta = 3$ (green). It is clearly visible that at the critical point $1/(2\beta)$ the function is not analytic, similarly to a second order phase transition. For large values of $|J|$, the field approaches $2J$. (Figure from Ref. [6].)

which approximates the state where all spins are pointing in the direction of the magnetic field.

The only other way to make the expression vanish is by a vanishing numerator. We thus look for solutions of the equation

$$h(J)\cosh(\beta h(J)) + 2J\sinh(\beta h(J)) = 0.$$

One solution is given by $h_1(J) = 0$. However, remembering that $J < 0$, we can simplify the equation to

$$h(J) = 2|J|\tanh(\beta h(J)). \tag{9.51}$$

Since $h \mapsto \tanh(\beta h)$ is a concave function for $h > 0$ and anti-symmetric, with the derivative $\beta > 0$ at the origin, a second solution $h_2(J)$ always exists if $2|J|\beta > 1$, or in other words if

$$|J| > \frac{1}{2}k_B T. \tag{9.52}$$

This solution can only exist if $J < 0$, explaining why the kink only appears in the anti-ferromagnetic case [4]. Furthermore, $h_2(J)$ always yields a larger entropy than $h_1(J)$ and is not continuously differentiable. To illustrate this, $h_2(J)$ is plotted in Figure 9.4.

Finally, we can see that $h_2(J)$ indeed approaches $2J$ for very large J since $\tanh(\beta h) \to 1$ as $h \to \infty$: Choosing $h(J) = 2J$ approximately solves equation 9.51 in the regime of large J.

[4] This also supports the optimality of the ferromagnetic protocol discussed in section 9.6.1

9.7   *Conclusions and Outlook*

The laws of phenomenological thermodynamics were originally formulated only for macroscopic bodies. Today we know that they also hold in a well-defined sense for small systems that can be brought in contact with heat baths. From this point of view it is very surprising that the fundamental bounds of thermodynamics, such as the Carnot-bound for the efficiency of a thermal machine, can essentially be saturated by controlling macroscopic systems only on a macroscopic level: It suffices to move the piston of a container with gas, but we never seem to need to control the molecules in the gas individually. This can be largely attributed to the fact that for macroscopic machines, we are also only interested in the macroscopic work-*density* and not the absolute value of the work, which in fact fluctuates widely.

For small machines, however, it is not the work-density but the absolute value of work that is important. We can then expect that the actual experimental control that one has over the system becomes more important. In this chapter, we have derived general corrections to the thermodynamic bound for work-extraction from non-equilibrium systems for situations

of limited experimental control. We have seen how such limited control can indeed lead to the fact that non-equilibrium states become *passive*: An experimenter with limited control could not extract any work from the states, although the free energy might be large.

We then continued with cyclically working thermal machines and de-rived the corresponding corrections to the efficiency of such machines. As a paradigmatic example, we discussed in great detail the situation of an Ising model where an experimenter can only control the external field in a homogeneous way. Surprisingly, we found that the ferromagnetic and anti-ferromagnetic regimes behave completely differently in the limit of very strong interactions and large systems: While the anti-ferromagnetic regime essentially allows to achieve Carnot efficiency with finite work-density even in the thermodynamic limit, the ferromagnetic model becomes essentially useless in the limit of large systems and strong interactions – both the work-density per cycle as well as the efficiency go to zero. All these results could be explained fully analytically.

In this chapter, we have essentially assumed throughout that heat baths are coupled weakly to working systems – in the sense that the thermal equilibrium of systems coupled to the heat bath can be defined by the local Gibbs state. For small systems, this assumption is debatable. It is therefore important to study the case of systems strongly coupled to heat baths. We will do this in the following two chapters 10 and 11, building up on the notion of equilibration in complex quantum systems discussed in chapters 7 and 8.

10 *Thermodynamics in closed quantum systems using Generalized Gibbs Ensembles*

IN CHAPTERS 7 AND 8, we discussed how complex quantum systems equilibrate. An important insight from the study of equilibration was that for most times and many physically relevant observables, the actual time-dependent state of a complex system $\rho(t)$ can be replaced with the *time-averaged* state[1]

$$\omega_{\text{t.a.}}(\rho, H) := \omega(\rho, H) = \sum_k P_k \rho P_k, \qquad (10.1)$$

[1] We introduce the additional subscript for later convenience.

where P_k are the projectors onto the different energy-subspaces with energies E_k. The time-averaged state hence provides an effective description of the state of the system, without containing all the time-dependent dynamical information. Depending on the Hamiltonian and the observables of interest, this effective description can be further simplified, for example to Gibbs states or Generalized Gibbs states (this will be discussed in more detail below).

IN THIS CHAPTER[2] we will connect this insight to thermodynamic operations with the goal to obtain interesting results about thermodynamics if we drop the assumptions of i) weak coupling to heat baths, ii) infinite heat baths, and iii) heat baths that are well described by Gibbs states. To do this, we will introduce a framework for thermodynamic operations similar to that of chapter 9, consisting of repeated quenches with subsequent equilibration steps in a closed quantum many-body system. We will thus explicitly incorporate the "bath" system in our description. In the sections 10.1 – 10.4 we will then study in great detail basic thermodynamic questions in this model, namely entropy production and reversibility as well as the validity of the minimum work principle.

[2] This chapter is based on joint-work with Martí Perarnau-Llobet, Arnau Riera, Rodrigo Gallego and Jens Eisert published in Refs. [3].

In the next chapter we will then focus on the setting where a good effective description is given by Gibbs-states of the system and bath together and use this to study the efficiency and power of thermal machines that are strongly coupled to large heat baths.

10.1 *Generalized Gibbs Ensembles as equilibration models*

I will now start with discussing the framework that we will be using for the rest of this chapter. As discussed before for equilibrating systems an effective description of the state of the system is given by the time-averaged state $\omega(\rho, H)$. This state describes to great precision the measurement statistics of a physically relevant set of observables for most times. This is the definition of equilibration.

A particular feature of the time-averaged state is that it is also the state that maximizes the von Neumann entropy given the expectation value of all *conserved quantities* (CQ), i.e., of all observables that commute with the Hamiltonian [242]. This can be seen as a dynamical emergence of Jaynes' maximum entropy principle [31, 32]. The proof of this statement goes as follows: Let ρ be some state with the same expectation value for all the conserved quantities as some stationary state τ. Then the time-average of ρ has to coincide with τ: $\omega_H(\rho) = \tau$. We can now use the data-processing inequality of the relative entropy to show that the entropy of ρ is smaller or equal to τ (see chapter 2). Since time-averaging leaves the maximally mixed state invariant, we have

$$D(\rho\|\mathbf{1}/d) \geq D(\omega_H(\rho)\|\mathbf{1}/d) = D(\tau\|\mathbf{1}/d), \qquad (10.2)$$

where D is the quantum relative entropy. Since furthermore

$$D(\rho\|\mathbf{1}/d) = \log(d) - S(\rho), \qquad (10.3)$$

we find that τ is the maximum entropy state with the given expectation values for all the conserved quantities.

Despite the fact that the time-averaged state provides an effective description, it is not very useful as an *efficient* description since in any many-body system there are exponentially many conserved quantities in the system size (the projectors onto the energy-levels). Therefore, even for moderately sized systems, one could not even store the classical description of the time-averaged state on a computer.

For *thermalizing* systems, a much simpler effective description holds: We can simply use the Gibbs-state

$$\omega_{\mathrm{Gibbs}}(\rho, H) := \omega_{\beta(\rho)}(H) = \frac{\mathrm{e}^{-\beta(\rho)H}}{Z_{\beta(\rho)}}, \qquad (10.4)$$

where $\beta(\rho)$ is chosen such that $\mathrm{Tr}(\rho H) = \mathrm{Tr}(\omega_{\mathrm{Gibbs}}(\rho, H)H)$. Similarly to the time-average state, which maximizes the entropy given all conserved quantities, the Gibbs-state maximizes the entropy given the expectation value of the Hamiltonian.

For many non-thermalizing systems, we can expect that a behaviour in-between the two is fulfilled: While it is not sufficient to fix the expectation value of the Hamiltonian alone, it is also not necessary to know the expectation value of *all* conserved quantities, but it suffices to know those of a set of m conserved quantities Q_j with m not being exponentially large in the system size. In this case we obtain the *Generalised Gibbs ensemble* (GGE)[3]

$$\omega_{\mathrm{GGE}}(\rho, H, \{Q_j\}) := \frac{\mathrm{e}^{-\sum_j \lambda_j(\rho)Q_j}}{Z_{\lambda_j(\rho)}}, \qquad (10.5)$$

where the Lagrange-multipliers λ_j are chose such that

$$\mathrm{Tr}(\rho Q_j) = \mathrm{Tr}\left(\omega_{\mathrm{GGE}}(\rho, H, \{Q_j\})Q_j\right), \quad j = 0, \dots, m \qquad (10.6)$$

and we have set $Q_0 := H$. Indeed, this description is known to give good predictions after a single quench in *integrable* systems, where the Q_j can be taken as quasi-local operators [249]. In general, it is not obvious how the constants of motion Q_j should be chosen to get the best description. In the rest of this chapter we will simply assume that they are chosen for any Hamiltonian H and study the connection to thermodynamics in full generality.

10.1.1 A simple example

In the next section, I will introduce the precise thermodynamic framework that we will consider. Before coming to that, let us briefly discuss an example of equilibration towards GGEs instead of Gibbs states. For simplicity, consider a chain of N fermionic sites f_j with a hopping Hamiltonian of the form

$$H^{(0)} = \sum_{j=1}^{N} \epsilon_j f_j^\dagger f_j + g \sum_{j=1}^{N-1} \left(f_j^\dagger f_{j+1} + f_{j+1}^\dagger f_j \right). \qquad (10.7)$$

[3] See, for example, Refs. [196, 198, 201, 202, 243–247] for discussions of GGEs in the context of equilibration of closed quantum systems and Refs. [53, 92, 93, 248] for recent discussions of the role of GGEs in (quantum) thermodynamics.
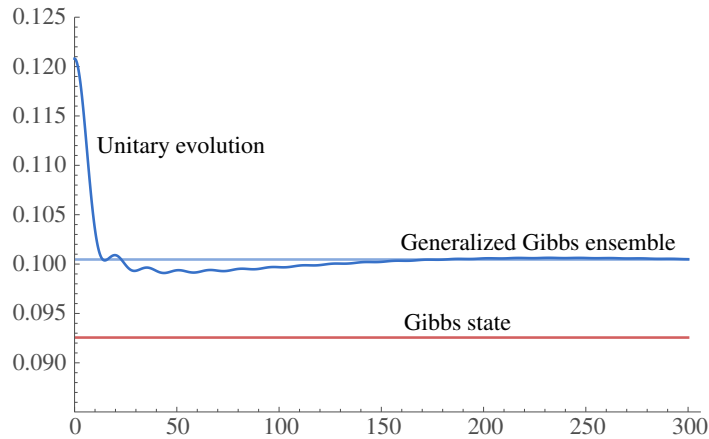
Figure 10.1: Example for equilibration towards a generalized Gibbs ensemble in the free fermionic system of $n = 100$ sites described in the main-text. The initial state is a thermal state at inverse temperature $\beta = 2$ on the whole system corresponding to a Hamiltonian for which $e_j = 1$ and $g = 0.1$. A quench on the first site is then applied, changing ϵ_1 from $\epsilon_1 = 1$ to $\epsilon_1 = 1.15$, and the expectation value $n_1(t)$ is plotted. Clearly the system equilibrates to a Generalized Gibbs ensemble and not to a thermal state. (Figure adapted from Ref. [3].)

Suppose the system is initially in thermal equilibrium and described by a Gibbs-state at inverse temperature β. Then, the potential on the first site is suddenly changed and the system evolves under the new Hamiltonian $H^{(1)} = H^{(0)} + \Delta f_1^\dagger f_1$. We are interested in how the population on the first site, given by

$$n_1(t) := \mathrm{Tr}\left(f_1^\dagger f_1 e^{-iH^{(1)}t} \omega_\beta(H^{(0)}) e^{iH^{(1)}t} \right), \tag{10.8}$$

evolves in time. The result is plotted[4] in Fig. 10.1. As expected, the population $n_1(t)$ relaxes to a steady-state value. Furthermore, this steady-state value is not predicted by the thermal state of $H^{(1)}$ corresponding to the initial state, but by the GGE of $H^{(1)}$ corresponding to the initial state, where the conserved quantities have been chosen as the N normal-modes of $H^{(1)}$.

[4] Since the Hamiltonian is quadratic in the fermionic operators, the time-evolution can simply be computed numerically even for large system size.

10.2   Thermodynamic protocols

We will now use the effective description of equilibrating systems in terms of GGEs to build up a thermodynamic framework similar to that in chapter 9. Again, we will consider a sequence of N Hamiltonians $H^{(j)}$ that connect an initial Hamiltonian $H(0)$ with a final Hamiltonian $H(1)$. We think of these Hamiltonians as taken from a trajectory of Hamiltonians $H(u) : u \in [0, 1] \mapsto H(u)$ such that $H^{(j)} = H(j/N)$. We will assume that all the Hamiltonians are taken from some set $\mathcal{H}$ of Hamiltonians that can be realized. The paradigmatic example is given by the case where we have a many-body system which we partition into two parts, a system S and a bath B. The total Hamiltonian is of the form

$$H_{SB}^{(0)} = H_S^{(0)} + V + H_B, \tag{10.9}$$

and the set of Hamiltonians $\mathcal{H}$ is obtained by varying the local Hamiltonian $H_S^{(0)}$ on the system while the interaction remains fixed. In the next chapter, will also consider the ability to turn on and off the interaction V, but in this section most results will be independent of the specific form of $\mathcal{H}$. This allows us to also study situations where, for example, we can only control external fields uniformly over the system, but no system-bath partition can be made.

A thermodynamic protocol then consists of the specification of the $N + 1$ Hamiltonians and an initial state $\rho(0)$. Just as before, we will assume that the system first evolves under Hamiltonian $H^{(0)}$ for some time t_0 after which a quench to Hamiltonian $H^{(1)}$ is applied. After the quench, the system evolves for a time t_1 under $H^{(1)}$ until the system is quenched

to $H^{(2)}$. This procedure is continued until the system reaches the final Hamiltonian $H^{(N)} = H(1)$ and a consecutive time-evolution for time t_N. Importantly, we will assume that the time-evolution after each quench is long enough so that the system equilibrates. A thermodynamic protocol thus again is given by an arbitrary combination of two operations:

1. A quench from $H^{(j)}$ to $H^{(j+1)}$.

2. Letting the system equilibrate.

Note that again that here we are considering a closed system, so that equilibration has to be understood in the sense of chapter 7.

   Up to now, we have not made use of the effective description of equilibrium states in terms of maximum entropy ensembles. We will now introduce an effective description of a thermodynamic protocol in terms of Generalized Gibbs ensembles. Of course the true time-evolution of the whole system is always unitary. That means that at physical time t_i in the protocol, right before the $(i+1)$-th quench, the true quantum state of the system is given by

$$\rho(t_i) = U_i U_{i-1} \cdot U_1 \rho(0) U_1^\dagger \cdot U_{i-1}^\dagger U_i^\dagger. \tag{10.10}$$

Here, U_i denotes the unitary evolution under Hamiltonian $H^{(i)}$ for a time $t_i - t_{i-1}$. However, for large classes of observables, we can expect that the state $\rho(t_i)$ can be well approximated by some GGE since we assume that the system equilibrates. We could therefore think of introducing an effective description in terms of GGEs by replacing

$$\rho(t_i) \mapsto \omega_{\text{GGE}}^{(i)} := \omega_{\text{GGE}}(\rho(t_i), H^{(i)}, \{Q_j^{(i)}\}). \tag{10.11}$$

In fact this prescription would, by definition, always yield a correct description of all conserved quantities $Q_j^{(i)}$. However, to compute the effective description, we would have to keep track of the actual state of the system $\rho(t_i)$ to compute the required expectation value.

   To avoid this problem, we will make the assumption that the GGE at step i not only describes correctly the conserved quantities at step i, but that it also gives good predictions for the expectation value of the conserved quantities at step $i+1$. That is, we assume

$$\text{Tr}\left(Q_j^{(i+1)}\rho(t_i)\right) \approx \text{Tr}\left(Q_j^{(i+1)}\omega_{\text{GGE}}^{(i)}\right), \quad j = 0, \dots, m. \tag{10.12}$$

This amounts to saying that the conserved quantities belong to the set of physically relevant observables that are well described by the effective description. Further note that we also have

$$\text{Tr}\left(Q_j^{(i+1)}\rho(t_{i+1})\right) \approx \text{Tr}\left(Q_j^{(i+1)}\rho(t_i)\right), \quad j = 0, \dots, m, \tag{10.13}$$

since $\rho(t_{i+1})$ is obtained form evolving $\rho(t_i)$ under the Hamiltonian $H^{(i+1)}$ with associated conserved quantities $Q_j^{(i+1)}$. Thus, by definition of conserved quantities, the two states $\rho(t_{i+1})$ and $\rho(t_i)$ must yield the same expectation values for $Q_j^{(i+1)}$ (remember that $Q_0^{(i+1)} = H^{(i+1)}$). Now the state $\omega_{\text{GGE}}^{(i+1)}$ is obtained form maximizing the von Neumann entropy, subject to the condition that the expectation values of the conserved quantities $Q_j^{(i+1)}$. Combining the above two relations we thus arrive at

$$\omega_{\text{GGE}}^{(i+1)} \approx \omega_{\text{GGE}}\left(\omega_{\text{GGE}}^{(i)}, H^{(i+1)}, \{Q_j^{(i+1)}\}\right). \tag{10.14}$$

Equality in the above relation provides our working assumption for the rest of the chapter.

Assumption 10.1 (Effective description of thermodynamic protocols). *We assume that in an equilibrating system thermodynamic protocols can be effectively described by calculating the GGE after a quench from the GGE prior to the quench,* repeatedly *along the whole protocol:*

$$\omega_{\text{GGE}}^{(i+1)} = \omega_{\text{GGE}}\left(\omega_{\text{GGE}}^{(i)}, H^{(i+1)}, \{Q_j^{(i+1)}\}\right), \quad \forall i = 1, \dots, N. \tag{10.15}$$

By assumption, we therefore have

$$\mathrm{Tr}\left(\omega_{\mathrm{GGE}}^{(i+1)} Q_j^{(i+1)}\right) = \mathrm{Tr}\left(\omega_{\mathrm{GGE}}^{(i)} Q_j^{(i+1)}\right), \quad j = 1, \ldots, m. \tag{10.16}$$

Using this assumption, we then arrive at an effective description in terms of a chain of GGEs:

$$\cdots \to \omega_{\mathrm{GGE}}^{(i)} \to \omega_{\mathrm{GGE}}^{(i+1)} \to \cdots, \tag{10.17}$$

with $\omega_{\mathrm{GGE}}^{(i)}$ denoting the effective description at protocol step i. This procedure, which determines the new effective description from the previous effective description, is standard in most treatments of equilibrium thermodynamics since one simply always uses the Gibbs-equilibration model and never talks about the true microscopic state of the system. Nevertheless, as emphasized above, there is a crucial assumption involved. Whether this assumption is fulfilled in concrete situations depends on the concrete models at hand in the same way as the assumption in usual equilibrium thermodynamics that the Gibbs state provides an accurate description of the situation. In the rest of the chapter, we will simply take this for granted and study the consequences of this assumption.

10.2.1 Work-cost of thermodynamic protocols

A thermodynamic protocol is composed of quenches $H^{(i)} \to H^{(i+1)}$. Each such quench has an associated work-cost given by

$$W(i \to i+1) = \mathrm{Tr}\left(\rho(t_i)\left(H^{(i)} - H^{(i+1)}\right)\right). \tag{10.18}$$

This is the "true" work-cost as computed from the actual, time-dependent quantum state $\rho(t)$. In our effective description, however, the associated work-cost of a single quench is given by

$$W_{\mathrm{GGE}}(i \to i+1) = \mathrm{Tr}\left(\omega_{\mathrm{GGE}}^{(i)}\left(H^{(i)} - H^{(i+1)}\right)\right). \tag{10.19}$$

It may at first glance seem that W_{GGE} would coincide with W by definition, since the Hamiltonians are conserved quantities. Indeed, if the effective description at step i in the protocol would be computed from $\rho(t_i)$ instead of $\omega_{\mathrm{GGE}}^{(i-1)}$, then we would have

$$\mathrm{Tr}(\omega_{\mathrm{GGE}}^{(i)} H^{(i)}) = \mathrm{Tr}(\rho(t_i) H^{(i)}). \tag{10.20}$$

However, this would still not justify computing the expectation value of $H^{(i+1)}$ using the effective description $\omega_{\mathrm{GGE}}^{(i)}$ since there is no guarantee that the two results would coincide. Note however, that this is exactly what our working assumption, expressed in Eq. (10.12), predicts. Consistency of our assumptions therefore implies that the actual work $W(i \to i+1)$ is well approximated by $W_{\mathrm{GGE}}(i \to i+1)$. Therefore, we will henceforth only consider the work in the effective description given by W_{GGE} when we discuss work-extraction in thermodynamic protocols in section 10.4. The total work-cost of a thermodynamic protocol $\mathcal{P}_N$ with N quenches is then given by

$$W_{\mathrm{GGE}}(\mathcal{P}_N) = \sum_i W_{\mathrm{GGE}}(i \to i+1)$$

$$= \mathrm{Tr}\left(\omega_{\mathrm{GGE}}^{(0)} H^{(0)}\right) - \mathrm{Tr}\left(\omega_{\mathrm{GGE}}^{(N-1)} H^{(N)}\right) + \sum_{i=1}^{N-1} \mathrm{Tr}\left(H^{(i)}\left(\omega_{\mathrm{GGE}}^{(i)} - \omega_{\mathrm{GGE}}^{(i-1)}\right)\right). \tag{10.21}$$

We can now use our defining equation (10.16) in the form

$$\mathrm{Tr}\left(H^{(i)} \omega_{\mathrm{GGE}}^{(i)}\right) = \mathrm{Tr}\left(H^{(i)} \omega_{\mathrm{GGE}}^{(i-1)}\right). \tag{10.22}$$

This yields

$$W_{\mathrm{GGE}}(\mathcal{P}_N) = \mathrm{Tr}\left(\omega_{\mathrm{GGE}}^{(0)} H^{(0)}\right) - \mathrm{Tr}\left(\omega_{\mathrm{GGE}}^{(N)} H^{(N)}\right). \tag{10.23}$$

We therefore obtain that the total work is simply given by the change of energy. This is what we should obtain, since we are dealing with a closed system.

10.2.2 Time-dependent Lagrange multipliers and strong-coupling corrections

The second important point to realize is that we use the effective description in terms of
GGEs for *all systems involved*. In particular, in the setting where the total system is parti-
tioned into a system S and a bath B (or, more generally, multiple baths), we use the effective
description for S and B together. Since the total system is finite, and quenches inject or ex-
tract energy from the system, this has as a consequence that the Lagrange multipliers $\lambda_j^{(i)}$
are not constant over the protocol. For example, if we are in a situation where we use the
Gibbs equilibration model, the inverse temperature $\beta^{(i)}$ will generally depend on time. In
particular in the system-bath setting, the state on the system at time t_i is given by

$$\rho_S^{(i)} = \text{Tr}_B \left(\omega_{\beta^{(i)}}(H_{SB}^{(i)}) \right), \tag{10.24}$$

whereas usually one associates the thermal state of the local Hamiltonian $H_S^{(i)}$ to the sys-
tem, at a fixed inverse temperature β. That is, one assumes

$$\rho_S^{(i)} = \omega_\beta(H_S^{(i)}). \tag{10.25}$$

In general the two descriptions (10.24) and (10.25) will differ. However, there are clear
cases in which we expect them to coincide.

First, if the system B is very large and we only operate with quenches on system S,
we can expect that the total work done on the system is negligible compared to the total
energy in S and B together. In this case we can expect that the inverse temperature stays
approximately constant, $\beta^{(i)} \approx \beta$.

Second, in the limit where the interaction V between S and B is negligibly small, we
can expect that the total state of SB becomes uncorrelated between S and B and $\rho_S^{(i)} \approx$
$\omega_{\beta^{(i)}}(H_S^{(i)})$. Combining this weak-coupling limit with the limit of an infinite bath, we then
obtain again (10.25).

We thus see that our framework can be seen as providing corrections to the weak-
coupling, infinite heat bath setting usually considered in thermodynamics. In the next
chapter, we will use this to study in detail strong-coupling corrections to general ther-
modynamic figures of merit, such as work extraction bounds and efficiencies of thermal
machines.

10.3 Entropy production and reversibility

Having introduced the concrete framework, we can now start to analyse thermodynamic
protocols. An important concept in thermodynamics is that of entropy production and re-
versibility. In phenomenological thermodynamics, the second law states that the entropy
of a closed system is non-decreasing over a thermodynamic protocol. If the entropy even
remains constant, then the process is said to be *reversible*. This is justified by the fact that
if one would actually implement the reversed process starting from the final condition of
the original process, then the total system would return to the initial condition of the orig-
inal process. Thus, reversibility in terms of entropy production coincides with operational
reversibility.

In this section, we will discuss in detail the role of entropy production and reversibility
in closed quantum systems when we use the effective description in terms of GGEs at all
times.

As emphasized above, the actual time-evolution of a closed system is always unitary.
Therefore any (quantum) information-theoretic entropy of the actual quantum state of the
system remains constant over time. This does not mean, however, that the von Neumann
entropy of our effective description, in terms of the states $\omega_{\text{GGE}}^{(i)}$, also remains constant.
Indeed, in every step of the protocol, the new state $\omega_{\text{GGE}}^{(i+1)}$ is obtained from the state $\omega_{\text{GGE}}^{(i)}$
by *maximizing* the von Neumann entropy given the expectation values of the conserved
quantities $Q_j^{(i+1)}$ in the state $\omega_{\text{GGE}}^{(i)}$. The state $\omega_{\text{GGE}}^{(i+1)}$ thus necessarily has at least as much

entropy as $\omega_{\text{GGE}}^{(i)}$. We therefore conclude that the entropy, in terms of the effective description, is indeed non-decreasing over a thermodynamic protocol:

$$S\left(\omega_{\text{GGE}}^{(i+1)}\right) \geq S\left(\omega_{\text{GGE}}^{(i)}\right). \tag{10.26}$$

This relation shows that in terms of the effective description, thermodynamic protocols are in general irreversible: Suppose we start with some initial state $\omega_{\text{GGE}}^{(1)}$ and then perform a protocol to arrive at $\omega_{\text{GGE}}^{(N)}$. If afterwards we implement the reversed protocol (i.e., go back along the reversed trajectory of Hamiltonians), we will in general not be able to come back to the initial state $\omega_{\text{GGE}}^{(1)}$.

From phenomenological thermodynamics, we might expect that thermodynamic protocols become reversible in a *quasi-static* limit, i.e., when the initial and final Hamiltonian are connected by a trajectory of infinitely many infinitesimal quenches. More precisely, in our context, we will say that a quasi-static process consists of an initial Hamiltonian $H^{(\text{initial})}$, a final Hamiltonian $H^{(\text{final})}$ and a series of $N \to \infty$ Hamiltonians $H^{(i)}$ such that $H^{(0)} = H^{(\text{initial})}$, $H^{(N)} = H^{(\text{final})}$ and $H^{(i)} - H^{(i-1)} = O(1/N)$. If we take $N \to \infty$, we can then describe the process simply by a continuous trajectory of Hamiltonians

$$u \in [0,1] \mapsto H(u), \tag{10.27}$$

with $H^{(i)} := H(i/N)$ and, given some initial condition $\rho(0)$, also a corresponding continuous trajectory of effective descriptions

$$u \in [0,1] \mapsto \omega_{\text{GGE}}(u), \tag{10.28}$$

with $\omega_{\text{GGE}}^{(i)} := \omega_{\text{GGE}}(i/N)$. It is important to keep in mind that a quasi static process as defined above will in principle take an infinite amount of time since it consists of infinitely many steps each of which takes a finite amount of time. This is in complete analogy to phenomenological thermodynamics. This does however not imply that quasi-static cannot be well approximated by a process that takes place in finite time. If we fix any error $\epsilon > 0$ on thermodynamic quantities such as the work extracted in the process and the entropy that is produced, there will be a finite number $N(\epsilon)$ of quenches which reproduces the quasi-static process to the given approximation in finite time. In general $N(\epsilon)$ will diverge as $\epsilon \to 0$.

Nevertheless, it might in fact be possible to perform a quasi-static process in finite time. This would be possible if the equilibration time after every quench would go to zero as $1/N$. Since the change of the Hamiltonian and the corresponding change of the effective description is only of order $1/N$ such a behaviour seems conceivable. We will leave this interesting open question for future work, however.

In the following, we will consider a fixed trajectory of Hamiltonians $H(u)$ and discuss the behaviour of the entropy along the trajectory given by

$$S(u) := -\text{Tr}\left(\omega_{\text{GGE}}(u) \log(\omega_{\text{GGE}}(u))\right). \tag{10.29}$$

In particular, we want to know under which conditions the entropy remains constant. We will discuss the behaviour of the entropy separately for the three cases of i) time-averaged states (GGE including all constants of motion), ii) the Gibbs equilibration model and iii) GGEs with a non-maximal number of conserved quantities. We will find that a quasi-static process is not automatically reversible in all the equilibration models and discuss examples of this behaviour which suggest that this only happens if the effective description is also not a good description of the system. We will also discuss conditions which guarantee that a quasi-static process has vanishing entropy production.

10.3.1   *Entropy production in the time-average model and more general GGEs*

We will now start to discuss entropy-production and reversibility in the time-average model. Let us first discuss a very simple toy-example that shows that there can be entropy production in the quasi-static limit. Consider the following Hamiltonian trajectory, which is

continuous but not differentiable:

$$u \in [-1, 1] \mapsto H(u) = \begin{cases} u\sigma_x, & -1 \le u < 0, \\ u\sigma_z, & 0 \le u \le 1. \end{cases} \tag{10.30}$$

Now suppose the initial state is an eigenstate of σ_x, e.g. $\rho(0) = |+\rangle\langle+|$. Then for $u \le 0$, time-averaging does not alter the change at all and we have $\omega_{\text{t.a.}}(u) = |+\rangle\langle+|$ for all $u \le 0$. For any $u > 0$ the eigenbasis of the Hamiltonian $H(u)$ is constant orthogonal to $|+\rangle$. Taking the time-average of the state $|+\rangle\langle+|$ in this new eigenbasis therefore completely mixes the state and we have $\omega_{\text{t.a.}}(u) = \mathbf{1}/2$ for any $u > 0$. We conclude that the entropy of the effective description changes from $S(-1) = 0$ to $S(1) = \log(2)$. This dissipation can be attributed to the fact that while the eigenvalues of the Hamiltonian can be described by smooth functions, the eigenbasis cannot. It should be clear that *in principle* a similar effect can occur in interacting many-body systems (even though it may be unlikely, e.g., due to the phenomenon of avoided crossings). Thus, the example shows that we cannot expect to prove the absence of entropy production in general.

But the example also hints at the sufficient condition for absent of entropy production, namely the smoothness of the Hamiltonian trajectory. This is indeed the case, as shown by the following lemma.

Lemma 10.2 (Absence of entropy production in time-average model). *Let $u \in [0, 1] \mapsto H(u)$ be a differentiable trajectory of Hamiltonians, in the sense that the eigenvectors of $H(u)$ are continuous and the eigenbasis of $H(u)$ is differentiable. Then there is no entropy-production in the time-average equilibration model: For any initial equilibrium state $\omega_{\text{t.a.}}(0)$ its spectrum is conserved during the process and therefore $S(0) = S(1)$.*

Proof. Let the eigenbasis of $H(u)$ be given by $|E_k(u)\rangle$. Then the eigenvalues $p_k(u + \delta u)$ of the equilibrium state at parameter value $u + \delta u$ can be written as

$$p_k(u + \delta u) = \langle E_k(u + \delta u) | \omega_{\text{t.a.}}(u) | E_k(u + \delta u) \rangle \tag{10.31}$$

$$= \sum_l p_l(u) \, |\langle E_k(u + \delta u)|E_l(u)\rangle|^2, \tag{10.32}$$

because they are simply the diagonal elements of $\omega_{\text{t.a.}}(u)$ in the new eigenbasis $|E_k(u + \delta u)\rangle$. Using differentiability of the eigenbasis, $|E_k(u)\rangle$ we can write them as

$$|E_k(u + \delta u)\rangle = |E_k(u)\rangle + |X_k(u)\rangle \, \delta u + O(\delta u^2). \tag{10.33}$$

Since the bases are ortho-normalized, we have $\text{Re}\,(\langle E_k(u)|X_k(u)\rangle) = 0$. This implies

$$|\langle E_k(u + \delta u)|E_l(u)\rangle|^2 = \delta_{l,k} + O(\delta u^2).$$

Hence we have $p_k(u + \delta u) = p_k(u) + O(\delta u^2)$. Taking the limit $\delta u \to 0$ we see that the eigenvalues $p_k(u)$ remain constant. $\qquad\square$

Lemma 10.2 is essentially a consequence of the adiabatic theorem of quantum mechanics. While in the quantum mechanical context, the qualifier "adiabatic" simply means "very slow", in this context we see that it also indeed acquires the meaning of vanishing entropy-production.

From the discussion of the time-average model we already learn that in the context of GGEs, vanishing entropy production in the quasi-static limit requires smoothness conditions on the thermodynamic protocols.

In the general case of arbitrary GGEs, we should expect that such conditions can be phrased in terms of the Lagrange-multipliers $\lambda_j(u)$ determining the equilibrium states. We will now provide such a characterization. To understand this characterization, first recall that in the case of a discrete thermodynamic protocol, the Lagrange-multipliers are implicitly determined by the equation

$$\text{Tr}\left(\omega_{\text{GGE}}^{(i)} Q_j^{(i)}\right) = \text{Tr}\left(\omega_{\text{GGE}}^{(i-1)} Q_j^{(i)}\right), \quad j = 1, \ldots, m. \tag{10.34}$$

We then take a continuum limit to obtain the functions $\lambda_j(u)$ in the quasi-static limit. The following Lemma shows that the entropy production vanishes as long as the functions $\lambda_j(u)$ are smooth.

Lemma 10.3 (Absence of entropy production in GGEs). *Consider a quasi-static process along a trajectory of Hamiltonians $u \in [0,1] \mapsto H(u)$ (with associated m conserved quantities $Q_j(u)$) giving rise to the equilibrium states $\omega_{\mathrm{GGE}}(u)$. Then the entropy of $\omega_{\mathrm{GGE}}(u)$ in the quasi-static limit if the functions $u \mapsto \lambda_j(u)$ determined by (10.16) are smooth.*

Proof. If the functions $\lambda_j(u)$ are smooth, we can take the continuum limit of equation (10.16) to obtain

$$\mathrm{Tr}\left(\frac{\mathrm{d}\omega_{\mathrm{GGE}}(u)}{\mathrm{d}u} Q_j(u)\right) = 0, \quad j = 1, \ldots, m. \qquad (10.35)$$

But, due to normalization of the states $\omega_{\mathrm{GGE}}(u)$, the derivative of $S(u)$ is given by

$$
\begin{aligned}
\frac{\mathrm{d}S(u)}{\mathrm{d}u} &= -\mathrm{Tr}\left(\frac{\mathrm{d}\omega_{\mathrm{GGE}}(u)}{\mathrm{d}u} \log(\omega_{\mathrm{GGE}}(u))\right) \\
&= \sum_{j=1}^{m} \lambda_j(u) \mathrm{Tr}\left(\frac{\mathrm{d}\omega_{\mathrm{GGE}}(u)}{\mathrm{d}u} Q_j(u)\right) = 0.
\end{aligned} \qquad (10.36)
$$

Thus the entropy remains constant. $\square$

To illustrate this Lemma, let us discuss a simply counter-example in which the quasi-static limit does not give rise to smooth Lagrange-multipliers. To do this, we will work in the Gibbs-equilibration model, which is a special case of GGEs and consider again a two-level system with Hamiltonian trajectory

$$u \in [0,1] \mapsto H(u) = (1-u)\,|1\rangle\langle 1|. \qquad (10.37)$$

Let the initial state be given by $\omega_{\beta(0)}(H(0))$ with $\beta(0) > 0$. We will now see that this example has the following properties:

1. The inverse temperature $\beta(u)$ diverges as $u \to 1$, therefore it is not smooth on the whole interval $[0,1]$.

2. The entropy does not remain constant.

3. The Gibbs model does not provide a good description of the system.

First note that all Hamiltonians $H(u)$ have the same eigenbasis. Therefore the actual quantum state of the system remains constant and equal to the initial state $\omega_{\beta(0)}(H(0))$. This in turn implies that the effective description in terms of a Gibbs-state requires that the effective inverse temperature and the Hamiltonian $H(u)$ fulfill

$$\beta(u)H(u) = \beta(0)H(0). \qquad (10.38)$$

Since $H(u) = H(0)(1-u)$ we then deduce $\beta(u) = \beta(0)/(1-u)$, which proves the first claim. To see that the entropy increases, note that the final Hamiltonian is fully degenerate. Hence any Gibbs-state of the final Hamiltonian is equal to the maximally mixed state and has entropy $\log(2)$, which shows the second claim. Note that the entropy production only happens if we follow the trajectory all the way to $u = 1$, which is the only point at which $\beta(u)$ is not smooth (since it is ill-defined). The reason for this behaviour is that at $u = 1$, the Hamiltonian is degenerate and cannot "resolve" the two different eigen-levels of the actual quantum state of the system. This also explains the last claim: Any additional, independent observable that commutes with $H(1)$ would distinguish the two different eigen-levels of the quantum state and would lead to smooth Lagrange-multipliers and vanishing entropy-production.

Summarizing, we find that quasi-static processes result in a constant entropy if and only if the process is sufficiently smooth and all Lagrange-multipliers are well-behaved. Let us now turn to the Gibbs-equilibration model.

10.3.2   Entropy-production in the Gibbs equilibration model

We have already seen in the toy-example at the end of the last section that also in the Gibbs-equilibration model entropy-production can happen in the quasi-static limit. In the particular example, the reason was easy to see: The final Hamiltonian $H(1)$ only admitted a single Gibbs-state with entropy $\log(2)$. Therefore, any initial state with different entropy necessarily had to lead to entropy-production. In this section, we will derive results that show that this is essentially the only reason that entropy production can happen in the Gibbs-equilibration model. The core result is the following Lemma.

Lemma 10.4 (General condition for absence of entropy-production)**.** *Consider a quasi-static process along a smooth trajectory of Hamiltonians $H(u)$ and an initial equilibrium state $\omega_{\beta(0)}(H(0))$ with $\beta(0) \neq 0$. Suppose there exists any function $u \mapsto f(u) \neq 0$ with $f(0) = \beta(0)$ such that*

$$S\left(\omega_{f(u)}(H(u))\right) = S(0). \tag{10.39}$$

Then the quasi-static process along $H(u)$ has no entropy-production and $\beta(u) = f(u)$.

The Lemma states that to check whether a given quasi-static process has entropy production, we do not need to actually solve the differential equations that define the quasi-static process. It is enough to check whether it could be possible to keep the entropy constant by any smooth choice of inverse temperature along the Hamiltonian trajectory. Also note that the condition $f(u) \neq 0$ is necessary: If $f(u) = 0$ for any u, the entropy at this point is given by $\log(d)$, where d is the dimension of the Hilbert-space. Then the only way to keep the entropy constant would be to have $\beta(u) = 0$ for all u, which is impossible since by assumption $\beta(0) \neq 0$.

Proof of Lemma 10.4.   In the following proof, we write $\omega_f(u)$ as a shorthand for the effective description $\omega_{f(u)}(H(u))$. Then, by the assumption on f, we have

$$\frac{\mathrm{d}S(\omega_f(u))}{\mathrm{d}u} = f(u)\mathrm{Tr}\left(\frac{\mathrm{d}\omega_f(u)}{\mathrm{d}u}H(u)\right) = 0. \tag{10.40}$$

But since $f(u) \neq 0$, this implies that $\omega_f(u)$ is a solution to (10.16) in the continuum limit. But these solutions are unique and therefore $\omega_f(u) = \omega_{\mathrm{Gibbs}}(u)$ is the actual solution of quasi-static process. Hence $\beta(u) = f(u)$. The uniqueness of the solution can be easily seen by nothing that for a fixed sign of the inverse temperature, the entropy is in one-to-one correspondence with β.   □

In the above Lemma, the Hamiltonian trajectory is fixed. In general, one might only know the initial and final Hamiltonian. We can use the Lemma to derive general conditions for Hamiltonian trajectories that ensure that the entropy-production vanishes. Let us discuss two examples.

Example 10.5 (Entropy-production and ground state degeneracy)**.** *Let the ground-state degeneracy of an Hamiltonian H be given by $d_g(H)$ and consider initial and final Hamiltonians such that $d_g(H(0)) \geq d_g(H(1))$ and an initial state $\omega_{\beta(0)}(H(0))$ with $\beta(0) > 0$. Then any smooth quasi-static process that fulfills*

$$d_g(H(0)) \geq d_g(H(u)) \geq d_g(H(1)), \quad \forall u \in [0,1], \tag{10.41}$$

keeps the entropy constant.

Proof.   The entropy of Gibbs-states of the Hamiltonian $H(u)$ can take any value in the interval $[\log(d_g(H(u))), \log(d)]$. Therefore, the condition ensures that the entropy can be made constant by a suitable choice of inverse temperature $f(u)$. Since the Hamiltonian trajectory is smooth, $f(u)$ can be chosen smoothly. Then Lemma 10.4 proves the claim.   □

In fact, we can leverage this example to show that it is essentially only necessary to be able to find a final inverse temperature that matches the initial entropy. This is shown in the next example.

Example 10.6 (No entropy-production in Gibbs ensembles). *Consider initial and final Hamiltonians $H(0)$, $H(1)$, respectively, and an initial state $\omega_{\beta(0)}(H(0))$ with finite $\beta(0) > 0$. Suppose there exists any finite $\beta^* > 0$ such that*

$$S(0) = S\left(\omega_{\beta^*}(H(1))\right). \tag{10.42}$$

Then any *smooth trajectory with $d_g(H(u)) = 1$ for all u in the open interval $(0, 1)$ connecting the initial and final Hamiltonian results in vanishing entropy-production.*

Proof. By the condition on the Hamiltonians, we can find some smooth function $f(u)$ that matches the entropy for all $u < 1$. But since by assumption β^* exists, we know that $f(1) = \beta^*$. Hence $f(u)$ exists for all $u \in [0, 1]$ and the claim follows from Lemma 10.4. $\qquad\square$

The last two examples show that in the case of the Gibbs equilibration model, smoothness of the Hamiltonian trajectory is essentially sufficient to ensure the absence of entropy-production in the quasi-static limit. Essentially the only condition is that the Hamiltonians along the trajectory allow for entropies that match the initial entropy. If this condition is fulfilled, the entropy does not increase in a quasi-static process.

10.3.3   Reversibility of quasi-static processes

In the last two sections we have studied in great detail when we can and cannot expect that a quasi-static process features entropy-production. In essence we found that there is no entropy-production as long as all involved quantities become sufficiently smooth in the quasi-static limit. Let us now discuss how this connects to *reversibility*.

The essential equation that governs thermodynamic protocols in terms of GGEs is equation (10.16). The continuum limit of this equation takes the form

$$\mathrm{Tr}\left(\frac{d\omega_{\mathrm{GGE}}(u)}{du}Q_j(u)\right) = 0. \tag{10.43}$$

Given a trajectory $H(u)$ and an initial state $\omega_{\mathrm{GGE}}(0)$, it determines the solution $\omega_{\mathrm{GGE}}(u)$. This equation is invariant under "time-reversal" of the "parameter-time" u: Consider the reversed process given by $H(\tilde{u})$ and initial state $\omega_{\mathrm{GGE}}(1)$ with $\tilde{u} \in [0, 1] \mapsto 1 - u$. If a process fulfills (10.43), then formally the process $\omega_{\mathrm{GGE}}(\tilde{u})$ fulfills

$$\mathrm{Tr}\left(\frac{d\omega_{\mathrm{GGE}}(\tilde{u})}{d\tilde{u}}Q_j(\tilde{u})\right) = 0. \tag{10.44}$$

Therefore, the final state of the reversed-process is $\omega_{\mathrm{GGE}}(0)$, the initial state of the original process. We therefore conclude that the process is indeed reversible int he operational sense: If after we have implemented the forward process $H(u)$ quasi-statically we implement the backward process $H(\tilde{u})$ quasi-statically, then we get back to the initial state. As emphasized before, this reasoning holds provided that the Lagrange-multipliers $\lambda_j(u)$ are smooth functions on $[0, 1]$, which also ensures that no entropy is produced.

It is also important to remember, that we always operate on the level of effective descriptions. Since thermodynamic protocols include letting the system equilibrate for an arbitrary time, the actual quantum state at the end of the thermodynamic protocol is not well defined unless one specifies exactly the equilibration times. Therefore, we cannot claim (or, in fact know) whether the actual quantum state comes back to its initial state if we reverse the process unless we precisely match the timing in all of the equilibration steps.

Finally, if we think of a system-bath setting, our results connect reversibility and entropy production in the same sense as in phenomenological thermodynamics: No entropy is produced on system and bath together in a reversible protocol and if no entropy is produced, then the protocol is reversible. Unlike in phenomenological thermodynamics, however, the

entropy that we consider cannot in general be attributed to either the system or the bath since the entropy is not additive, but sub-additive (cf. chapter 5):

$$S(\omega_{\text{GGE}}) = S(\text{Tr}_B(\omega_{\text{GGE}})) + S(\text{Tr}_S(\omega_{\text{GGE}})) - I(S : B), \qquad (10.45)$$

where $I(S : B) \geq 0$ denotes the *mutual information* between S and B and measures the correlations between S and B. In our set-up it is perfectly compatible that i) the entropy of S increases, ii) the entropy of B increases, but at the same time iii) the total entropy remains constant. This can happen due to the built-up of correlations between SB (which increases $I(S : B)$) while increasing the local entropies. Indeed, such behaviour is well-known to lead to thermodynamic entropy production of the system and is a cause of irreversibility if one later on cannot control these correlations anymore (also see discussions in chapter 5 and Refs. [89]).

10.4   The minimum work principle

We have seen in the last sections that our framework reproduces, except for pathological cases, the close links between quasi-static processes, entropy-production and reversibility known from phenomenological thermodynamics. In phenomenological thermodynamics, reversibility is also closely linked to work-extraction through the *minimum work principle*. We will now discuss the minimum work principle in the context of Generalized Gibbs ensembles. To do that, we will use the following formulation of the minimum work principle [250]:

> Given an initial equilibrium state and a path of Hamiltonians, the work performed on the system is minimal for the slowest realisation of the process.

In our context, this means the following. Given an initial equilibrium state a trajectory of Hamiltonians $H(u)$ and an initial equilibrium state $\omega_{\text{GGE}}(0)$, consider the discretization of the trajectory $H(u)$ into N steps given by $H^{(i)} = H(i/N)$ as thermodynamic protocol $\mathcal{P}_N$. This determines the states $\omega_{\text{GGE}}^{(i)}$ through our usual procedures. Note that these states are *not* the same as the states obtained from discretizing the trajectory of states $\omega_{\text{GGE}}(u)$ in the quasi-static limit. The minimum work principle then says that the minimum work cost over all the protocols $\mathcal{P}_N$ is given in the quasi-static limit $N \to \infty$.

While this formulation of the minimum work principle fits well into our framework, it is different from the one usually encountered in text-books on thermodynamics [22, 251]. There, the minimum work principle states that among all processes between two *fixed equilibrium states* on some system, reversible ones minimize the work-cost (equivalently, maximize the extracted work). The argument for this statement is fairly simple and it is useful to have in mind for the following considerations. Let me therefore briefly recapitulate how it works. Consider a system that can be connected to a heat bath at temperature T and a work-storage device. First assume that the system undergoes some process from state A to state B, extracting work $W(A \to B)$ and absorbing heat $Q(A \to B)$. We can now come back to the initial state A by a reversible process[5], extracting work $W(B \to A)$ and absorbing heat $Q(B \to A)$. The total process is then cyclic. Hence, if $W(A \to B) + W(B \to A) \geq 0$, we could cyclically extract work from a single heat bath, which is forbidden by the second law. More formally, we have $Q(B \to A) = T\Delta S(B \to A)$ by reversibility and $Q(A \to B) \leq -T\Delta S(B \to A)$ from Clausius' inequality ($\Delta S(B \to A)$ denotes the change of entropy). From the first law we then get

$$W(A \to B) + W(B \to A) = Q(A \to B) + Q(B \to A) \leq 0.$$

Indeed, in phenomenological thermodynamics once we assume the first law, the minimum work principle is equivalent to the second law. So why can't we use the usual formulation of the minimum work principle in our setting? The reason is that in phenomenological thermodynamics, equilibrium states are in one-to-one correspondence with the thermodynamic variables, which in our case are given by the Hamiltonian. In such a case, our formulation and the traditional one are in fact equivalent, since fixing the initial and final Hamiltonian would fix the initial and final state. This is, for example, the case if one would

<hr/>

[5] In phenomenological thermodynamics, it is assumed that this can always be done.

take as equilibration model Gibbs states with a fixed inverse temperature – as is usually done in thermodynamics. In our context, however, in general the final state of a process depends on the specific trajectory between the initial and final Hamiltonian. Therefore, different reversible protocols from $H(0)$ to $H(1)$ can lead to different final states and we cannot connect them to cyclic protocols. Consequently, the above traditional arguments do not hold in our context if we consider Generalized Gibbs ensembles in finite systems as equilibration models. We thus have to study in which sense the minimum work principle remains true by more specific arguments. This is what we will do in the following sections. Indeed we will find that the minimum work principle can fail and we will provide a concrete example where it does.

Before we come to that, however, let us first discuss the minimum work principle in general in the Gibbs equilibration model and the time-average equilibration model.

10.4.1 Minimum work principle in the Gibbs equilibration model

To discuss the minimum work principle in the Gibbs equilibration model, consider a smooth trajectory of Hamiltonians $H(u)$ and the initial equilibrium state $\omega_{\beta(0)}(H(0))$. Now consider the N-step discretization $\mathcal{P}_N$ of the process. The work associated to this protocol is given by (see section 10.2.1)

$$W_{\text{GGE}}(\mathcal{P}_N) = \text{Tr}\left(\omega_{\beta(0)}(H(0))\right) - \text{Tr}\left(\omega_{\beta(N)}(H(1))\right). \tag{10.46}$$

The work-cost is therefore minimized if the energy of the final state is minimized. If $\beta(N) > 0$, this is the case if the von Neumann entropy is minimized, because we are dealing with Gibbs states. Since the von Neumann entropy can only increase during a thermodynamic protocol, we therefore conclude that the work-cost is minimized in a reversible process. We thus obtain again the minimum work-principle just as in phenomenological thermodynamics as long as the inverse temperature does not change sign during the process. This is the case as long as the final energy fulfills

$$\text{Tr}\left(\omega_{\beta(N)}H(1)\right) \leq \frac{1}{d}\text{Tr}(H(1)), \tag{10.47}$$

where d is the Hilbert-space dimension of the total Hamiltonian $H(1)$. In a system-bath setting with a large bath, this is fulfilled as long as we do not have a work-cost that scales extensively with size of the bath. If we fix an initial and final Hamiltonian on the system, this is true as long as the bath is large enough, since the work cost is bounded by $\|H(0)\| + \|H(1)\|$ but the total energy diverges with the size of the bath.

Also note that if the final temperature in the reversible process $\beta(1)$ is positive, then the initial temperature also must have been positive. This is due to the fact that the temperature $\beta(u)$ is a smooth function. Thus if it would change sign in the process it would have to take the value $\beta(u) = 0$ in which case the entropy would take the maximum possible value and therefore cannot be preserved.

If we consider a cyclic process, $H(0) = H(1)$ and the initial temperature is positive, the above arguments also imply that no positive work can be extracted since the system can only "heat up" in a cycle by creating entropy. We thus obtain the second law of thermodynamics in terms of work extraction. The above discussion is not very surprising, but mostly shows that our framework is adequate and reproduces standard results in the limit where they should apply. We will now go further and discuss the minimum work principle in the case of GGEs, starting with the time-average equilibration model.

10.4.2 Minimum work principle in the time-average equilibration model

Let us again fix a smooth trajectory of Hamiltonians $H(u)$ and assume some initial equilibrium state $\omega_{\text{t.a.}}(0)$. We already know from Lemma 10.2 that in the quasi-static limit, the spectrum of the density matrix $\omega_{\text{t.a.}}(u)$ is conserved. The question is whether this also implies that the work-cost of the process is minimized. We will now see that this is in general only the case if the final state of the quasi-static process, $\omega_{\text{t.a.}}(1)$ is a so-called *passive* state with respect to arbitrary unitaries. To explain what this means, consider the

final Hamiltonian with energies $E(1)_k$. The state $\omega_{\text{t.a.}}(1)$ is diagonal in this basis. Let $\omega_{\text{t.a.}}(1)^\downarrow$ denote the vector with the eigenvalues of $\omega_{\text{t.a.}}(1)$ ordered in non-increasing order. Then $\omega_{\text{t.a.}}(1)$ is called *passive* if its eigenvalues decrease with increasing energies: $(\omega_{\text{t.a.}}(1))^\downarrow_k \geq (\omega_{\text{t.a.}}(1))^\downarrow_l$ implies $E_k(1) \leq E_l(1)$ for all k and l. Passive states have the property that their average energies can only be *increased* using arbitrary unitary operations [236, 237] (however, also see the recent work [238]):

$$\text{Tr}(\rho H) \leq \text{Tr}(U\rho U^\dagger H), \quad \forall U \iff \rho \text{ is passive w.r.t. } H. \qquad (10.48)$$

With this definition at hand we can now show the following Lemma.

Lemma 10.7 (Passive quasi-static protocols are optimal). *Given a smooth trajectory of Hamiltonians $H(u)$ and an initial state $\omega_{\text{t.a.}}(0)$, if the final state in the quasi-static realisation of the process is passive, then the work-cost is minimized in the quasi-static realisation of the protocol.*

Proof. In the quasi-static limit, the final state $\omega_{\text{t.a.}}(1)$ is related to the initial state by a unitary operation W: $\omega_{\text{t.a.}}(1) = W\omega_{\text{t.a.}}(0)W^\dagger$. This follows since their spectra are identical. Now consider any discretization of the process with final state $\omega_{\text{t.a.}}^{(N)}$ and note that the time-averaging process can be seen as a mixture of unitaries. Since $\omega_t a^{(j+1)}$ is obtained from $\omega_t a^{(j)}$ by time-averaging and this holds for all j, the final state $\omega_{\text{t.a.}}^{(N)}$ is to the initial state and to $\omega_{\text{t.a.}}(1)$ by a mixture of unitaries:

$$\omega_{\text{t.a.}}^{(N)} = \sum_\alpha p_\alpha V_\alpha \omega_{\text{t.a.}}(0) V_\alpha^\dagger = \sum_\alpha p_\alpha (V_\alpha W^\dagger) \omega_{\text{t.a.}}(1) (V_\alpha W^\dagger)^\dagger, \qquad (10.49)$$

where U_α are some unitary matrices. Since $\omega_{\text{t.a.}}(1)$ is passive, the state $\omega_{\text{t.a.}}^{(N)}$ can therefore only have higher energy than $\omega_{\text{t.a.}}(1)$. $\square$

The Lemma establishes that the minimum work principle holds if the final state in the quasi-static realisation is passive. Given any two Hamiltonians $H(0)$ and $H(1)$ one can in fact always construct a smooth trajectory such that the final state is passive (for an explicit construction, I refer to Ref. [3]). Furthermore, if the spectrum of the initial state is non-degenerate all such trajectories are equivalent in the sense that they have the same work-cost, since the ordering into a passive state is unique. However, in general, such trajectories require changing the Hamiltonian over time globally. Thus, in a system-bath setting it might be impossible to find a trajectory of the local Hamiltonian of the system so that the final state is passive and in principle it can become beneficial to implement a protocol rapidly instead of in a quasi-static way to extract the most work.

In the case of cyclic processes, the above considerations show that the optimal protocol is one where the final state is passive and has the same spectrum as the initial state. We then conclude from (10.48) that we can extract work in a cyclic protocol if and only if the initial state is non-passive. Thus non-passive states can be considered "non-equilibrium" states in the framework of the time-average equilibration model, even if they are diagonal in the energy eigenbasis: Work can be extracted from them, but only once, since they end up being passive.

Lemma 10.7 also can be seen as a generalisation of previous results in Ref. [250], where the minimum work principle was studied for cyclic unitary processes. In Ref. [250], the authors showed that the minimum work principle is valid if i) the initial state is passive with respect to $H(0)$ *and* ii) the trajectory of Hamiltonians does not induce level crossing. This means that $E_i(0) \geq E_j(0)$ then also $E_i(1) \geq E_j(1)$ (due to the smoothness of the trajectory, the labelling of the basis is fixed throughout the trajectory). Now, if the initial state is passive then the absence of level-crossings is equivalent to having a final passive state. Therefore the Lemma naturally generalizes this result from Ref. [250].

Finally, let us note that in the case of the Gibbs-equilibration model, the condition that the final temperature remains positive can be seen as enforcing passivity of the final state since Gibbs states are passive if and only if they have a positive temperature. We can then formulate the results of section 10.4.1 in complete analogy to Lemma 10.7:

Lemma 10.8 (Passive quasi-static protocols are optimal in the Gibbs equilibration model)**.**
Consider a smooth trajectory of Hamiltonians $H(u)$ and an initial state $\omega_{\beta(0)}(H(0))$. If the final state in the quasi-static realisation of process in the Gibbs equilibration model is passive, then the quasi-static realisation is optimal.

We can thus summarize that both for the time-average and the Gibbs equilibration model, we can check whether the minimum work-principle is fulfilled by calculating the final state of the quasi-static protocol. If it is passive, then the minimum work principle holds. Let us now connect the minimum work principle to the problem of work-extraction from non-equilibrium states and then turn to GGEs with an intermediate number of conserved quantities.

10.4.3   Minimum work principle and work extraction protocols

The minimum work-principle can also be used to study optimal work-extraction protocols from *non-equilibrium states*. For the purpose of this section, let us therefore assume that the equilibration model is given by either the time-average model or the Gibbs model. Then suppose the system is initially in some initial state $\rho(0)$, which is in general *not* an equilibrium state and we want to extract work from the system in a cyclic thermodynamic protocol. We will assume that it is possible to perform a quench before the system equilibrates the first time. Then we can decompose any cyclic work-extraction protocol into two steps. First, the first quench with work-cost

$$W_{\mathrm{GGE}}(0 \to 1) = \mathrm{Tr}\left(\rho(0)(H^{(0)} - H^{(1)})\right). \tag{10.50}$$

After this first quench, the system equilibrates and a thermodynamic protocol back to the initial Hamiltonian $H^{(0)} = H^{(N)}$ follows. Crucially, for this latter part of the protocol, the system starts in the equilibrium state

$$\omega^{(1)}_{\mathrm{t.a.,Gibbs}} = \omega_{\mathrm{t.a.,Gibbs}}(\rho(0), H(1)) \tag{10.51}$$

and leads to some work-cost $W_{\mathrm{GGE}}(1 \to N)$. Here, the notation $\omega_{\mathrm{t.a.,Gibbs}}$ means that we either consider the time-average or Gibbs equilibration model. Clearly, after fixing $H(1)$ we want to minimize the work-cost of this second part of the protocol. In particular, we know from the last section that if the final state of the reversible realisation of this protocol has a final passive state, then it is optimal to perform the reversible process back to the initial Hamiltonian. To optimize the work extraction we then have to optimize the Hamiltonian $H^{(1)}$ after the first quench and the trajectory back to the initial Hamiltonian. This shows that the optimal protocol for work-extraction consists of two steps:

1. A quench to some optimal Hamiltonian H^*, which depends on the initial state $\rho(0)$.

2. A reversible process back to the initial Hamiltonian $H(0)$.

We can now use this to show how much work can be extracted *in principle* from a non-equilibrium state in both the Gibbs-equilibration model and the time-average equilibration model (similar results were obtained in [250]) if we assume that all Hamiltonians can be reached by quenches, i.e., $\mathcal{H}$ consists of all bounded Hamiltonians. First, we obtain for the total work in the process

$$\begin{aligned}
W_{\mathrm{GGE}} &= \mathrm{Tr}\left(\rho(0)H(0)\right) - \mathrm{Tr}\left(\rho(0)H^*\right) \\
&\quad + \mathrm{Tr}\left(\omega^{(1)}_{\mathrm{t.a.,Gibbs}}H^*\right) - \mathrm{Tr}\left(\omega_{\mathrm{t.a.,Gibbs}}(1)H(0)\right) \\
&= \mathrm{Tr}\left(\rho(0)H(0)\right) - \mathrm{Tr}\left(\omega_{\mathrm{t.a.,Gibbs}}(1)H(0)\right),
\end{aligned} \tag{10.52}$$

where $\omega_{\mathrm{t.a.,Gibbs}}(1)$ is the final state of the reversible process with the initial state

$$\omega^{(1)}_{\mathrm{t.a.,Gibbs}} = \omega_{\mathrm{t.a.,Gibbs}}(\rho(0), H^*)$$

and initial Hamiltonian H^* back to the initial Hamiltonian $H(0)$. We therefore have to minimize the energy of $\omega_{\text{t.a.,Gibbs}}(1)$ subject to the condition that its entropy is larger or equal to that of $\rho(0)$. This is achieved if we choose H^* such that

$$\rho(0) = \omega_{\text{t.a.,Gibbs}}(\rho(0), H^*) \tag{10.53}$$

and such that $\rho(0)$ is passive with respect to H^*. This simply means that the first quench is such that $\rho(0)$ is an equilibrium state with respect to the Hamiltonian H^* after the first quench and no entropy production occurs upon equilibration with respect to this Hamiltonian (note the similarity with the optimal work-extraction protocols in chapter 9).

If we define $W^{\max}_{\text{t.a.,Gibbs}}(\rho(0), H(0))$ to be the supremum of the work that can be extracted in the two settings, respectively, then we obtain

$$W^{\max}_{\text{Gibbs}}(\rho(0), H(0)) = \text{Tr}\left(\rho(0)H(0)\right) - \text{Tr}\left(\omega_\beta(H(0))H(0)\right) \tag{10.54}$$

$$= \Delta F_\beta(\rho(0), H(0)), \tag{10.55}$$

$$W^{\max}_{\text{t.a.}}(\rho(0), H(0)) = \text{Tr}\left(\rho(0)H(0)\right) - \text{Tr}\left(\rho(0)^\downarrow H(0)\right), \tag{10.56}$$

where $\beta > 0$ is chosen such that $S(\rho(0)) = S(\omega_\beta(H(0)))$ and $\rho(0)^\downarrow$ is the state that is passive with respect to $H(0)$ and has the same spectrum as $\rho(0)$. The quantity $\text{Tr}((\rho(0)H(0)) - \text{Tr}\left(\rho(0)^\downarrow H(0)\right)$ is also known as *ergotropy* [252] and coincides with the amount of work that can be extracted under purely unitary operations, as discussed in the last section.

In summary we conclude that the maximum amount of work that can be extracted from a non-equilibrium state is given by the energy-difference to the equilibrium state with the same entropy. We will now see that this principle fails in the case of arbitrary GGEs.

10.4.4   *The minimum work principle and work extraction in the GGE equilibration model*

For the Gibbs equilibration model and the time-average equilibration model, we can check whether the minimum work principle is valid in a given thermodynamic protocol by looking at whether the final state in the quasi-static realisation is passive. This is possible in the Gibbs equilibration model due to the close link between energy, entropy and temperature and in the case of the time-average model due to its close connection to passivity and random unitary quantum channels. In the case of arbitrary GGEs both these links are not available to us. It is therefore not surprising that we won't be able to formulate general statements in the sense of the previous two sections. In this section, we will therefore discuss numerical examples using free Fermions instead. In particular, we will see examples for the following behaviours:

1. Initial conditions and a Hamiltonian trajectory where the minimum work principle remains valid.

2. Initial conditions and a Hamiltonian trajectory where the minimum work principle fails, even though the effective description in terms of Gibbs states would suggest that it is valid.

We will further discuss in detail how these behaviours can be understood and connect them, in the spirit of the last two sections, to the concept of passivity. The reasons for discussing free fermionic systems plentiful:

   i) They are known to be well described by GGEs, but not by Gibbs states, due to their integrability [40, 42].

  ii) They can be efficiently simulated on a computer, even for large system sizes.

 iii) They can be simulated with ultra-cold atom in optical lattices, making our predictions testable in experiments [170, 173, 253, 254].

While focussing on free fermionic systems here, we can expect that similar results can be derived for free bosonic systems as well, which have a very similar structure. Here, the aim

is primarily to present examples that show different behaviour, including a violation of the minimum work principle, and which can be understood conceptually. For more detailed numerical investigation that appeared as follow-up to the work presented in this chapter see Ref. [255].

In the following, we will consider a simple one-dimensional chain of n free fermions, with an initial Hamiltonian of the form

$$H(0) = \sum_{j=1}^{n} \epsilon_j f_j^\dagger f_j + g \sum_{j=1}^{n-1} (f_j^\dagger f_{j+1} + f_{j+1}^\dagger f_j), \qquad (10.57)$$

where the ϵ_j denote on-site potentials and g determines the amplitude for hopping between neighboring sites. Any such Hamiltonian can be brought into the normal-form

$$H(0) = \sum_{k=1}^{n} \mu_k \eta_k^\dagger \eta_k = \sum_{k=1}^{n} \mu_k n_k, \qquad (10.58)$$

where $\eta_k, \eta_k^\dagger$ are again fermionic operators that are related to the f_j by a canonical transformation (a $n \times n$ unitary matrix). The number operators n_k mutually commute and hence also commute with the Hamiltonian $H(0)$. They provide the relevant set of conserved quantities in the case of initial states that are Gaussian. For a more detailed discussion of this point see Ref. [249]. Gaussian states are those states that are fully determined by the correlation functions $C(\rho)_{ij} := \mathrm{Tr}(\rho f_i f_j)$ through Wick's theorem and include eigenstates and Gibbs-states of quadratic Hamiltonians such as $H(0)$. Importantly, if we consider the GGE with the conserved quantities n_k it takes the form

$$\omega_{\mathrm{GGE}}(\rho, H(0), \{n_k\}) = \frac{\mathrm{e}^{-\sum \lambda_k n_k}}{Z}. \qquad (10.59)$$

It therefore can be understood as a Gibbs state of the quadratic Hamiltonian $\sum_k \lambda_k n_k$ and is also a Gaussian state. This is even the case if the state ρ from which the GGE is determined is not Gaussian. This is well in accordance with recent results that show that a non-Gaussian initial state evolves towards a Gaussian one under the dynamics of a short-ranged quadratic Hamiltonian [189, 196]. Furthermore, it implies that we can completely focus our attention to Gaussian states even if the initial state of the thermodynamic protocol is not Gaussian, since nothing changes on the level of the effective description if we replace this initial state with the Gaussian state which has the same correlation functions $C(\rho)_{ij}$. This is even true on the level of exact, time-dependent, unitary dynamics, since time-evolution under quadratic Hamiltonians maps Gaussian states to Gaussian states and can be expressed solely on the level of the correlation matrices in closed form.

Since we can restrict to Gaussian states, all the analysis of thermodynamic protocols can be reduced to the level of correlation matrices $C(\rho)$. These are $n \times n$-matrices in contrast to the full density matrices of size $2^n \times 2^n$, which enables us to compute the full, unitary time-evolution of system and bath together in an efficient manner and compare it with the effective description.

In the following I will discuss two exemplary cases of work-extraction protocols from different initial conditions. To do that let us split the Hamiltonian $H(0)$ into three parts: a system Hamiltonian on the first site $H_S = \epsilon_1 f_1^\dagger f_1$, an interaction term $V = g(f_1^\dagger f_2 + f_2^\dagger f_2)$ and the bath Hamiltonian $H_B = H - H_S - V$. We now assume that we can only change the Hamiltonian by adjusting the potential ϵ_1 on the first site and that $\epsilon_j = \epsilon$ is fixed for all $j \neq 1$. The two examples will be optimal work-extraction protocols from different initial conditions, leading to completely different behaviour. Before discussing the results in the specific examples, let us first discuss optimal work-extraction protocols in the case of free fermionic systems with GGEs as effective description.

As discussed above, the whole analysis can be carried out on the level of correlation matrices. On the level of correlation matrices, going from a state ρ to its associated Generalized Gibbs ensemble corresponds to projecting the correlation matrix $C(\rho)$ to its diagonal in the basis of the normal-modes η_k. It is therefore very similar from a formal point of view to the case of the time-average equilibration model on the level of density matrices. To make this clearer, let us call a correlation matrix C *passive* with respect to the Hamiltonian
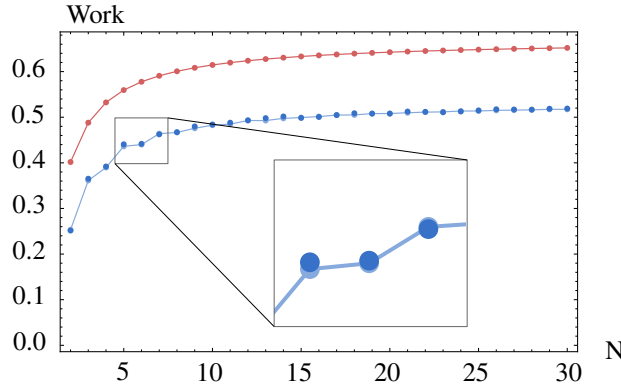
Figure 10.2: Extracted work with quenches only on a single site of the chain of Fermions in the first example. Dark blue correspond to the work W computed from the exact unitary evolution, light blue points to the work W_{GGE} computed from the effective description in terms GGE states, and red points to the effective description using Gibbs states. As an initial state we take, $\beta = 1/2$, $\mathrm{Tr}(f_1^\dagger f_1 \rho(0)_S) = 0.1$, $n = 100$. For the initial Hamiltonian, $\epsilon_0 = 0.1$, $\epsilon_i = 1\ \forall i \neq 1$, $g = 0.5$. The protocol consists of a first quench to $\epsilon_1 = 4.3$, followed by $N-1$ equidistant quenches back to the original Hamiltonian. The exact evolution is obtained by letting system and bath interact for a random time between $20/g$ and $100/g$ much larger than the equilibration time. (Figure adapted from Ref. [3].)

$H = \sum_k \mu_k n_k$, if it is diagonal in the basis of the normal-modes and the populations of the normal modes decrease with increasing energy of the mode:

$$C_{kk} \geq C_{ll} \implies \mu_k \leq \mu_l. \tag{10.60}$$

It then follows by essentially the same arguments as in the time-average equilibration models that optimal protocols have the property that the final correlation matrix has the same spectrum as the initial correlation matrix and is passive. Similarly, the minimum work principle holds in the case that the final correlation matrix in the reversible limit is passive.

Not again however, that optimal protocols in general require being able to quench to arbitrary free fermionic Hamiltonians and hence one usually cannot expect to reach this bound in the system-bath setting. Similarly, in the system-bath setting it might be impossible to find protocols with final states that have a passive correlation matrix.

A further point that will become important later is that a passive correlation matrix does *not* imply that the state from which it is calculated is passive as a quantum state with respect to the Hamiltonian on the full many-particle Hilbert-space. A simple example of such behaviour is discussed in section 14.7.1. Let us now come to the specific examples.

10.4.5   Example: Work extraction in accordance with the minimum work principle

In the first example, we consider as initial state a state of the form

$$\rho(0) = \rho(0)_S \otimes \omega_\beta(H_B), \tag{10.61}$$

where $\rho(0)$ is out of equilibrium. We will here choose $\rho(0)$ as a thermal state of H_S with an inverse temperature that differs from β. Then, we numerically implement a work extraction protocol that consists of a first rapid quench before letting the system equilibrate and then a thermodynamic protocol back to the initial Hamiltonian. We can then calculate both the actual work done in the process W and the work as calculated from the effective description in terms of GGEs and Gibbs states. The results are presented in Fig. 10.2 and demonstrate that the minimum work principle holds in this case.

10.4.6   Example: Work extraction violating the minimum work principle

For the second example, instead of taking a Gibbs state as the initial state of the bath-part of the system, we choose a GGE initial state $\omega_{\mathrm{GGE}}^{(B)}$, which we want to choose to find an example that violates the minimum work principle. From the previous discussions, we
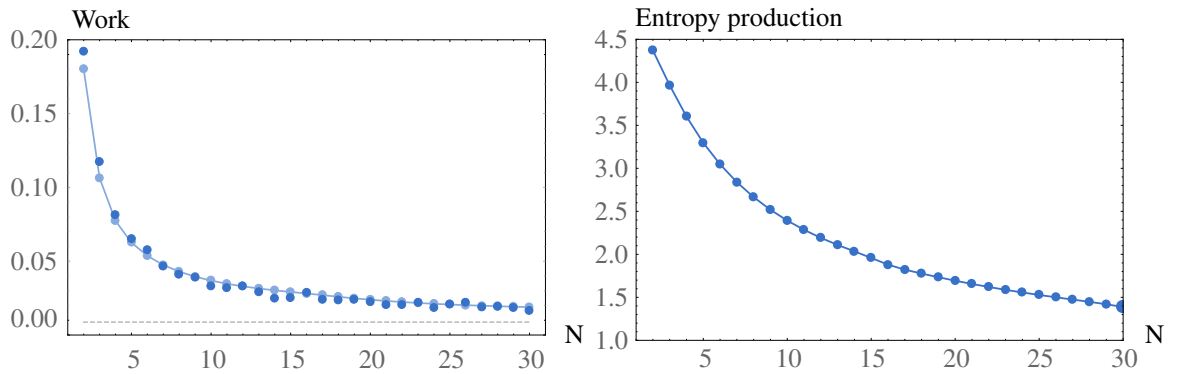
Figure 10.3: *Left:* Extracted work with quenches only on a single site of the chain of Fermions in the second example. Green points correspond to the work W computed from the exact unitary evolution, yellow points to the work W_{GGE} computed from the effective description in terms GGE states. As an initial state we take the one specified by $K = 32$, $\text{Tr}(f_1^\dagger f_1 \rho(0)_S) = 0.1$, $n = 150$. For the initial Hamiltonian, $\epsilon_0 = 0.1$, $\epsilon_i = 1$ $\forall i \neq 1$, $g = 0.5$. The protocol consists of a first quench to $\epsilon_1 = 1.6$, followed by $N - 1$ equidistant quenches back to the original Hamiltonian. The exact evolution is obtained by letting system and bath interact for a random time between $20/g$ and $100/g$ much larger than the equilibration time. *Right:* The entropy production in the same protocol. (Figures adapted from Ref. [3].)

learn that we should try to engineer it such that the final state of the reversible limit is not passive. We therefore choose the GGE in the following way:

$$\text{Tr}\left(\omega_{\text{GGE}}^{(B)} n_k^{(B)}\right) = \begin{cases} 1 & \text{if} \quad k \leq K, \\ 0 & \text{if} \quad k > K. \end{cases} \qquad (10.62)$$

Here, $n_k^{(B)}$ denote the number operators of the normal modes of the bath Hamiltonian H_B and we assume that the corresponding single-particle energies $\mu_k^{(B)}$ are ordered in decreasing fashion. Therefore only the K most energetic normal modes are populated in the initial state $\omega_{\text{GGE}}^{(N)}$. This ensures that the initial correlation matrix of the system is not passive. Note that such behaviour would be impossible to have in a Gibbs state with positive temperature. Nevertheless, if we would use an effective description in terms of Gibbs states, then the effective temperature of this state would be positive for large enough system sizes and fixed K since the energy-density in the state $\omega_{\text{GGE}}^{(B)}$ is well below the critical energy density. This implies that from the point of view of the Gibbs equilibration model, we would predict that the minimum work principle would be fulfilled in a work-extraction protocol.

Fig. 10.3 shows the results of a work-extraction protocol from these initial condition. As can clearly be seen, the minimum work principle is violated: the extracted work monotonically decreases with the number of steps in the second part of the protocol and thus in the reversible limit. Indeed, we find that the final state in the reversible limit is highly non-passive, explaining the breakdown of the minimum work principle. This finishes our discussion of entropy production and the minimum work principle in Generalized Gibbs ensembles.

10.5   Summary

It is generally assumed that sufficiently interacting, non-integrable systems equilibrate and even thermalize after a quench of the Hamiltonian. This means that local observables can be described by global Gibbs states with the same energy-density as the initial state. Nevertheless, there are class of many-body systems for which this is not true, such as integrable systems [192–202] or many-body localizing systems, which equilibrate but do not thermalize [224–226]. The equilibrium states of such systems have to be described by Generalized Gibbs ensembles.

In this chapter, we built up a framework for thermal machines and studied its predictions for different kind of equilibration behaviour of many-body systems on the basis of an effective description in terms of Generalized Gibbs ensembles. In this framework, we assume that Generalized Gibbs ensembles provide a good description not only for single

quenches but, importantly, also for many consecutive quenches. This allowed us to derive general results about entropy production and the minimum work principle in a unifying language. Importantly, the statement that the entropy can only increase along a thermodynamic protocol of several quenches in an isolated system follows straightforwardly without any additional assumptions.

As expected, systems that thermalize allow us to derive standard statements of phenomenological thermodynamics, such as the minimum work principle. This is even true for finite baths, where standard thermodynamic arguments fail since the bath changes its temperature over the course of a thermodynamic protocol. Importantly, these results also hold true if we cannot clearly separate systems into "working systems" and "heat baths". What we call "system" and what we call "bath" is simply determined by our control capabilities, as in chapter 9.

For the case of arbitrary GGEs we found that the usual thermodynamic statements are not automatically true. In particular, we discussed an explicit example where the minimum work principle fails. This could be attributed to the fact that GGE states need not be passive states, even though the description on the basis of the constants of motion seems to be passive. Similarly, in the case of the time-average equilibration model and for Gibbs-states we also found that the minimum work principle holds, in a suitable formulation, when the relevant effective descriptions are given by passive states. This suggests that "equilibrium states", in the sense of the second law, are not simply maximum entropy ensembles, given the expectation values of the relevant conserved quantities, but also have to be passive. It is an intriguing open problem to study the role of passivity for Generalized Gibbs ensembles for more general systems then just free fermionic or free bosonic systems.

As discussed above, the framework introduced in this chapter assumes that we can use the effective description after a first quench also to predict the new effective description after a second quench. This goes beyond of what is usually considered in theory and experiments and deserves further studies both from a theoretical point of view, but also from the experimental point of view. Indeed, using optical lattices or trapped ion platforms, it seems entirely plausible that this assumption can be tested experimentally with similar set-ups as for single-quench experiments [169–174].

I already mentioned above that the results and the framework in this chapter are not restricted to any weak coupling limit. In the next chapter, we will now use this framework to study thermal machines in the strong coupling regime and derive general corrections to work-extraction bounds and efficiencies of heat engines.

11   Corrections to work and efficiency under strong coupling

IN MACROSCOPIC THERMODYNAMICS, the interaction energies between different macroscopic bodies can usually be safely neglected. This is due to the fact that interactions usually act *locally* in space. Therefore the interaction energy between two macroscopic bodies scales only like the contact area between the two bodies, while thermodynamic quantities such as the internal energy, entropy, free energy and extractable work scale with the volume of the bodies. For microscopic systems, this is no longer true and the corrections due to finite coupling strengths need to be considered. Deriving the thermodynamic bounds that are applicable in the strong coupling regime is a well-known problem and recently numerous studies dealt with thermodynamics in the strong-coupling regime (see Refs. [3, 76, 77, 235, 256–270]). Nevertheless, general and concrete corrections to thermodynamic bounds such as the optimal efficiency of a quantum thermal machine which can be evaluated for arbitrary models have been lacking. In this chapter[1], we use the thermodynamic framework developed in the last chapter to derive such corrections for extractable work and the efficiency of a thermal machine operating between two heat baths. In particular, we will obtain corrections to second order in the coupling strength, which can in principle be evaluated for any model. To derive the results, we will make the following assumptions:

[1] This chapter is based on joint-work with Martí Perarnau-Llobet, Arnau Riera, Rodrigo Gallego and Jens Eisert published in Ref. [9].

1. We consider a working system that can be coupled to heat baths.

2. When (and only when) the working system is coupled to a heat bath, it equilibrates.

3. We assume that the Gibbs equilibration model is applicable to any working system in contact with a heat bath. This assumption should be understood as a condition on possible heat baths and means that observables on the working system are correctly predicted by the Gibbs state on the full compound.

4. We consider arbitrarily large heat baths, so that all effective temperatures remain constant. We also assume that all effective temperatures are positive.

5. The interaction operator between the system and each of the heat baths is fixed and can only be turned on or off.

Apart from these assumptions, we will introduce no general additional assumptions on top of the framework for thermodynamic operations from the last chapter. We will then derive the fundamental bounds for the tasks of i) work-extraction from a single heat bath using a non-equilibrium working system and ii) optimal efficiencies of thermal machines operating between two heat baths.

11.1   *Work extraction in the strong coupling regime*

Since work-extraction bounds are necessary to derive bounds on efficiencies, let us begin with the task of extracting work from a system that is initially out of equilibrium and uncoupled from the heat bath with Hamiltonian H_B. We can then apply quenches to the Hamiltonian that consist of either changing the system Hamiltonian or turning on or off the coupling to the heat bath. The possible Hamiltonians during a work-extraction protocol are then of the form

$$H^{(j)} = H_S^{(j)} + H_B(+V), \tag{11.1}$$

where the interaction term V is only present when the system is contact with the bath. As before we associate a work-cost

$$W(j \to j+1) = \mathrm{Tr}(\rho^{(j)} \left(H^{(j)} - H^{(j+1)} \right) \tag{11.2}$$

to a quench of the Hamiltonian. Apart from the quenches that involve turning on or off the coupling to the heat bath, the work-cost can be computed from the state on subsystem S only, since we only change the Hamiltonian on S. If the system is coupled to the heat bath and we let the system and bath equilibrate, the final effective description is given by the Gibbs-state at temperature β. Thus, after the quench from $H^{(j)}$ to $H^{(j+1)}$, the state of the system changes according to

$$\rho_S^{(j)} \mapsto \mathrm{Tr}_B \left(\omega_\beta(H^{(j+1)}) \right). \tag{11.3}$$

As discussed in the previous chapter, if we fix initial Hamiltonian $H(0)$ and a final Hamiltonian $H(1)$ (with interaction turned on) and consider an initial equilibrium state $\omega_\beta(H(0))$, we can perform a quasi-static and reversible protocol from $H(0)$ to $H(1)$. This isothermal reversible process has work cost

$$W = F_\beta(\omega_\beta(H(0)), H(0)) - F_\beta(\omega_\beta(H(1)), H(1)). \tag{11.4}$$

The proof is exactly the same as for the isothermal process in the weak-coupling regime as presented in section 14.6.1 or in Ref. [235], due to the fact that we assume that β is fixed (since the bath is arbitrarily large). Note however, that the work-cost is here expressed in terms of the state and Hamiltonian on system and bath together, even though the quenches only occur on the system S.

   We have seen already in chapters 9 and 10 that optimal work-extraction bounds minimize dissipation, which is ultimately due to the minimum work principle. In this context, this principle implies that an optimal work-extraction protocol from the initial system state $\rho(0)$ and Hamiltonian $H_S(0)$ consists of the following steps:

1. A quench to a Hamiltonian $H_S^{(1)}$ while remaining uncoupled to the bath and turning on the interaction afterwards. This step has a work-cost

$$W_1 = \mathrm{Tr} \left(\rho(0) \otimes \omega_\beta(H_B) \left(H(0)_S - H_S^{(1)} - V \right) \right). \tag{11.5}$$

2. An isothermal reversible process to some Hamiltonian $H_S^{(f)}$ with work-cost

$$W_2 = F_\beta(\omega_\beta(H^{(1)}), H^{(1)}) - F_\beta(\omega_\beta(H^{(f)}), H^{(f)}). \tag{11.6}$$

3. Turning off the interaction and returning to the initial Hamiltonian, with work-cost

$$W_3 = \mathrm{Tr} \left(\omega_\beta(H^{(f)}) \left(H_S^{(f)} + V - H(0)_S \right) \right). \tag{11.7}$$

As we have seen in chapter 9, in the weak coupling limit it is optimal to take $H_S^{(f)} = H(0)_S$ and one obtains, upon optimizing $H_S^{(1)}$, the weak coupling bound

$$W^{(\mathrm{weak})}(\rho(0), H(0)_S) = \Delta F_\beta(\rho(0), H(0)_S). \tag{11.8}$$

We will now express the total work in the optimal protocol of the strong coupling case as corrections to this weak coupling limit. The result (see section 14.8.1) is

$$W(\rho(0)_S, H(0)_S) = W_1 + W_2 + W_3 \tag{11.9}$$

$$= W^{(\text{weak})}(\rho(0), H(0)_S) - \Delta F^{(\text{irrev})}(H_S^{(1)}) - \Delta F^{(\text{res})}(H_S^{(f)}), \tag{11.10}$$

with the correction terms

$$\Delta F^{(\text{irrev})}(H_S^{(1)}) = \frac{1}{\beta} D\left(\rho(0) \otimes \omega_\beta(H_B) \| \omega_\beta(H_S^{(1)} + V + H_B)\right), \tag{11.11}$$

$$\Delta F^{(\text{res})}(H_S^{(f)}) = \frac{1}{\beta} D\left(\omega_\beta(H_S^{(f)} + V + H_B) \| \omega_\beta(H(0)_S) \otimes \omega_\beta(H_B)\right). \tag{11.12}$$

Since both terms are expressed as relative entropies, it is clear that they are positive. The optimal work that can be extracted from the initial state is then given by optimizing the Hamiltonians $H_S^{(1)}$ and $H_S^{(f)}$. Further below, we will derive the exact conditions that characterize the solutions of this minimization problem. We will also provide the exact solutions to second order in the coupling strength. Before we come to that, however, let us first discuss the two penalty terms heuristically.

The first penalty term, given by $\Delta F^{(\text{irrev})}$, describes the dissipation that occurs when the system is coupled to to the bath and compares the state after equilibration with the state before equilibration. Due to the interaction, there is an unavoidable build-up of correlations, which are not present in the initial state. Hence, in general this term cannot be made zero for any finite interaction. In the weak coupling limit, on the other hand, the term can be made vanishingly small by choosing $H_S^{(1)}$ as the modular Hamiltonian $H_{\rho(0)}^\beta$ (cf. section 5).

The second penalty term, given by $\Delta F^{(\text{res})}$, can be interpreted as the residual free energy with respect to the initial, uncoupled Hamiltonian that is left at the end of the isothermal protocol. This free energy could in principle be extracted, if the Hamiltonian could be quenched globally instead of only on the system. Again, in the weak-coupling limit this term becomes arbitrarily small by taking $H_S^{(f)} = H(0)_S$.

Observe that the above discussion already shows that, if we treat the interaction as a perturbation, all corrections to the weak coupling bound vanish to first order in the interaction strength. Since the relative entropy vanishes if and only if the two arguments coincide, we can also expect already that corrections to first order in the coupling strength vanish as well. We will later verify these statements explicitly. Let us now discuss general properties of the correction terms.

11.1.1   General properties of the correction terms

The correction terms formally depend on the whole state and Hamiltonian of the system and bath together. In principle, they could therefore scale extensively with the size of the bath, as is typically the case for non-equilibrium free energies. The first result that we will discuss shows that this is not the case.

Lemma 11.1 (Scaling of corrections). *For any $\rho(0) > 0, H(0)_S, V$ and H_B the correction terms fulfill*

$$\Delta F^{(\text{res})}(H_S^{(f)}) \leq 2\left(\|V\| + \left\|H(0)_S - H_S^{(f)}\right\|\right), \tag{11.13}$$

$$\Delta F^{(\text{irrev})}(H_S^{(1)}) \leq 2\left(\|V\| + \left\|H_S^{(1)} - H_{\rho(0)}^\beta\right\|\right). \tag{11.14}$$

Proof. For any two Hamiltonians A, B, we have

$$\frac{1}{\beta} D(\omega_\beta(A) \| \omega_\beta(B)) = \text{Tr}\left((A - B)(\omega_\beta(B) - \omega_\beta(A))\right) - \frac{1}{\beta} D(\omega_\beta(B) \| \omega_\beta(A))$$

$$\leq 2 \|A - B\|, \tag{11.15}$$

where the last inequality follows from the positivity of the relative entropy, the triangle inequality and the definition of the operator norm. The claim then follows by inserting the corresponding Hamiltonians and interpreting $\rho(0)$ as thermal state of the modular Hamiltonian $H^{\beta}_{\rho(0)}$. $\qquad\square$

In particular, the lemma shows that the optimal corrections fulfill

$$\Delta F^{(\text{res})}_{\min} := \min_{H^{(f)}_S} \Delta F^{(\text{res})}(H^{(f)}_S) \leq 2\,\|V\|, \qquad (11.16)$$

$$\Delta F^{(\text{irrev})}_{\min} := \min_{H^{(1)}_S} \Delta F^{(\text{irrev})}(H^{(1)}_S) \leq 2\,\|V\|. \qquad (11.17)$$

These relations show that the strong coupling corrections are indeed negligible in the macroscopic limit if the interaction between two macroscopic bodies is given by a local interaction, since in such a case the norm of the interaction $\|V\|$ only scales like the contact area between the two bodies. The weak coupling work $W^{(\text{weak})}$, on the other hand, generally scales like the volume of the system S and hence dominates.

All of the above could be deduced without knowing the exact Hamiltonians that minimize the correction terms. We will now derive conditions on the optimal Hamiltonians, which will be instrumental for determining the perturbative expansion of the correction terms.

Lemma 11.2 (Minimizing dissipation). *Let $\rho(0)$ have full support. Then any Hamiltonian X_S that minimizes $\Delta F^{(\text{irrev})}$ has to fulfill*

$$\rho(0) = \text{Tr}_B\left(\omega_\beta(X_S + V + H_B)\right). \qquad (11.18)$$

Proof. Since we consider arbitrary large, but finite systems, $\Delta F^{(\text{irrev})}(H^{(1)}_S)$ is a smooth, positive function of $H^{(1)}_S$. Let one minimum be attained by X_S and consider the Hamiltonians

$$X_S(t) := X_S + t Y_S, \quad X(t) = X_S(t) + V + H_B, \qquad (11.19)$$

where Y_S is an arbitrary perturbation. Then we have

$$\left.\frac{\mathrm{d}\Delta F^{(\text{irrev})}(X(t))}{\mathrm{d}t}\right|_{t=0} = 0. \qquad (11.20)$$

Calculating the derivative yields

$$\begin{aligned}
\left.\frac{\mathrm{d}\Delta F^{(\text{irrev})}(X(t))}{\mathrm{d}t}\right|_{t=0} &= \text{Tr}\left(\left(\rho(0)\otimes\omega_\beta(H_B) - \omega_\beta(X(0))\right) Y_S \otimes \mathbf{1}\right) \\
&= \text{Tr}\left(\left(\rho(0) - \text{Tr}_B(\omega_\beta(X_S + V + H_B))\right) Y_S\right) = 0. \quad (11.21)
\end{aligned}$$

Since this relation has to fold for arbitrary Y_S, the claim follows. $\qquad\square$

The lemma can easily be interpreted. It tells us that to optimize the dissipation in the protocol, we have to do the first quench in such a way that the initial state can be interpreted as the marginal of the corresponding *interacting* Hamiltonian. Even though this minimizes the dissipation, the dissipation does not vanish, but is given by

$$\Delta F^{(\text{irrev})}_{\min} = \frac{1}{\beta} D\left(\text{Tr}_B(\omega_\beta(X(0)))\otimes\omega_\beta(H_B)\|\omega_\beta(X(0))\right). \qquad (11.22)$$

This quantity vanishes only if $\omega_\beta(X(0))$ is a product-state and hence non-interacting. It can thus be seen as a measure of the correlations which are induced by interaction between S and B. It is, however, not a standard-measure of correlations, even though it might easily be confused with the mutual information

$$I(S:B)_{\omega_\beta(X(0))} = D\left(\omega(X(0))\|\omega(X(0))_S \otimes \omega(X(0))_B\right). \qquad (11.23)$$

There are two important differences: First, the order of the arguments in the relative entropy is reversed, somewhat similar in how the "vacancy" in chapter 3 behaves in comparison to the non-equilibrium free energy. Second, the first argument in (11.22) is not given by the product of the marginals of $\omega_\beta(X(0))$ in contrast to the mutual information.

Similar to the condition on the optimal Hamiltonian to minimize dissipation, we can also derive a condition on the Hamiltonian $H_S^{(f)}$ that minimizes the residual free energy. This condition is stated in the following Lemma. Its proof is technically more demanding and therefore given in section 14.8.2. To state the result, we define the following function on arbitrary Hermitian operators Y, X:

$$Y_X := \int_0^1 e^{-\beta\tau X} Y e^{\beta\tau X} d\tau. \tag{11.24}$$

Y_X can be interpreted as the operator Y averaged over the imaginary-time evolution under the operator X.

Lemma 11.3 (Minimizing residual free energy). *Let R_S be a Hamiltonian that minimizes the residual free energy $\Delta F^{(\mathrm{res})}\left(H_S^{(f)}\right)$ and define $R = R_S + V + H_B$. Then R has to fulfill*

$$\mathrm{Tr}_B\left(\omega_\beta(R)\right) = \frac{\mathrm{Tr}_B\left(\omega_\beta(R)\left(R_S + V - H(0)_S\right)_{-R}\right)}{\mathrm{Tr}\left(\omega_\beta(R)(R_S + V - H(0)_S)\right)}. \tag{11.25}$$

In particular, we find that for $V = 0$ we can choose $R_S = H(0)_S$ as expected. Unfortunately, in general and for finite V, the optimal choice R_S is much more difficult to interpret than the corresponding X_S minimizing the dissipation. Let us therefore now consider the perturbative expansion of the correction terms in powers of the coupling strength, which will yield a clearer interpretation.

11.1.2 Perturbative expansion of correction terms

Since the general expression for the strong coupling corrections are somewhat complicated, we will now consider their expansion in terms of the coupling strength. We will therefore replace V with gV and calculate the correction terms to leading order in g. The optimal Hamiltonians that minimizes the two dissipation terms will of course depend on the coupling strength g. We will denote them by $X(g)$ and $R(g)$, respectively. In particular, we already know that

$$X(0) = H_{\rho(0)}^\beta + H_B \tag{11.26}$$

$$R(0) = H(0)_S + H_B. \tag{11.27}$$

Using these results, we can express the dissipation terms as a function of g as

$$\Delta F^{(\mathrm{irrev})}(g) = \frac{1}{\beta} D\left(\omega_\beta(X(0))\|\omega_\beta(X(g))\right), \tag{11.28}$$

$$\Delta F^{(\mathrm{res})}(g) = \frac{1}{\beta} D\left(\omega_\beta(R(g))\|\omega_\beta(R(0))\right). \tag{11.29}$$

Unfortunately, the relative entropy is not a symmetric function of its two arguments. Therefore, in general the two terms show a different behaviour as a function of g. It turns out, however, that to leading order in g, this is not the case.

This is shown by the following lemma. Since the proof consists of key calculations which will be used later, we present it here in full generality.

Lemma 11.4 (Perturbative symmetry of relative entropy). *Let $H(t)$ be smooth one-parameter family of Hamiltonians and define*

$$\Delta(t) := D\left(\omega_\beta(H(t))\|\omega_\beta(H(0))\right) - D\left(\omega_\beta(H(0))\|\omega_\beta(H(t))\right). \tag{11.30}$$

Then for small t, $\Delta(t) = O(t^3)$.

To prove the result, we will make use of the Duhamel's formula for the derivative of an exponential. If $H(t)$ is a smooth, operator-valued function, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{-\beta H(t)} = -\beta \int_0^1 \mathrm{e}^{-\beta s H(t)} H'(t) \mathrm{e}^{-\beta(1-s)H(t)} \mathrm{d}s. \qquad (11.31)$$

Here, and in the following we use the notation $f'(t)$ to denote the derivative of a function $f(t)$.

Proof. Let us define the partition function $Z_t := \mathrm{Tr}(\mathrm{e}^{-\beta H_t})$ and introduce the short-hand $\omega_\beta(t)\omega_\beta(H(t))$ for the proof. Using (11.31), we obtain $\log(Z_t)' = -\beta \mathrm{Tr}(\omega_\beta(t)H'(t))$. We can then calculate the derivative of $D(\omega_\beta(0)\|\omega_\beta(t))$ as

$$D(\omega_\beta(0)\|\omega_\beta(t))' = -\mathrm{Tr}\left(\omega_\beta(0)\log(\omega_\beta(t))'\right) = \beta\mathrm{Tr}\left(\omega_\beta(0)H'(t)\right) + \log(Z_t)'$$
$$= \beta\mathrm{Tr}\left((\omega_\beta(0) - \omega_\beta(t))H'(t)\right). \qquad (11.32)$$

To compute the derivative of $D(\omega_\beta(t)\|\omega_\beta(0))$, let us first compute the derivative of the entropy of $\omega_\beta(t)$:

$$S(\omega_\beta(t))' = -\mathrm{Tr}\left(\omega_\beta(t)\log(\omega_\beta(t))'\right) - \mathrm{Tr}\left(\omega'_\beta(t)\log(\omega_\beta(t))\right)$$
$$= \beta\mathrm{Tr}\left(\omega_\beta(t)H'(t)\right) + \log(Z_t)' - \mathrm{Tr}\left(\omega'_\beta(t)\log(\omega_\beta(t))'\right)$$
$$= -\mathrm{Tr}\left(\omega'_\beta(t)\log(\omega_\beta(t))\right). \qquad (11.33)$$

We then obtain for the derivative of the relative entropy:

$$D(\omega_\beta(t)\|\omega_\beta(0))' = -S(\omega_\beta(t))' - \mathrm{Tr}\left(\omega_\beta(t)\log(\omega_\beta(0))\right)' \qquad (11.34)$$
$$= -\mathrm{Tr}\left(\omega'_\beta(t)\left(\log(\omega_\beta(0)) - \log(\omega_\beta(t))\right)\right) \qquad (11.35)$$
$$= \beta\mathrm{Tr}\left(\omega'_\beta(t)(H(0) - H(t))\right), \qquad (11.36)$$

and for the first derivative of $\Delta(t)$:

$$\Delta'(t) = \beta\left[\mathrm{Tr}\left((\omega_\beta(0) - \omega_\beta(t))H'(t)\right) - \mathrm{Tr}\left(\omega'_\beta(t)(H(0) - H(t))\right)\right]. \qquad (11.37)$$

From this expression we can easily compute the second derivative as

$$\Delta''(t) = \beta\left[\mathrm{Tr}\left((\omega_\beta(0) - \omega_\beta(t))'H'(t)\right) + \mathrm{Tr}\left((\omega_\beta(0) - \omega_\beta(t))H'(t)\right)\right]$$
$$- \beta\left[\mathrm{Tr}\left(\omega''_\beta(t)(H(0) - H(t))\right) + \mathrm{Tr}\left(\omega'_\beta(t)(H(0) - H(t))'\right)\right]$$
$$= \beta\left[\mathrm{Tr}\left((\omega_\beta(0) - \omega_\beta(t))H'(t)\right) - \mathrm{Tr}\left(\omega''_\beta(t)(H(0) - H(t))\right)\right]. \qquad (11.38)$$

In particular, we get $\Delta'(0) = 0$ and $\Delta''(0) = 0$, which proves the claim. $\qquad \square$

The Lemma shows that if we are interested in corrections up to second order in g, it essentially suffices to calculate one of the two expansions, since the functional form of the two is identical. We will therefore now calculate the perturbative expansion of $\Delta F^{(\mathrm{irrev})}(g)$ to second order in g.

From the calculation in the proof of the Lemma, we find

$$\Delta F^{(\mathrm{irrev})'}(g) = \mathrm{Tr}\left((\omega_\beta(X(0)) - \omega_\beta(X(g)))X'(g)\right).$$

Consequently, the second derivative is given by

$$\Delta F^{(\mathrm{irrev})''}(g) = -\mathrm{Tr}\left(\omega_\beta(X(g))'X'(g)\right) + \mathrm{Tr}\left((\omega_\beta(X(0)) - \omega_\beta(X(g)))X'(g)\right).$$

In particular, at $g = 0$ we find

$$\Delta F^{(\text{irrev})'}(0) = 0,$$

$$\Delta F^{(\text{irrev})''}(0) = -\text{Tr}\left(\left.\frac{\mathrm{d}\omega_\beta(X(g))}{\mathrm{d}g}\right|_{g=0}(X_S'(0) + V)\right),$$

where we have used that $X'(g) = X_S'(g) + V$. Importantly, this second order correction only depends on the first order corrections to the optimal Hamiltonian. We can now use (11.31) to obtain

$$\Delta F^{(\text{irrev})''}(0) = \beta \, \text{cov}_{\omega_\beta(X(0))}\left(X'(0), X'(0)\right), \tag{11.39}$$

where we have introduced the *generalized covariance* [206], which for any state ρ and observables A, B is given by

$$\text{cov}_\rho(A, B) := \int_0^1 \text{Tr}\left(\rho^s A \rho^{(1-s)} B\right) \mathrm{d}s - \text{Tr}(\rho A)\text{Tr}(\rho B). \tag{11.40}$$

What is left to do is to determine the first order-correction $X'(0)$ as prescribed by Lemma 11.2. Using again (11.31) and the fact that $\omega_\beta(X(0)) = \omega_\beta(X_S(0)) \otimes \omega_\beta(H_B)$, we find that $\text{Tr}_B(\omega_\beta(X(g)))$ is given to first order by

$$\text{Tr}_B\left(\omega_\beta(X(g))\right) = \omega_\beta(X_S(0))$$
$$- \beta g \left[\text{Tr}_B\left((X'(0))_{X(0)}\omega_\beta(X(0))\right) - \text{Tr}\left((X'(0))\omega_\beta(X(0))\right)\omega_\beta(X_S(0))\right].$$

Since $X(0) = X_S(0) + H_B$, the zero-order solution is given by $X_S(0) = H_{\rho(0)}^\beta$ as we expect. We thus now set $X_S(g) = H_{\rho(0)}^\beta + gX_S$ and determine X_S. The condition in Lemma 11.2 is satisfied to first order if and only if the first order term in the above equation vanishes. Since $X'(0) = X_S + V$, we can re-write this condition as

$$\text{Tr}_B\left((X_S + V)_{X(0)}\omega_\beta(X(0))\right) = \text{Tr}\left((X_S + V)\omega_\beta(X(0))\right)\omega_\beta\left(H_{\rho(0)}^\beta\right). \tag{11.41}$$

Since X_S is supported only on S and $X(0) = H_{\rho(0)}^\beta + H_B$, we can re-write this equation as

$$(X_S)_{H_{\rho(0)}^\beta} - \text{Tr}\left(X_S\omega_\beta\left(H_{\rho(0)}^\beta\right)\right)\mathbf{1} = -\left((V_S)_{H_{\rho(0)}^\beta} - \text{Tr}\left(V_S\omega_\beta\left(H_{\rho(0)}^\beta\right)\right)\mathbf{1}\right),$$

where we defined the operator $V_S := \text{Tr}_B(V\omega_\beta(X(0)))$. Introducing the linear operator

$$L(X_S) := (X_S)_{H_{\rho(0)}^\beta} - \text{Tr}\left(X_S\omega_\beta\left(H_{\rho(0)}^\beta\right)\right)\mathbf{1}, \tag{11.42}$$

we then find that any solution X_S of the linear equation $L(X_S) = -L(V_S)$ provides a first-order solution to the condition in Lemma 11.2. We can thus conclude that Lemma 11.2 is satisfied up to first order by choosing

$$X(g) = H_{\rho(0)}^\beta + gV - g\text{Tr}_B\left(\omega_\beta(H_B)V\right) + H_B := H_{\rho(0)}^\beta + g\tilde{V} + H_B, \tag{11.43}$$

where we have introduced the *renormalized interaction*

$$\tilde{V} = V - V_S = V - \text{Tr}_B(\omega_\beta(H_B)V). \tag{11.44}$$

Plugging this solution back into the second order correction of the dissipation, we finally obtain

$$\Delta F^{(\text{irrev})}(g) = \frac{\beta g^2}{2}\text{cov}_{\omega_\beta(H_{\rho(0)}^\beta + H_B)}\left(\tilde{V}, \tilde{V}\right). \tag{11.45}$$

The solution is unique up to additive terms which fulfill $L(Y_S) = 0$. This implies

$$(Y_S)_{H^\beta_{\rho(0)}} = \mathrm{Tr}(\omega_\beta(H^\beta_{\rho(0)})Y)\mathbf{1}$$

and due to the explicit form of the generalized covariance such terms do not affect the value of the correction.

We can proceed similarly with the residual free energy $\Delta F^{(\mathrm{irrev})}(g)$. Due to Lemma 11.4, we again obtain

$$\Delta F^{(\mathrm{res})}(g) = \frac{\beta g^2}{2}\mathrm{cov}_{\omega_\beta(R(0))}\left(R'(0), R'(0)\right). \tag{11.46}$$

Remembering the form of $R(0)$, similar calculations as above show that the condition for a solution at first order in g is now given by $\tilde{L}(R_S) = \tilde{L}(V_S)$, with

$$\tilde{L}(R_S) := (R_S)_{H(0)_S} - \mathrm{Tr}\left(R_S\omega_\beta\left(H(0)_S\right)\right)\mathbf{1}. \tag{11.47}$$

Hence, the choice $R_S(g) = H(0)_S - g\mathrm{Tr}_B(\omega_\beta(H_B)V)$ solves the equation in Lemma 11.3 to first order in g. We then obtain

$$\Delta F^{(\mathrm{res})}(g) = \frac{\beta g^2}{2}\mathrm{cov}_{\omega_\beta(H(0)_S+H_B)}\left(\tilde{V}, \tilde{V}\right). \tag{11.48}$$

The solution is again unique up to terms which fulfill $(Y_S)_{H(0)_S} \propto \mathbf{1}$, which do not affect the correction.

We conclude that both corrections have the same form, the only difference being the state with respect to which the generalized covariance is evaluated. In both cases, the states are given by the optimal weak-coupling choices. Let us summarize these results.

Result 11.5 (Perturbative corrections). *To leading order in the coupling strength g, the perturbative corrections in the optimal work-extraction protocol from a system $(\rho(0), H(0)_S)$, which can be coupled via the interaction gV to with Hamiltonian H_B at inverse temperature β, are given by*

$$\Delta F^{(\mathrm{irrev})}(g) = \frac{\beta g^2}{2}\mathrm{cov}_{\omega_\beta(H^\beta_{\rho(0)}+H_B)}\left(\tilde{V}, \tilde{V}\right) + O(g^3), \tag{11.49}$$

$$\Delta F^{(\mathrm{res})}(g) = \frac{\beta g^2}{2}\mathrm{cov}_{\omega_\beta(H(0)_S+H_B)}\left(\tilde{V}, \tilde{V}\right) + O(g^3). \tag{11.50}$$

11.1.3 Interpretation of perturbative corrections

In the context of linear response theory, the generalized covariance is also known as the Kubo-Mori inner product [271–273] or Bogoliubov inner product [274]. Indeed as these names suggests, it is an inner product in the proper sense. Therefore the two correction terms are indeed positive also to leading order in the perturbative expansion. The connection to linear response theory also allows us to give physical interpretation to the correction terms. Indeed, the generalized covariances above are the *isothermal static admittance* to leading order in g and hence measure the difference of the expectation value of $\tilde{V}$ in the Hamiltonians $X(g)$ and $X(0)$ to first order in g (and similarly for $R(g)$). For example we have

$$\mathrm{Tr}\left(\left(\omega_\beta(X(g)) - \omega_\beta(X(0))\right)g\tilde{V}\right) = g^2\mathrm{cov}_{\omega_\beta(X(0))}(\tilde{V}, \tilde{V}) + O(g^3). \tag{11.51}$$

to leading order in g. Thus, the dissipation terms can be understood to simply measure the part of the work that is required to turn on or off the renormalized interaction which is dissipated into the environment – measured in units of the bath's temperature.

11.1.4 Heat

So far we have determined the strong-coupling corrections to the extracted work in an optimal protocol. Below we want to apply these results to the case of a thermal machine operating between two heat baths and find general corrections to the efficiency. As a prerequisite for this, let us now calculate that heat in an optimal work-extraction protocol. We can define the heat using energy-conservation whenever a quench on the system happens while the system is coupled to a heat bath. In a quench from $H_S^{(j)}$ to $H_S^{(j+1)}$, the heat that flows from the bath into the system is then simply given by the change of energy on the system due to the equilibration process under the new Hamiltonian $H^{(j+1)}$.

$$Q(j \to j+1) = \text{Tr}\left(\left(\omega_\beta(H^{(j+1)}) - \rho^{(j)}\right) H_S^{(j+1)}\right). \tag{11.52}$$

This quantity does in general not vanish since the local energy of S is not conserved in the equilibration process if the interaction between S and B is chosen such that it actually leads to equilibration and thermalization.

In the optimal work-extraction protocol, heat flow only occurs after the first coupling to the heat bath and during the isothermal part of the protocol. We therefore only need to track the heat until the interaction is turned after the isothermal reversible process. The final Hamiltonian of the relevant process is then given by $H^{(f)} = H_S^{(f)} + H_B$. In the optimal protocol, the Hamiltonian $H_S^{(f)}$ is given by $R(g)_S$ in the notation from the previous section. The process to consider therefore now starts in the non-interacting Hamiltonian $H(0)_S + H_B$ and ends in the non-interacting Hamiltonian $H_S^{(f)} + H_B$.

The total absorbed heat can then be calculated from energy-conservation as

$$Q = W - \Delta E_S, \tag{11.53}$$

where ΔE_S is the change of energy on the system. The work in this part of the protocol is given by

$$
\begin{aligned}
W &= F_\beta\left(\rho(0) \otimes \omega_\beta(H_B), H(0)_S + H_B\right) - F_\beta\left(\omega_\beta(H_S^{(f)} + V + H_B), H_S^{(f)} + V + H_B\right) \\
&\quad - \Delta F^{(\text{irrev})} + \text{Tr}\left(\omega_\beta(H_S^{(f)} + V + H_B)V\right) \\
&= F_\beta(\rho(0) \otimes \omega_\beta(H_B), H(0)_S + H_B) - F_\beta\left(\omega_\beta(H_S^{(f)} + V + H_B), H_S^{(f)} + H_B\right) \\
&\quad - \Delta F^{(\text{irrev})} \\
&= \Delta E_S + \Delta E_B - T\left(S(\rho(0)) + S(\omega_\beta(H_B)) - S(\omega_\beta(H_S^{(f)} + V + H_B))\right) \\
&\quad - \Delta F^{(\text{irrev})},
\end{aligned}
$$

where the change of energy on the system is given by

$$\Delta E_S = \text{Tr}(\rho(0)H(0)_S) - \text{Tr}\left(\omega_\beta(H_S^{(f)} + V + H_B)H_S^{(f)}\right). \tag{11.54}$$

Introducing the mutual information in the final state of the isothermal process and free energies on the bath, a small calculation shows that we can then write the total absorbed heat as

$$Q = T\Delta S_S - \Delta F_\beta(\omega_B^{(f)}, H_B) - TI(S:B)_{\omega_\beta(H_S^{(f)}+V+H_B)} - \Delta F^{(\text{irrev})}, \tag{11.55}$$

with $\omega_{B/S}^{(f)} = \text{Tr}_{B/S}\left(\omega_\beta(H_S^{(f)} + V + H_B)\right)$ the final state on the bath and system at the end of the isothermal process, respectively, and

$$\Delta S_S = S(\rho(0)) - S\left(\omega_S^{(f)}\right) \tag{11.56}$$

the change of entropy on the system.

The result (11.55) can easily be interpreted. First, we have the customary term $T\Delta S_S$ which relates the heat to the change of entropy on the system. However, the final entropy is not calculated from the Gibbs-state of the local Hamiltonian $H_S^{(f)}$ but from the reduced state of the global, interacting Hamiltonian. Apart from the first term, we have three correction terms:

1. The term $\Delta F_\beta(\omega_B^{(f)}, H_B)$ measures how far the final state of the *bath* is out of equilibrium with respect to its initial equilibrium state.

2. The term $I(S : B)_{\omega_\beta(H_S^{(f)}+V+H_B)}$ measures the amount of correlations that have been build up in the isothermal process.

3. The term $\Delta F^{(\text{irrev})}$ is the same as in the expression for work and measures the amount of dissipation when the system is brought in contact with the bath.

Let us briefly discuss the general properties of the correction terms. First, all the correction terms are positive since they are proportional to relative entropies. Since all the correction terms are positive, the heat fulfills the Clausius inequality

$$Q \le T\Delta S_S. \tag{11.57}$$

Second, the correction terms are all bounded by $2\,\|V\|$ and hence are negligible in the macroscopic limit (cf. section 11.1.1): The two free energy terms can be bounded by $2\,\|V\|$ by the same techniques as the correction terms for the work. The mutual information can bounded by $2\,\|V\|$ using the following general result about area laws in Gibbs states.

Lemma 11.6 (Area Law in Gibbs states[275, 276])**.** *Consider a bipartite system SB with Hamiltonian $H = H_S + H_B + V$, with local Hamiltonians H_S and H_B and interaction V. Then*

$$I(S : B)_{\omega_\beta(H_S+V+H_B)} \le 2\,\|V\|. \tag{11.58}$$

Proof. This follows from the extremality of Gibbs-states, by considering the "reversed" mutual information:

$$
\begin{aligned}
0 &\le D\left(\omega_\beta(H)_S \otimes \omega_\beta(H)_B \| \omega_\beta(H)\right) \\
&= -S_S - S_B - \text{Tr}\left(\omega_\beta(H)_S \otimes \omega_\beta(H)_B \log(\omega_\beta(H))\right) \\
&= -S_S - S_B + \beta\left(\text{Tr}(\omega_\beta(H)_S \otimes \omega_\beta(H)_B H) + \log(Z_\beta(H))\right) \\
&= -I(S : B)_{\omega_\beta(H)} + \beta\text{Tr}\left((\omega_\beta(H)_S \otimes \omega_\beta(H)_B - \omega_\beta(H))\, V\right),
\end{aligned}
$$

where the last line follows by explicit calculation of the mutual information. The triangle inequality then yields the claim. $\qquad\square$

Finally, all the correction terms vanish exactly in the weak coupling limit. Since they are all smooth functions if we consider finite dimensional systems, we can already deduce that they are of second order in the coupling strength.

We will now use these results on work extraction and the absorbed heat to derive bounds on the efficiencies of cyclically running thermal machines in the weak coupling regime.

11.2 Efficiency: Thermal machines in the strong coupling regime

We will now derive general bounds on the efficiency of a thermal machine running between two heat baths at inverse temperatures β_c and β_h. We will make the following two assumptions: i) the system can only brought in contact with one of the baths at a given point in time and ii) while the system is in contact with one bath, it looses all correlations with the second bath. The first assumption is simply for technical convenience. We could simulate a situation where a system is in contact with both baths at once by iterating between the two baths in many small steps. The second assumption can be justified by noting that in a given step the system only becomes correlated with a small part of the bath. In the time when the

system interacts with the other bath, these correlations get distributed throughout the first bath if it is sufficiently interacting. Since we anyway assume that the bath is interacting, equilibrating and thermalizing, we therefore do not make a strong additional assumption.

As in the case of chapter 9, where we considered restrictions on the possible Hamiltonians in the weak coupling regime, optimal protocols for such machines consists of four steps:

1. An isothermal process in contact with the hot bath,

2. an adiabatic quench while disconnected from the bath,

3. an isothermal process in contact with the cold bath,

4. an adiabatic quench back to the initial Hamiltonian.

Such protocols optimize the efficiency since they minimize the number of dissipation events by only coupling and decoupling with each bath once. A formal proof of this statement can be done using the same arguments as for the case in chapter 9 (see section 14.6.6).

Let us now calculate the efficiency of such protocols. To that, let us assume that our protocol starts after we have decoupled from the cold bath and denote the final Hamiltonian on S at the end of the isothermal with the cold bath as $H_S^{(f),c}$. We will call the corresponding initial Hamiltonian $H_S^{(i),c}$ and similarly for the hot bath ($H_S^{(i/f),h}$). In the following we will also denote the Hamiltonians and interaction terms of the two baths by $H_B^{h/c}$ and $V^{h/c}$, respectively. Note that once $H_B^{h/c}$ and $V^{h/c}$ are fixed, the efficiency in the optimal protocol is fixed as a function of the Hamiltonians $H_S^{(f),h/c}$. For notational simplicity, let us use the notation

$$H_{SB}^{(f/i),c/h} = H_S^{(i/f),c/h} + gV^{c/h} + H_B^{c/h} \tag{11.59}$$

in the following.

The total cycle of the thermal machine can be decomposed into two parts, each of which has the form considered in the previous section: A first part using the hot bath, with initial Hamiltonian $H_S^{(f),c}$ and final Hamiltonian $H_S^{(f),h}$. The initial state of this part is given by

$$\rho_S^{(c)} = \text{Tr}_B\left(\omega_{\beta_c}(H_{SB}^{(f),c})\right). \tag{11.60}$$

After this part, we have a second part using the cold bath back to Hamiltonian $H_S^{(f),c}$ with initial Hamiltonian $H_S^{(f),h}$ and initial state

$$\rho_S^{(h)} = \text{Tr}_B\left(\omega_{\beta_h}(H_{SB}^{(f),h})\right). \tag{11.61}$$

In the first part, a positive amount of heat Q_h is absorbed from the hot bath and in the second part the heat Q_c is exchanged with the cold bath. From energy-conservation we obtain that the efficiency is given by

$$\eta = \frac{W}{Q_h} = \frac{Q_h + Q_c}{Q_h} = 1 + \frac{Q_c}{Q_h}.$$

Using (11.55), the quantities Q_c and Q_h are given by

$$Q_h = T_h \Delta S - \Delta F_{\beta_h}(\omega_B^{(f),h}, H_B^h) - T_h I(S:B)_{\omega_{\beta_h}(H_{SB}^{(f),h})} - \Delta F^{(\text{irrev}),h},$$

$$Q_c = -T_c \Delta S - \Delta F_{\beta_c}(\omega_B^{(f),c}, H_B^c) - T_c I(S:B)_{\omega_{\beta_c}(H_{SB}^{(f),c})} - \Delta F^{(\text{irrev}),c},$$

with

$$\Delta S = S(\rho_S^{(c)}) - S(\rho_S^{(h)}) \tag{11.62}$$

and

$$\omega_{S/B}^{(f),c/h} = \mathrm{Tr}_{B/S}\left(\omega_{\beta_{c/h}}(H_{SB}^{(f),c/h})\right). \tag{11.63}$$

From the formulas, it is obvious that a positive amount of heat Q_h implies $\Delta S \geq 0$ and $Q_c < 0$ (and vice versa). Putting all the formulas together, we obtain the following result for the efficiency.

Result 11.7 (Efficiency in the strong coupling regime). *Given bath Hamiltonians $H_B^{c/h}$ and interactions $V^{c/h}$, respectively, the optimal efficiency is given by*

$$\eta\left(H_S^{(f),h}, H_S^{(f),c}\right) = 1 - \frac{T_c \Delta S + C_c}{T_h \Delta S - C_h} \leq 1 - \frac{T_c}{T_h}. \tag{11.64}$$

The positive semi-definite correction terms $C_{c,h}$ are given by:

$$C_{c/h} = \Delta F_{\beta_{c/h}}\left(\mathrm{Tr}_S\left(\omega_{\beta_{c/h}}(H_{SB}^{(f),c/h})\right), H_{SB}^{(f),c/h}\right) + T_{c/h}I(S:B)_{\omega_{\beta_{c/h}}\left(H_{SB}^{(f),c/h}\right)}$$
$$+ \min_{H_S^{(i),c/h}} \Delta F\left(\omega_{\beta_{c/h}}\left(H_S^{(f),h/c} + H_B\right), H_{SB}^{(i),c/h}\right), \tag{11.65}$$

and the change of entropy by

$$\Delta S = S\left(\omega_S^{(f),c}\right) - S\left(\omega_S^{(f),h}\right). $$

We have arrived at the full, non-perturbative expression for the efficiency of an optimal thermal machine operating between the given set of Hamiltonians and temperatures. Similarly to the work-extraction setting, let us briefly discuss its behaviour in dependence of the coupling strength g.

11.2.1 Dependence of efficiency on coupling strength

To discuss the dependence of the efficiency in (11.64), let us assume that the machine extracts a finite amount of work per cycle, so that $\Delta S > 0$. We can then re-write the efficiency as

$$\eta(g) = 1 - \frac{T_c}{T_h}\frac{1 + x_c(g)}{1 - x_h(g)}, \tag{11.66}$$

with $x_{c/h}(g) = C_{c/h}(g)/(T_{c/h}\Delta S(g))$, where we have already inserted the explicit g-dependence of all the quantities. In particular, the change of entropy ΔS depends on the coupling strength. In the limit of $g \to 0$, it is given by

$$\Delta S^{(\mathrm{weak})} = S\left(\omega_{\beta_c}(H_S^{(f),c})\right) - S\left(\omega_{\beta_h}(H_S^{(f),h})\right), \tag{11.67}$$

as expected from the weakly coupled engines in chapter 9. For small but finite g, we can expand ΔS as

$$\Delta S(g) = \Delta S^{(\mathrm{weak})} + K^{(1)}g + K^{(2)}g^2 + O(g^3). \tag{11.68}$$

We have already remarked in section 11.1.4 that the correction terms $C_{c/h}$ are of order g^2,

$$C_{c/h}(g) = \frac{1}{2}C_{c/h}^{(2)}g^2 + O(g^3). \tag{11.69}$$

In particular, $C_{c/h}^{(2)}$ can be lower bounded by the contribution stemming from the dissipation term $\Delta F^{(\mathrm{irrev})}$, which we explicitly calculated in section 11.1.2. Since $C_{c/h}(g)$ are of second order in g with $C_{c/h}(0) = 0$, the first derivative of $x_{c/h}(g)$ at $g = 0$ vanishes.

Therefore, the corrections $x_{c/h}(g)$ are also of second order with $x_{c/h}(0) = 0$. We then obtain

$$
\begin{aligned}
\frac{1 + x_c(g)}{1 - x_h(g)} &= 1 + \frac{1}{2}(x_c''(0) + x_h''(0))g^2 + O(g^3) \\
&= 1 + \frac{1}{2}\left(\frac{C_c^{(2)}}{Q_c^{(\text{weak})}} + \frac{C_h^{(2)}}{Q_h^{(\text{weak})}}\right)g^2 + O(g^3),
\end{aligned}
\qquad (11.70)
$$

where $Q_{c/h}^{(\text{weak})} = T_{c/h}|\Delta S^{(\text{weak})}|$ is absolute value of the heat exchanged with the respective bath in the weak-coupling limit. It is somewhat surprising that the corrections do not depend on the first-order correction $K^{(1)}$ on the change of entropy on the system. This has the advantage that, to leading order in the coupling strength, the correction terms can be minimized almost independently of the term ΔS as long as one ensures that this term remains positive and finite.

To summarize, we have derived that the Carnot-efficiency remains as the theoretical optimal efficiency of any thermal machine, even in the strong coupling regime. For machines that are coupled to the bath with a finite, but small coupling strength, the corrections to the Carnot efficiency are of second order in g and we can explicitly calculate a lower bound to these corrections using the results in section 11.1.2.

11.3   Power

We have seen that a finite coupling strength is detrimental to work-extraction for any thermal machine, because it leads to dissipation and hence irreversibility. Nevertheless, a finite coupling strength is of course necessary for any thermal machine since the thermal machine has to be able to thermalize with a heat bath to be able to act as a thermal machine at all. For practical purposes, one is furthermore often not so much interested in the efficiency of a thermal machine, but in the power P of the machine, which we can define as the ratio of the extracted work per cycle $W^{(\text{cycle})}$ divided by the time $t^{(\text{cycle})}$ that it takes to run the machine for one cycle. Both quantities depend on g in a non-trivial way. Nevertheless, the optimal power is achieved for some finite value of g. First, the cycle-time $t^{(\text{cycle})}$ has to scale at least like $1/g$ since the maximal rate of energy-exchange of the system with the baths is proportional to g. In particular, as $g \to 0$, the cycle-time has to diverge. Therefore

$$
P(g) \to 0, \quad \text{as} \quad g \to 0. \qquad (11.71)
$$

Secondly, in the limit $g \to \infty$, we can argue that the extracted work becomes at most zero. To see this, consider the correction term $\Delta F^{(\text{irrev})}$, given by

$$
\Delta F^{(\text{irrev})}(g) = \sup_{H_S^{(1)}} \frac{1}{\beta} D\left(\rho(0) \otimes \omega_\beta(H_B) \| \omega_\beta(H_S^{(1)} + gV + H_B)\right). \qquad (11.72)
$$

Let P denote the (projector onto the) ground state subspace of the interaction V. Then in the limit $g \to \infty$ we have

$$
\lim_{g \to \infty} \omega_\beta(H_S^{(1)} + gV + H_B) = \omega_\beta((H_S^{(1)} + H_B)_P) \otimes 0, \qquad (11.73)
$$

where $(H_S^{(1)} + H_B)_P$ denotes the restriction of $H_S^{(1)} + H_B$ onto P and the direct sum is over P and its complement. Therefore, the support of $\rho(0) \otimes \omega_\beta(H_B)$ is not contained in the support of $\lim_g \omega_\beta(H_S^{(1)} + gV + H_B)$. But as we have seen in chapters 2 and 3, the relative entropy diverges in this case. Therefore, the dissipation $\Delta F^{(irrev)}(g)$ diverges as well as $g \to \infty$ and hence the optimal work per cycle goes to zero. Hence the power goes to zero as well:

$$
P(g) \to 0, \quad \text{as} \quad g \to \pm\infty. \qquad (11.74)
$$

Since we expect that finite power can be achieved for some value of g, we conclude that there is some optimal value of g which maximizes power. It should be clear, however, that we cannot expect to derive the general form $P(g)$ without specific assumptions on the systems involved, since this would require solving the problem of equilibration time-scales first (see chapter 7).

We can, however, give a heuristic discussion for the behaviour for small g. First, since the cycle time is at least of the order $1/g$ (this will be re-fined below) and the work is of the form $W(0) - Kg^2$, we find that the power fulfills

$$P(g) \leq \alpha|g| - O(g^3), \tag{11.75}$$

for some non-universal constant $\alpha \geq 0$. We can refine this bound by observing that the dominant contribution to the cycle-time has to come from the equilibration process after each quench. There is a simple heuristic argument, which suggests that the equilibration time should scale as g^2 instead of g from the point of view of perturbation theory: If the bath is large and its spectral density sufficiently smooth, there are many more second-order processes that can contribute to the transfer of energy from the system to the bath than there are first order processes, since these have to be resonant. For example, for a bath of harmonic oscillators, a process that changes the energy of the system by some amount has to couple resonantly to a single oscillator of the bath with the corresponding eigenfrequency. On the other hand, at second order two oscillators of the bath can be combined to exchange the same amount of energy with the system if their frequencies add up correspondingly. Hence, we can expect that processes which are at least of second order in g dominate the equilibration process and lead to an equilibration time that scales as $1/g^2$. This is consistent with the second-order expansion that one employs in the theory of open systems to derive Markovian, equilibrating system dynamics in the weak-coupling regime [81]. The assumption that the equilibration time scales as $1/g^2$ would lead to an expansion of the power of the form

$$P(g) \sim \alpha|g|^2 - O(g^3). \tag{11.76}$$

We will now present an example which shows exactly this behaviour.

11.4   Example: Quantum Brownian Motion

Let us finally discuss a simple example to illustrate our findings. We will consider *Quantum Brownian Motion* in the *Caldeira-Leggett-* or *Ullersma-model* [64, 65]. It consists of a single central harmonic oscillator which is linearly coupled to a bath of n harmonic oscillators. The spectral density of the bath is assumed to be Ohmic with some cut-off $\omega_{\max}$. The total Hamiltonian takes the form

$$H(0) = H(0)_S + gV(g) + H_B + H_L, \tag{11.77}$$

with

$$H(0)_S = \frac{1}{2}m\omega^2 x^2 + \frac{p^2}{2m}, \tag{11.78}$$

$$H_B = \sum_k \frac{1}{2}m_k\omega_k^2 x_k^2 + \frac{p_k^2}{2m_k}, \tag{11.79}$$

$$V(g) = x\sum_k g_k x_k + H_L, \tag{11.80}$$

$$H_L = gx^2 \sum_k \frac{g_k^2}{m_k\omega_k^2}, \tag{11.81}$$

where the *Lamb-shift* H_L renormalizes the system's oscillator frequency in presence of the interactions to the bath. It ensures that the Hamiltonian remains bounded from below and that the system thermalizes to the Hamiltonian $H(0)_S$ in the Markovian dynamics that
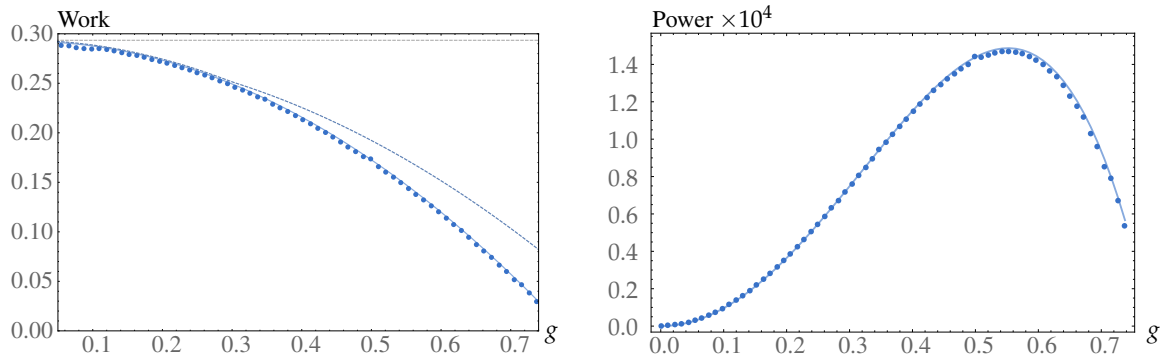
Figure 11.1: *Left figure:* The extracted work is plotted versus interaction strength g. The blue dots show the exact extracted work by computing the unitary evolution of SB for a protocol which becomes optimal in the weak coupling regime. The solid, light blue line shows the corresponding predictions using our framework. The dashed blue line shows the maximum extractable work as a function of g. The dashed grey line shows the maximum extractable work in the weak coupling limit $W^{(\text{weak})}$. Parameters are given by $\omega = 1$, $m_k = m = 1$, $\omega_{\max} = 1.2$ and $\beta = 3.5, \beta_S = 1$. The bath consists of $n = 165$ oscillators. The protocol consists of 200 quenches, with a waiting time $10/g^2$ when computing the unitary dynamics. For values of g smaller than the ones shown larger bath sizes are required for proper thermalization of the system and lead to larger errors for the bath considered here. *Right figure:* Plot of power versus interaction strength. The blue dots again show the exact results obtained from unitary evolution and the solid, light blue line the effective description using our framework. The parameters are the same as in the left figure.

results from taking the limit of weak interactions if the bath is thermal. The ohmic spectral density means that the frequencies ω_k and couplings g_k take the values

$$\omega_k = \frac{k}{n}\omega_{\max}, \quad g_k = \omega_k\sqrt{\frac{2\omega_{\max}}{\pi n}}. \tag{11.82}$$

This choice ensures that the spectral density

$$J(\omega) := \frac{2}{\pi}\sum_k \frac{g_k^2}{\omega_k}\delta(\omega - \omega_k) \tag{11.83}$$

approaches a linear function $J(\omega) \propto \omega$ with cut-off $\omega_{\max}$ in the continuum limit $N \to \infty$. This model is an important in the field of open quantum systems [277], in particular because it is both exactly solvable and efficiently simulable on a computer, because it is quadratic in the bosonic annihilation and creation operators (see, e.g., Refs. [278, 279]). Therefore it is also a good test-bed in quantum thermodynamics and has found numerous applications there-in (see, e.g., Refs. [256, 258, 261, 267, 269, 270, 280–282]).

Importantly, this model is known to have the property that the state on S equilibrates to the reduced state of the thermal state of the full system in the limit of large baths independently of the coupling strength and hence fulfills our basic assumption of thermalization after a single quench [283, 284]. Note however, that this does not automatically imply that the effective description in terms of iterated Gibbs-states is valid. Furthermore, the interaction is not of the form gV, due to the presence of the Lamb-shift. It formally hence does not fit our framework developed in the previous sections. It can easily be seen from the derivations, however, that to leading order in g, this latter discrepancy does not make a difference. In particular, to leading order the optimal Hamiltonian local Hamiltonians in work-extraction protocols are again determined by the renormalized interaction $\tilde{V}$, which is now given by

$$\tilde{V} = x\sum_k g_k x_k - x\text{Tr}\left(\omega_\beta(H_B)\sum_k g_k x_k\right) = x\sum_k g_k x_k, \tag{11.84}$$

which follows from the reflection-symmetry of the harmonic-oscillator. This result implies that, to leading order in g, the optimal work-extraction protocols are *identical to the weak-coupling protocols.*

Let us now discuss work-extraction protocols in the Caldeira-Leggett model. We assume an initial state $\omega_{\beta_S}(H_S) \otimes \omega_\beta(H_B)$ with $\beta_S \neq \beta$. In the weak coupling limit, the optimal work-extraction protocol first quenches the central oscillator by setting $\omega \mapsto \beta/\beta_S\,\omega$ and

$m \mapsto \beta_S/\beta \, m$. Note that often harmonic oscillators appear as effective desciptions in which the parameter m does not correspond to the actual, physical mass of some particle and can be modified. Not being able to change the parameter m will naturally result in protocols that are not optimal. This first quench is then followed by an isothermal process back to the initial Hamiltonian. This is nicely illustrated in Fig. 11.1, where we calculated numerically the results obtained from exact unitary time-evolution and the predictions in our model, showing good agreement. In particular, the extracted work decreases as g^2 with the coupling strength, as expected from our general results.

Turning to the power of this protocol, we then consider the protocol with a fixed number of steps N and study the power as a function of g. We numerically found that the equilibration time in this model is proportional to $1/g^2$ (for $g < 1$), in agreement with our considerations in the previous sections. Hence, we expect that the power shows a behaviour of the form $P(g) \propto W^{(\text{weak})} g^2 - O(g^3)$. This behaviour can indeed be seen in Fig. 11.1 as expected. The example thus shows good agreement with our predictions.

11.5 Summary

Standard thermodynamic bounds usually hold for systems that are weakly coupled. This weak coupling behaviour can be well justified for macroscopic systems which interact locally, due to the vanishing surface-to-volume ratio. For truly small systems this argument does not hold and we hence have to study more specifically in how far thermodynamic bounds remain valid or have to be corrected. In this chapter, we have derived fully general expressions for the strong-coupling corrections of a thermal machine coupled to a heat bath. To arrive at our results, we employed the framework developed in the previous chapter. We have explicitly shown that the corrections to the weak coupling bounds become irrelevant for macropscopic systems. Our expressions are completely general and only rely on the assumption that the systems in questions actually thermalize in the sense of closed systems when coupled to a heat bath. Interestingly, the correction terms can be expressed succinctly in terms of relative entropies, and essentially measure irreversible dissipation in terms of correlations which are caused from the interaction between the system and bath.

In the case of weak, but finite coupling strength, we derived the explicit leading corrections to work-extraction bounds and the efficiency of optimal thermal machines as a function of the system and bath Hamiltonians and their interaction operator. The leading terms are at second order in the interaction strength and take a fairly simple form. The fact that the corrections are of second order in the coupling strength may not come as a surprise, since it can be argued from general grounds: Assuming that the universal weak-coupling bounds also apply for strongly coupled machines (which is necessary for thermodynamic consistency and matches our results) and that optimal work-extraction bounds vary smoothly in g already implies that corrections have to be at second order without making any calculation. The merit of the results in this chapter is thus that they provide the explicit functions which can in principle be evaluated for any model.

The results in this chapter show that, in terms of efficiency of thermal machines, finite coupling strength are detrimental when compared to the weak-coupling limit. We also discussed that our results imply that the power vanishes in the limit of arbitrarily strong interactions and weak interactions. Hence optimal power is given by a finite coupling strength as expected.

All our results can be demonstrated in the paradigmatic example of Quantum Brownian motion, which is an integrable (exactly solvable) model. It would be interesting to study the explicit corrections for models which are fully interacting. In this chapter, we had to assume that the thermal machine never interacts with both heat baths at the same time. An interesting avenue for further research would be to study autonomous thermal machines in the strong coupling regime, which are simultaneously coupled to both thermal baths and find out whether similar bounds can be derived.

12 On spontaneous symmetry breaking in dissipative systems

In the previous parts of this thesis, I discussed in great detail the thermodynamics of (potentially small) quantum systems that can be manipulated by acting unitarily and can be brought in contact with heat baths. The treatment was in parts somewhat unorthodox, in particular in the strong emphasis of resource theoretic formulations of thermodynamics and the close connection drawn to results about equilibration in *closed* many body systems.

IN THIS LAST CHAPTER OF THIS THESIS[1], I will come back to a classic question in statistical mechanics, namely the emergence of continuous phase transitions in large many-body systems; more precisely, I will discuss the dynamical emergence of spontaneous symmetry breaking in lattice systems. Unlike in previous chapters, I will not assume that the system is a closed system, but want to study the dynamical effects of dissipation on the system.

Traditional treatments of spontaneous symmetry breaking in thermal phase-transitions mostly work on a *kinematic* level. They show that in certain systems with internal symmetries, such as spin-flip symmetry, the Gibbs state at high temperatures is unique and symmetric in the thermodynamic limit [206], but there exists a critical temperature below which thermal states are not unique in the thermodynamic limit. The different (extremal) thermal states can usually be distinguished by a local order parameter, such as the magnetization density in a certain direction. These states thus break the symmetry of the Hamiltonian.

In practice, to show this one usually introduces an arbitrarily small symmetry breaking field of strength h and shows that a finite magnetization density persists if one first takes the thermodynamic limit and then sends the field strength h to zero. The direction of the corresponding magnetization density then depends on the direction of the small symmetry breaking field. A similar treatment is of course also possible if the order parameter is not given by the magnetization. For example, in an anti-ferromagnet the order-parameter could be the staggered magnetization.

In finite systems, the thermal state is always unique. Thus, from a formal perspective, phase transitions as considered above cannot exist in finite system. However, it is well known that a phase transition in the thermodynamic limit can be associated to finite fluctuations in the *density* of an order parameter in large, but finite system. In particular, the value that an order parameter takes in the symmetry-broken phases that are constructed by introducing infinitesimal symmetry-breaking fields can be lower bounded by the magnitude of fluctuations in systems of large but finite volumes in which no symmetry-breaking field has been applied [285–287]. As a simple example consider the two-dimensional ferromagnetic Ising model at extremely low temperature. The Gibbs state of this model is essentially

an equal weight mixture of the state where all spins point up and the state where all spins point down. This Gibbs state evidently has finite fluctuations in the magnetization density and infinite correlation length. The individual symmetry-broken phases, on the other hand, have finite magnetization density, arbitrarily small fluctuations in the magnetization density and a finite correlation length.

In this chapter, I will connect such finite density-fluctuations with the *dynamical emergence* of symmetry-breaking states in lattice models, which are described by dissipative dynamics. Roughly speaking, I will demonstrate that if local, purely dissipative dynamics (which can, for example, result from a weak coupling to a heat bath) in detailed balance has a symmetric state with finite fluctuations in the density of an order parameter as a steady-state, then it necessarily also has states as approximate steady-state that explicitly break the symmetry associated to the order parameter if the system size becomes big. The survival time of these approximate steady states diverges with the system size. In the thermodynamic limit, they therefore behave exactly as steady states.

To remain in the example of the two-dimensional Ising model, any detailed balance Markovian stochastic dynamics that has the Gibbs state at low temperature as steady state, also has two symmetry-broken states where all the spins point in a single direction as approximate steady states.

Suppose now that a system is initialized in a state with very few correlations. Then to dynamically build up the correlations necessary to have finite fluctuations in an order parameter at least takes a time that grows linearly with the system size (due to Lieb-Robinson bounds) as long as the interactions are local. This is true also for dissipative dynamics [223, 288–290]. It therefore seems likely that the dissipative dynamics first quickly brings the system to a state that explicitly breaks the symmetry but has only short-range correlations. Once the system reaches such a state, it will remain in this state for a time that diverges with the system size. This later phenomenon of diverging equilibration time-scales will be shown explicitly using Lieb-Robinson bounds. The results in this chapter therefore explicitly show a phenomenon of critical slowing down in connection with spontaneous symmetry breaking and add to the recent discussions about equilibration time-scales and the closing of the dissipative gap in open lattice models [291–296].

In other words, the results can also be interpreted by saying that spontaneous symmetry breaking is in a sense stable to any dissipative dynamics that prepares a symmetric steady-states with finite fluctuations in the density of an order parameter. Nevertheless, the results do of course not explain, which one of the symmetry braking states appears as a steady state.

12.1   Formal set-up

In the following, we consider sequences of systems defined on finite lattices $\Lambda \subset \mathbb{Z}^D$ of increasing volume L^D. For simplicity, I will only discuss square lattices, but any regular lattice will do. To every site x of the lattice we associate the local Hilbert-space $\mathscr{H}_x$, so that the total Hilbert-space is given by $\otimes_{x \in \Lambda} \mathscr{H}_x$. We thus think in terms of spins attached to the lattice sites. All the results can, however, also be straight-forwardly transferred to the setting of fermionic lattice systems.

A special role will be played by the magnetization in z direction on a region $X \subseteq \Lambda$, which is defined as

$$S_X^z = \sum_{x \in X} S_{\{x\}}^z, \tag{12.1}$$

where $S_{\{x\}}^z$ is the Spin-operator in z direction on site x. For a system of spin-1/2 we therefore have $S_{\{x\}}^z = \sigma_x^z / 2$. The *magnetization density* in a region X is simply the average value of the magnetization over the region $S_X / |X|$, where $|X|$ denotes the number of sites in region X. In general, I will call any operator that is a sum of local operators, each supported around a lattice site, an *extensive operator* and the corresponding average over a region the *density* of the corresponding operator.

As emphasized in the introduction, in this chapter we are interested not in unitary dynamics generated by a Hamiltonian as in the previous chapters of this thesis, but in Marko-

vian dynamics in general. This allows for dissipation effects which are due to a coupling to an environment (for example a heat bath or the coupling to a weak continuous measurement device). We will assume that the dynamics is *local*, i.e., assume that it is given by the sum of many terms acting on a small region of the system. Furthermore, in this chapter we will work in the Heisenberg-picture, i.e., consider the dynamics on the observables instead of the states.

The conditions of Markovianity and locality then forces the dynamics to be of the form[2]

$$A(t) = e^{t\mathcal{L}^\Lambda}[A], \quad \mathcal{L}^\Lambda = \sum_{x \in \Lambda} \mathcal{L}_x^\Lambda. \tag{12.2}$$

where square brackets denote the action of a super-operator and each *Liouvillian* $\mathcal{L}_x^\Lambda$ acts on observables as [83, 84]

$$\mathcal{L}_x^\Lambda[A] = i[H_x, A] + \sum_j 2\left((L_x^j)^\dagger A L_x^j - \left\{(L_x^j)^\dagger L_x^j, A\right\}\right), \tag{12.3}$$

with L_x^j being the so-called *Lindblad operators*. Throughout this chapter, I will for simplicity assume that the operators L_x^j and H_x are strictly local. This means that they only act non-trivially on sites within a Ball $B_r(x)$ of radius r centered on x, where the radius is measured in terms of the lattice-distance. However, the general proofs, which appear in section 14.9, are formulated for more general approximately local Liouvillians.

I will furthermore assume periodic boundary conditions and that the dynamics is uniformly bounded, i.e., there is a constant b such that $\left\|\mathcal{L}_x^\Lambda[A]\right\| \leq b\|A\|$ for all Λ and $x \in \Lambda$.

In the former chapters of this thesis, we have always formalized quantum states as density matrices. In this chapter, I will stick to the usual formulation in mathematical physics, where states are understood as positive, linear functions on the algebra observables [297]. A state ω on an algebra of observables $\mathcal{A}$ is then simply any linear functional, which fulfills the following two conditions:

1. *Positivity:* $\omega(A^\dagger A) \geq 0$ for all operators $A \in \mathcal{A}$,

2. *Normalization:* $\omega(\mathbf{1}) = 1$ for the identity operator $\mathbf{1} \in \mathcal{A}$.

Working with states in this Heisenberg picture will turn out to be more appropriate when we consider the dynamics in the thermodynamic limit. The reason is that we cannot easily construct a Hilbert-space in the thermodynamic limit. However, we can define the *algebra of local observables* $\mathcal{A}$ as the closure of the set of observables that can be approximated arbitrarily well by local observables [297]. A state in the thermodynamic limit is then simply a state on the algebra of local observables in the above sense. Note that on a finite system, any operator is part of the algebra of local observables. In particular, the state ω can then be represented by a density matrix ρ^ω in the usual way. Its matrix elements in some basis $|i\rangle$ are given by

$$\rho_{i,j}^\omega = \omega(|i\rangle\langle j|). \tag{12.4}$$

A particularly important kind of states are *steady states* or *stationary states* with respect to $\mathcal{L}^\Lambda$. These are states that give rise to static expectation values:

$$\omega(\mathcal{L}^\Lambda[A]) = 0, \tag{12.5}$$

for any observable A supported in Λ. Steady states play a similar role in dissipative dynamics as do ground states in closed systems. The key difference is that in dissipative systems, the system evolves towards a steady state as long as the latter is unique. In particular, if we consider a finite system with a unique steady state, then any observable approaches the operator $\omega(A)\mathbf{1}$ in the infinite-time limit.

The fact that we assume that the Liouvillian $\mathcal{L}^\Lambda$ is local ensures that the dynamics induced on (quasi-)local observables is well defined in the thermodynamic limit as

$$A(t) = \lim_{\Lambda \nearrow \mathbb{Z}^D} e^{t\mathcal{L}^\Lambda}[A] \in \mathcal{A}. \tag{12.6}$$

The proof of this statement follows from Lieb-Robinson bounds [223, 288–290]. A steady state in the thermodynamic limit then simply fulfills $\omega(A(t)) = \omega(A)$.

In the following, a particular role will be played by sequences of states that are not steady states on any finite system, but converge to steady states in the thermodynamic limit. More formally, they are defined in the following way.

Definition 12.1 (Asymptotically stationary states). *A sequence of states ω_Λ (one for each volume Λ) is called* asymptotically stationary *if it satisfies*

$$\lim_{\Lambda \nearrow \mathbb{Z}^d} \omega_\Lambda(\mathcal{L}^\Lambda[A]) = 0 \tag{12.7}$$

for all local operators A.

From the definition it follows that asymptotically stationary states define a steady state in the thermodynamic limit. In other words, asymptotic stationarity means that the time it takes to reach the actual steady state in a finite system diverges with the system size.

12.1.1 Detailed balance

In the following, I want to discuss dynamics that is at equilibrium. I will therefore assume that the system fulfills detailed balance. Furthermore, I will only consider the non-unitary part of the time-evolution since I do not want to make specific assumptions about the Hamiltonian. For such purely dissipative quantum dynamics on a finite system, we will say that a state ω_Λ is in detailed balance (or reversible) with respect to the Liouvillian $\mathcal{L}^\Lambda$ if [84, 298–303]

$$\omega_\Lambda(A\mathcal{L}^\Lambda[B]) = \omega_\Lambda(\mathcal{L}^\Lambda[A]B), \tag{12.8}$$

where A, B are any two local observables. In a finite system, if $\rho^\omega = \sum_j p_j |j\rangle\langle j|$ is the spectral decomposition of the steady-state, then the transition-probabilities $p(i \mid j, \tau)$ to go from state $|j\rangle$ to state $|i\rangle$ in time τ fulfill

$$p_i p(j \mid i, \tau) = p_j p(i \mid j, \tau). \tag{12.9}$$

We thus recover the classical definition of detailed balance in the eigenbasis of the steady-state. To see this, first note that

$$p(i \mid j, \tau) = \mathrm{Tr}\left(|i\rangle\langle i| \, e^{\tau\mathcal{L}^\Lambda}[|j\rangle\langle j|]\right).$$

Inserting this into the definition of detailed balance and using the eigen-decomposition of the steady-state, we then obtain

$$p_i p(j \mid i, \tau) = \omega_\Lambda\left(|i\rangle\langle i| \, e^{\tau\mathcal{L}^\Lambda}[|j\rangle\langle j|]\right) \tag{12.10}$$

$$= \omega_\Lambda\left(e^{\tau\mathcal{L}^\Lambda}[|i\rangle\langle i|] \, |j\rangle\langle j|\right) = p_j p(i \mid j, \tau).$$

Since any Liouvillian fulfills $\mathcal{L}^\Lambda[\mathbf{1}] = 0$, a state that is in detailed balance is automatically stationary. The property of being in detailed balance is quite common in the literature. In particular Liouvillians that are derived by a standard weak coupling limit to a thermal heat baths fulfill this property with the thermal state as steady state (see, for example, [84, 299, 303]). Nevertheless, it is a long-standing open problem to derive local Liouvillians in detailed balance that have the Gibbs state of a local Hamiltonian as steady state. While it is solved for local, commuting Hamiltonians [303], it is unsolved for non-commuting Hamiltonians in general. When we assume both locality and detailed balance, we are thus making a non-trivial assumption in general.

Similarly to asymptotically stationary states, we can then also define asymptotically reversible states as those sequences of states that become reversible in the thermodynamic limit.

Definition 12.2 (Asymptotically reversible states). *Let $\mathcal{L}^\Lambda$ be a sequence of Liouvillians and ω_Λ a sequence of states. We call ω_Λ asymptotically reversible if*

$$\lim_{\Lambda \nearrow \mathbb{Z}^d} \left(\omega_\Lambda(\mathcal{L}^\Lambda\,[A]\,B) - \omega_\Lambda(A\mathcal{L}^\Lambda\,[B]) \right) = 0 \qquad (12.11)$$

for any two local operators A, B.

Similarly as reversibility implies stationarity, asymptotic reversibility implies asymptotic stationarity. The main result of this chapter will be that finite density fluctuations of an order parameter in a symmetric steady state lead to asymptotically reversible symmetry-breaking states. Let us now start with the simplest setting, namely that of discrete symmetries.

12.2   Discrete symmetries

Consider a system of L^D spin-1/2 particles. For example, this could be a system that undergoes Markovian dynamics which bring it to the thermal state of the ferromagnetic Ising model without external field. If the temperature is below the phase transition temperature (i.e., we assume $D \geq 2$), then we can decompose this state as $\omega = (\omega^+ + \omega^-)/2$, where ω^+ describes the spin-up phase and ω^- describes the spin-down phase. In particular, at zero temperature ω^+ would correspond to a product-state where all spins point up, while ω^- corresponds to a state where all spins point down. I will now express $\omega^\pm$ in a different way, making use of the fact that the state $\omega = (\omega^+ + \omega^-)/2$ has finite fluctuations in the magnetization density. Then I will show that any local, dissipative dynamics in detailed balance with the state $\omega = (\omega^+ + \omega^-)/2$ necessarily also has the states $\omega^\pm$ as asymptotically stationary states.

From the arguments that we will use, it will be clear that a similar conclusion holds for any state with a discrete symmetry that has finite fluctuations in the density of an order parameter since no special properties of the Ising model will be used. For concreteness however, let us assume that $\omega^\pm$ are the product-states with all spins up or down, respectively. For notational simplicity I will in the following drop all the explicit system-size dependence and write $L^D = N$. For example, the total magnetization in the system is now denoted by S^z instead of S^z_Λ and its spectrum ranges from $-N/2$ to $N/2$.

Let us first re-write the states $\omega^\pm$. From the fact that they are the extremal eigenstates of S^z we have $\omega^\pm(S^z A) = \omega^\pm(A S^z) = \pm N \omega^\pm(A)/2$ for any operator A. Similarly, we have $\omega^\pm(S^Z A S^z) = N^2 \omega^\pm(A)/4$. In particular, $\omega((S^z)^2) = N^2/4$ with $\omega = (\omega^+ + \omega^-)/2$. Let us define the operators

$$\tilde{O}^\pm := \frac{1}{\sqrt{2}} \left(\mathbf{1} \pm \frac{S^z}{\sqrt{\omega\left((S^z)^2\right)}} \right). \qquad (12.12)$$

Then we find

$$\omega(\tilde{O}^\pm A \tilde{O}^\pm) = \frac{1}{2} \left[\omega(A) + \frac{\omega(S^z A S^z)}{\omega\left((S^z)^2\right)} \pm \frac{\omega\left(\{S^z, A\}\right)}{\sqrt{\omega\left((S^z)^2\right)}} \right] \qquad (12.13)$$

$$= \frac{1}{2} \left[2\omega(A) \pm \left(\omega^+(A) - \omega^-(A) \right) \right]$$

$$= \omega^\pm(A). \qquad (12.14)$$

Now suppose that ω is not given by our specific example, but by any state that fulfills $\omega((S^z)^2) \geq \mu^2 N^2/4$ and that is symmetric with respect to spin-flips. Then we can try to use (12.13) as a *definition* of a symmetry-breaking state. Indeed this is possible whenever the state ω is symmetric and has non-vanishing fluctuations in the magnetization. To see this, first note that it is clearly positive and in fact it is also already normalized:

$$\omega^\pm(\mathbf{1}) := \omega(\tilde{O}^\pm \mathbf{1} \tilde{O}^\pm) = \frac{1}{2} \left[1 + \frac{\omega\left((S^z)^2\right)}{\omega\left((S^z)^2\right)} \pm 2 \frac{\omega\left(S^z\right)}{\sqrt{\omega\left((S^z)^2\right)}} \right] = 1, \qquad (12.15)$$

where we have used that $\omega(S^z) = 0$ due to symmetry. Let us now calculate the magnetization in the states $\omega^{\pm}$. We find

$$\left|\omega\left(\tilde{O}^{\pm}S^z\tilde{O}^{\pm}\right)\right| = \frac{1}{2}\left|\omega(S^z) + \frac{\omega\left((S^z)^3\right)}{\omega\left((S^z)^2)\right)} \pm 2\frac{\omega\left((S^z)^2\right)}{\sqrt{\omega\left((S^z)^2\right)}}\right| \tag{12.16}$$

$$= \sqrt{\omega\left((S^z)^2\right)} \geq \mu N/2, \tag{12.17}$$

where we have again used that the state ω is symmetric while odd powers of S^z are anti-symmetric under spin-flips. Therefore, the states $\omega^{\pm}$ explicitly break the symmetry and their magnetization-density is lower bounded by the fluctuations in the magnetization density in the original state ω.

I will now explain how to see that the states $\omega^{\pm}$ are not only symmetry-breaking, but also asymptotically stationary provided that ω is reversible with respect to the local Liouvillian $\mathcal{L}$ (note that we omit the Λ dependence). Consider equation (12.13) with the observable A replaced by $\mathcal{L}[A]$. Since ω is, by assumption, reversible, the states $\omega^{\pm}$ are asymptotically stationary provided that i) $\omega(S^z\mathcal{L}[A]S^z)$ grows slower than N^2 and that ii) $\omega(\{S^z, \mathcal{L}[A]\})$ grows slower than N as we increase the system-size.

Let me show the first statement; the latter follows by a completely analogous argument. First, we rewrite a product of operators using the commutator:

$$\omega\left(S^z\mathcal{L}[A]S^z\right) = \omega\left(S^z[\mathcal{L}[A],S^z]\right) + \omega\left((S^z)^2\mathcal{L}[A]\right). \tag{12.18}$$

Now, due to locality of the Liouvillian and the fact that S^z is an extensive operator the first term on the right hand side is at most of order N. We can therefore neglect it. For the second term, we will use the following Lemma.

Lemma 12.3 (Approximate Leibniz rule). *Let $\mathcal{L}$ be a strictly local Liouvillian a on regular lattice with N sites and A a strictly local operator. Then*

$$\mathcal{L}\left[(S^z)^2A\right] = \mathcal{L}\left[(S^z)^2\right]A + (S^z)^2\mathcal{L}[A] + O(N). \tag{12.19}$$

Before giving the proof of the Lemma, let us show how this property implies that $\omega(S^z\mathcal{L}[A]S^z)$ grows slower than N^2. From the above equation and reversibility of ω, we get

$$\omega\left((S^z)^2\mathcal{L}[A]\right) = \omega\left(\mathcal{L}\left[(S^z)^2A\right]\right) - \omega\left(\mathcal{L}\left[(S^z)^2\right]A\right) + O(N)$$

$$= -\omega\left(\mathcal{L}\left[(S^z)^2\right]A\right) + O(N)$$

$$= -\omega\left((S^z)^2\mathcal{L}[A]\right) + O(N). \tag{12.20}$$

We therefore conclude

$$\omega\left((S^z)^2\mathcal{L}[A]\right) = O(N) \tag{12.21}$$

as desired.

Proof of Lemma 12.3. Let us, by slight abuse of notation, identify an operator A with the support of it and let $\tilde{A}$ be the smallest region such that $\mathcal{L}[A] = \mathcal{L}_{\tilde{A}}[A]$, where $\mathcal{L}_{\tilde{A}}$ only contains those terms $\mathcal{L}_x$ that are supported inside of $\tilde{A}$. Then the super-operator $\mathcal{L} - \mathcal{L}_{\tilde{A}}$ does not act non-trivially on any operator inside (the support of) A. In particular, from the explicit form of the Liouvillians $\mathcal{L}_x$ we have

$$(\mathcal{L} - \mathcal{L}_{\tilde{A}})\left[(S^z)^2A\right] = (\mathcal{L} - \mathcal{L}_{\tilde{A}})\left[(S^z)^2\right]A.$$

We can then write

$$\mathcal{L}\left[(S^z)^2A\right] = (\mathcal{L} - \mathcal{L}_{\tilde{A}})\left[(S^z)^2\right]A + \mathcal{L}_{\tilde{A}}\left[(S^z)^2A\right]$$

$$= \mathcal{L}\left[(S^z)^2\right]A + (S^z)^2\mathcal{L}[A]$$

$$+ \left(\mathcal{L}_{\tilde{A}}\left[(S^z)^2A\right] - \mathcal{L}_{\tilde{A}}\left[(S^z)^2\right]A - (S^z)^2\mathcal{L}_{\tilde{A}}[A]\right),$$

where I have added and subtracted $(S^z)^2 \mathcal{L}[A] = (S^z)^2 \mathcal{L}_{\tilde{A}}[A]$. To prove the Lemma, we thus have to show that the term in parenthesis is of order N. Let us now decompose S^z as $S^z_{\tilde{A}} + S^z_{\tilde{A}^c}$, where $\tilde{A}^c$ is the complement of $\tilde{A}$ in the lattice Λ. Clearly, only the terms quadratic in $S^z_{\tilde{A}}$ could pose a problem. But these terms cancel identically:

$$
\mathcal{L}_{\tilde{A}}\left[(S^z_{\tilde{A}})^2 A\right] - \mathcal{L}_{\tilde{A}}\left[(S^z_{\tilde{A}})^2\right] A - (S^z_{\tilde{A}})^2 \mathcal{L}_{\tilde{A}}[A]
$$

$$
= (S^z_{\tilde{A}})^2 \mathcal{L}_{\tilde{A}}[A] - (S^z_{\tilde{A}})^2 \mathcal{L}_{\tilde{A}}[\mathbf{1}] A - (S^z_{\tilde{A}})^2 \mathcal{L}_{\tilde{A}}[A]
$$

$$
= -(S^z_{\tilde{A}})^2 \mathcal{L}_{\tilde{A}}[\mathbf{1}] A = 0,
$$

since a $\mathcal{L}[\mathbf{1}] = 0$ for any Liouvillian. This finishes the proof. $\qquad\square$

Note that none of the arguments showing asymptotic stationarity required that the magnetization transforms in a particular way under a symmetry. The only requirement was that the original state ω has finite fluctuations in the magnetization density or, equivalently, that the fluctuations in the magnetization scale like N^2. The fact that the magnetization changes sign under a global spin-flip is only necessary to prove that the states $\omega^\pm$ have a finite magnetization density.

Here, we have only discussed the case of a strictly local Liouvillian and shown asymptotic stationarity of the states $\omega^\pm$. By slightly refining the arguments one can further show that the states $\omega^\pm$ are in fact asymptotically reversible and can also incorporate Liouvillians which are not strictly local but decay like a power-law with exponent β, as long as $\beta > 2D$. It thus holds for any purely dissipative, short-range Liouvillian. The full theorem is then the following:

Theorem 12.4 (Reversibility and discrete symmetry breaking from fluctuations). *Let ω_Λ be a sequence of states that has finite fluctuation in the density of some extensive quantity O:*

$$
\omega_\Lambda(O^2_\Lambda) \geq \mu^2 o^2, \quad \mu > 0, \tag{12.22}
$$

where $O_\Lambda = \sum_{x \in \Lambda} O^\Lambda_{\{x\}}$ and $O^\Lambda_{\{x\}}$ are local operators supported in Λ, which are uniformly bounded in norm: $\left\| O_{\{x\}} \right\| \leq o$. Furthermore assume $\omega_\Lambda(O_\Lambda) = 0$ for all system-sizes and define the states

$$
\omega^\pm_\Lambda(A) := \omega_\Lambda\left(\left(\mathbf{1} + \frac{O_\Lambda}{\omega_\Lambda(O^2_\Lambda)} \right) A \left(\mathbf{1} + \frac{O_\Lambda}{\omega_\Lambda(O^2_\Lambda)} \right) \right). \tag{12.23}
$$

Then for any sequence of purely dissipative, short-ranged Liouvillians $\mathcal{L}^\Lambda$ which have ω_Λ as reversible steady state, the states $\omega^\pm_\Lambda$ are asymptotically reversible.

If furthermore there is a sequence of unitaries U_Λ such that $U_\Lambda O_\Lambda U_\Lambda^{-1} = -O_\Lambda$ and ω_Λ is symmetric with respect to these unitaries, then the states $\omega^\pm_\Lambda$ fulfill

$$
\left| \omega^\pm_\Lambda(O_\Lambda) \right| \geq \mu o N. \tag{12.24}
$$

Proof. The proof is given in Section 14.9.1. $\qquad\square$

As a last remark, note that translational invariance is not required for any of the appearing quantities. In particular, the result applies both to ferromagnetic and anti-ferromagnetic systems (where X is the staggered magnetization). Let us now come to the case of continuous symmetries.

12.3   Continuous symmetries

To discuss continuous symmetries, we assume that there is an extensive quantity C, which generates a symmetry on the lattice. A particular example to have in mind could be the case of $C = S^z$. We will assume that the *charge* C is a sum of on-site terms, so that the unitaries that represent the symmetry are tensor-products of unitaries on each site. We will

require the existence of two extensive order parameters $O^{(1)}, O^{(2)}$ that constitute a vector $(O^{(1)}, O^{(2)})$ which transforms under $U(1)$ in the appropriate way:

$$[O_\Lambda^{(1)}, C_\Lambda] = -iO_\Lambda^{(2)}, \quad [O_\Lambda^{(2)}, C_\Lambda] = iO_\Lambda^{(1)}. \tag{12.25}$$

For example, if we choose $C = S^z$, then we can choose $O^{(1)} = S^x$ and $O^{(2)} = S^y$. In general, we will write $O_\Lambda^{(i)} = \sum_{x \in \Lambda} O_{\{x\}}^{(i)}$ and assume $\left\| O_{\{x\}}^{(i)} \right\| \leq o$. Again, we will consider symmetric steady-states. Since we are working with finite systems, we can express the symmetry of the state as

$$[\rho_\Lambda^\omega, C_\Lambda] = 0. \tag{12.26}$$

It then follows from the commutation relations above that the associated state ω does not break the symmetry in the sense that $\omega_\Lambda(O_\Lambda^{(i)}) = 0$ for $i = 1, 2$. Similarly as before, we again assume that the steady state ω has extensive fluctuations in the order parameters:

$$\omega_\Lambda\left(((O^{(i)})_\Lambda)^2\right) \geq \mu^2 o^2 N^2, \quad i = 1, 2, \tag{12.27}$$

with $\mu > 0$ independent of the system size.

Under the assumptions given above, Koma and Tasaki have shown [287] in the context of spontaneous symmetry breaking of ground states of local Hamiltonians how to construct states that explicitly break the symmetry. In the following, we will see that in the context of dissipative systems, this construction gives rise to asymptotically reversible states. To state the construction, let us introduce the raising and lowering operators

$$O_\Lambda^\pm = O_\Lambda^{(1)} \pm iO_\Lambda^{(2)}. \tag{12.28}$$

Let m, m' be integers such that $|m|, |m'| \leq M$. Then let us define the functionals

$$\chi_\Lambda^{(m,m')}(A) := \frac{\omega_\Lambda\left((O_\Lambda^-)^{m'} A (O_\Lambda^+)^m\right)}{Z(m)Z(m')}, \tag{12.29}$$

with $Z(m) = \omega_\Lambda\left((O_\Lambda^-)^m (O_\Lambda^+)^m\right)^{1/2}$ and the convention $(O_\Lambda^+)^m = (O_\Lambda^-)^{-m}$ if $m < 0$. In particular, observe that $\chi_\Lambda^{(m,m')}(\mathbf{1}) = \delta_{m,m'}$ if ω_Λ is symmetric with respect to the rotations generated by the charge C_Λ. With these definitions at hand, the construction of Koma and Tasaki yields the following statement[3].

[3] In Ref. [287], the statement is only shown for pure states. By following the original derivation, it can be seen to also hold for the mixed states $\omega_\Lambda^{(M)}$.

Theorem 12.5 (Symmetry breaking states [287]). *For any $M < |\Lambda|$ define the states*

$$\omega_\Lambda^{(M)}(A) := \frac{1}{2M+1} \sum_{m=-M}^{M} \sum_{m'=-M}^{M} \chi_\Lambda^{(m,m')}(A). \tag{12.30}$$

Assume that the ω_Λ are represented by density matrices commuting with the charge: $[\rho_\Lambda^\omega, C_\Lambda] = 0$. If the condition

$$\omega_\Lambda\left((O_\Lambda^{(1)})^2\right) = \omega_\Lambda\left((O_\Lambda^{(2)})^2\right) \geq (\mu o |\Lambda|)^2 \tag{12.31}$$

is fulfilled with a constant $\mu > 0$, the states $\omega_\Lambda^{(M)}$ are asymptotically symmetry breaking in the sense that

$$\omega_\Lambda^{(M)}\left(O_\Lambda^{(2)}\right) = 0, \tag{12.32}$$

$$\lim_{M \to \infty} \lim_{\Lambda \nearrow \mathbb{Z}^d} \frac{1}{|\Lambda|} \omega_\Lambda^{(M)}\left(O_\Lambda^{(1)}\right) \geq \sqrt{2}\mu o. \tag{12.33}$$

We thus see that the states $\omega_\Lambda^{(M)}$ break the symmetry arbitrarily well in the direction of the operator $O^{(1)}$ as M increases. Clearly, we can rotate the states in any direction in the plane spanned by the order parameters $O^{(1)}$ and $O^{(2)}$ by acting with the symmetry operator

generated by the charge C. In the thermodynamic limit, we then get a whole $U(1)$-manifold of symmetry-breaking states. As a side-remark, also note that the theorem puts a bound on the possible values of μ. It is then possible, by a generalization of the techniques used in the last section and techniques introduced by Koma and Tasaki in Ref. [287] to prove the following result, which is a direct generalization of the result for discrete symmetries.

Theorem 12.6 (Symmetry breaking from fluctuations for continuous symmetries). *Let ω_Λ be a sequence of states that fulfills conditions (12.26) and (12.27). If $\mathcal{L}_\Lambda$ is a sequence of purely dissipative, local Liouvillians in detailed balance with ω_Λ, then the states $\omega_\Lambda^{(M)}$ are asymptotically reversible for any M.*

Similarly to the case of discrete symmetries, the theorem also generalizes to Liouvillians that are not strictly local, but whose terms decay faster than any power. The basic proof steps are the same: First we use a generalization of the approximate Leibniz property in Lemma 12.3 and then we use reversibility of the steady state. The details of the proof are, however, technically quite involved. They are therefore presented in section 14.9.2.

12.4   Goldstone-modes

In the case of closed systems, Goldstone's theorem states that a spontaneously broken continuous symmetry in a system with local interactions gives rise to gapless excitations, which are often called Goldstone-modes [304, 305]. The basic physical mechanism for this is quite simple: Suppose the system is in a symmetry-broken ground state and we introduce a continuous rotation of the spins with a very long wavelength into the system (a magnon). Since the interactions are local, the energy introduced into the system can be made arbitrarily small by increasing the wave-length. This is due to the fact that any local interaction term only sees an arbitrarily small change in the direction into which the spins point.

Given this explanation, it is natural to ask, whether we can prove the existence "dissipative goldstone modes", i.e., asymptotically stationary states with an order-parameter that varies over space. Unsurprisingly, we can indeed do this, if we assume that the Liouvillian $\mathcal{L}^\Lambda$ is symmetric in the sense that

$$\mathcal{L}^\Lambda\left[[C_\Lambda, A]\right] = \left[C_\Lambda, \mathcal{L}^\Lambda[A]\right]. \tag{12.34}$$

Let me sketch how to see this. Suppose the unitary U_Λ inserts a spin-wave with wavelength L in each space-direction into the system. Such a unitary is a tensor-product over all the sites in the lattice. Then its restriction U_A to a finite region A can be arbitrarily well approximated, with an error of order $1/L$, by a unitary V_A that rotates all the spins in the region A by the same small amount. In particular, we have

$$\left\|U_{\tilde{A}}\mathcal{L}_{\tilde{A}}^\Lambda[A]U_{\tilde{A}}^\dagger - V_{\tilde{A}}\mathcal{L}_{\tilde{A}}^\Lambda[A]V_{\tilde{A}}^\dagger\right\| = O(1/L), \tag{12.35}$$

Now define the states $\sigma_\Lambda^{(M)}(A) := \omega_\Lambda^{(M)}(U_\Lambda A U_\Lambda^\dagger)$, which describe states with one spin-wave on top of the states $\omega_\Lambda^{(M)}$. Then we have

$$\begin{aligned}
\sigma_\Lambda^{(M)}\left(\mathcal{L}^\Lambda[A]\right) &= \omega_\Lambda^{(M)}\left(U_{\tilde{A}}\mathcal{L}_{\tilde{A}}^\Lambda[A]U_{\tilde{A}}^\dagger\right) \approx \omega_\Lambda^{(M)}\left(V_{\tilde{A}}\mathcal{L}_{\tilde{A}}^\Lambda[A]V_{\tilde{A}}^\dagger\right) \\
&= \omega_\Lambda^{(M)}\left(V_\Lambda\mathcal{L}^\Lambda[A]V_\Lambda^\dagger\right) \\
&= \omega_\Lambda^{(M)}\left(\mathcal{L}^\Lambda\left[V_A^\dagger A V_A\right]\right) \approx 0,
\end{aligned}$$

where $\approx$ means equality up to an error of the order $1/L$ and the last line follows from the symmetry of the Liouvillian. Thus the states $\sigma_\Lambda^{(M)}$ are asymptotically stationary as advertised. Furthermore, for large Λ and M, the local spin rotates by 2π as we transverse the system. Hence the state describes a spin-wave.

12.5   Time-scales

In the previous two sections we saw that it is possible to construct asymptotically stationary states which break a given symmetry from the assumption of finite-fluctuations in an order parameter together with detailed balance. This shows that in the thermodynamic limit, there exist stationary states that break the given symmetry. In a finite system, however, these states will in general not be stationary. It is therefore interesting to discuss how long it takes for such states to equilibrate to the actual, symmetric stationary state.

In this section, I will show how to estimate these time-scales. For simplicity, we will work again with strictly local Liouvillians and for concreteness consider the case of discrete symmetries. The discussion in section 12.2 showed that we can estimate the time-derivative of a local observables A as

$$\left| \omega_\Lambda^\pm \left(\mathcal{L}\left[A^\Lambda \right] \right) \right| \leq k_1 \frac{\|A\| \, |\tilde{A}|}{N}. \tag{12.36}$$

Note that for an observable that is supported on the whole lattice, this bound only shows that the derivative is upper bounded by $k_1 \|A\|$. To derive a bound on the equilibration time, we will use Lieb-Robinson bounds, which allow us to approximate time-evolved local observables by observables that are supported in the *Lieb-Robinson cone* (cf. section 7.5.4 in chapter 7). Lieb-Robinson bounds also exist in the case of dissipative, local dynamics [288–290]. To discuss their application let us in the following denote time-evolved observables in the finite lattice as

$$A^\Lambda(t) := e^{t\mathcal{L}^\Lambda}\left[A\right]. \tag{12.37}$$

Lieb-Robinson bounds give rise to a finite velocity $v > 0$ with which information can spread through the system. We can use this to approximate the observable $A^\Lambda(t)$ by an observable $A^\vee(t)$ which is supported within a finite region that contains sites at most a distance $\tilde{v}t < L$ away from A, where $\tilde{v}$ controls the error of the approximation. As long as $\tilde{v} > v$, the error in the approximation decays exponentially with $\tilde{v}$. More precisely, there exists a constant k_2, independent of the system size, such that

$$\left\| A^\Lambda(t) - A^\vee(t) \right\| \leq k_2 \|A\| \, (\tilde{v}t)^{D-1} e^{-(\tilde{v}-v)t}. \tag{12.38}$$

The constant k_2 depends on the range and strength of the Liouvillian as well as the spatial dimension. We combine this bound with the estimate on the time-derivatives above to obtain

$$\left| \omega_\Lambda^\pm \left(\mathcal{L}\left[A^\Lambda(t) - A^\vee(t) \right] \right) \right| \leq k_1 k_2 \left\| A^\Lambda(t) - A^\vee(t) \right\| (\tilde{v}t)^{D-1} e^{-(\tilde{v}-v)t} \tag{12.39}$$

$$\leq 2 k_1 k_2 \|A\| \, (\tilde{v}t)^{D-1} e^{-(\tilde{v}-v)t}, \tag{12.40}$$

where we have used that both $\left\| A^\Lambda(t) \right\|$ and $\left\| A^\vee(t) \right\|$ are upper bounded by $\|A\|$ since both arise from the time-evolution under a Liouvillian. Clearly, the right hand side of the bound can be made as small as we wish by increasing $\tilde{v}$.

We can now use these ingredients to derive an estimate of the equilibration time. To do this, let us ask for the minimum time it takes for the expectation value of A to change by a given amount ΔA. Denote this time by t_{eq}. For simplicity, we consider here an on-site observable A, like the magnetization of a single spin, however a similar argument holds for any local observable. We then have

$$\Delta A = \left| \omega_\Lambda^\pm (A^\Lambda(t_{\text{eq}}) - A) \right| \leq \int_0^{t_{\text{eq}}} \left| \omega_\Lambda^\pm \left(\frac{\mathrm{d}A^\Lambda(s)}{\mathrm{d}s} \right) \right| \mathrm{d}s = \int_0^{t_{\text{eq}}} \left| \omega_\Lambda^\pm \left(\mathcal{L}^\Lambda \left[A^\Lambda(s) \right] \right) \right| \mathrm{d}s$$

$$\leq \int_0^{t_{\text{eq}}} \left| \omega_\Lambda^\pm \left(\mathcal{L}^\Lambda \left[A^\Lambda(s) - A^\vee(s) \right] \right) \right| \mathrm{d}s + \int_0^{t_{\text{eq}}} \left| \omega_\Lambda^\pm \left(\mathcal{L}^\Lambda \left[A^\vee(s) \right] \right) \right| \mathrm{d}s.$$

Since the Liouvillian is local, and $A^\vee(s)$ has support on a region of size at most $(2\tilde{v}s + 1)^D$, we can upper bound the last term as

$$\int_0^{t_{\text{eq}}} \left| \omega_\Lambda^\pm \left(\mathcal{L}^\Lambda \left[A^\vee(s) \right] \right) \right| \mathrm{d}s \leq k_1' \frac{\|A\| \, (\tilde{v}t_{\text{eq}})^{D+1}}{L^D},$$

where k_1' is a suitable constant and we have inserted $N = L^D$. For the other term, we use the Lieb-Robinson bounds to get

$$
\begin{aligned}
\int_0^{t_{\text{eq}}} \left| \omega_\Lambda^\pm \left(\mathcal{L}^\Lambda \left[A^\Lambda(s) - A^\vee(s) \right] \right) \right| \mathrm{d}s &\le 2k_1 k_2 \|A\| \int_0^{t_{\text{eq}}} (\tilde{v}s)^{D-1} \mathrm{e}^{-(\tilde{v}-v)s} \mathrm{d}s \\
&\le 2k_1 k_2 \|A\| \int_0^\infty (\tilde{v}s)^{D-1} \mathrm{e}^{-(\tilde{v}-v)} \mathrm{d}s \\
&= 2k_1 k_2 \|A\| \frac{(D-1)!}{(1-v/\tilde{v})^D} \frac{1}{\tilde{v}} \\
&:= 2k_1 k_2 \|A\| \, \delta(\tilde{v}, v, D) > 0. \qquad (12.41)
\end{aligned}
$$

The function $\delta(\tilde{v}, v, D)$ becomes arbitrarily small as $\tilde{v}$ increases. Let us now fix some $\epsilon > 0$. Then we can always choose $\tilde{v}$ such that

$$
2k_1 k_2 \|A\| \, \delta(\tilde{v}, v, D) < \epsilon \Delta A. \qquad (12.42)
$$

Combining all the estimates, we then obtain

$$
\frac{\Delta A}{\|A\|} (1 - \epsilon) < k_1' \frac{(\tilde{v} t_{\text{eq}})^{D+1}}{L^D}. \qquad (12.43)
$$

Rearranging the terms, we then conclude

$$
t_{\text{eq}} > \frac{\Delta A}{\|A\|} (1 - \epsilon) \frac{1}{(k_1')^{1/(D+1)} \tilde{v}} L^{D/(D+1)} = c L^{D/(D+1)}, \qquad (12.44)
$$

for some constant $c > 0$. Thus the equilibration time diverges at last as $L^{D/(D+1)}$ with the system size.

12.6 Conclusions

Spontaneous symmetry breaking is one of the most interesting phenomena in physics. This phenomenon is intrinsically connected to fluctuations of order parameters across all scales. In Hamiltonian systems it has been rigorously shown by Koma and Tasaki that if a symmetric ground state of a local Hamiltonian has finite fluctuations in an order parameter, then there are also low-lying states that break the symmetry and converge to ground states in the thermodynamic limit [287]. In this chapter, I built on top of their results and derived the corresponding results for local, purely dissipative systems in detailed balance. The results hold for discrete as well as continuous symmetries and I also demonstrated the existence of dissipative Goldstone modes on top of a symmetry-broken steady-state. In finite systems, the symmetry-breaking states that we constructed are not steady-states. I explicitly calculated a lower bound to their lifetime. This adds to the recent discussion on equilibration times and the role of the gap in dissipative many-body systems [291–296].

Recently, dissipative many-body dynamics has received some attention due to the theoretical possibility of engineered dissipation, which could be used to prepare many-body states that are useful for quantum computation or simulation [306–308]. One aim of quantum simulators is to study the phase diagram of zero-temperature Gibbs states of complex quantum systems, which are beyond the computational power of classical computers [170, 309]. The results in this chapter show that purely dissipative dynamics can only uniquely prepare states with strong correlations on large scale if the dynamics is *not* in detailed balance with the target state.

To obtain the results in this chapter, we had to assume local dynamics that fulfills detailed balance. While the locality of the dynamics is certainly a necessary assumption, it is an interesting open question, whether the assumption on detailed balance can be overcome. In statistical mechanics it is well known that the symmetry-breaking states below the critical temperature show exponential decay of correlations with a correlation length that diverges if one approaches a continuous phase-transition (and leads to fluctuations on all scales at the critical point). A further interesting open question is whether the explicit

symmetry breaking states that we used – adapted from Ref. [287] – can be proven to show at least some decay of correlations. Finally it would be interesting to know whether the construction leads to KMS-states, representing thermal equilibrium, in the thermodynamic limit if applied to Gibbs-states [275].

13 Conclusions and Acknowledgments

In this thesis, I discussed the basic problem of relating thermodynamics and quantum mechanics from a variety of angles. First, using a resource theoretic approach, I presented novel ways of discussing the emergence of thermodynamic concepts from basic principles in chapters 2–6. This approach is essentially an information theoretic approach, building on concepts such as divergence measures for probability distributions. These tools are quite elegant, but require an existing notion of thermal equilibrium in one way or another. I then discussed how such equilibrium emerges from the point of view of many-body physics and, in chapter 8, connected this point of view with a resource theoretic point of view.

Then, we turned to a more hands-on approach to studying the thermodynamics of thermal machines in the quantum regime. We discussed in detail how experimental restrictions and strong coupling to heat baths influence fundamental thermodynamic bounds. Interestingly, the corrections to thermodynamic bounds can be expressed in terms of purely information theoretic quantities, namely relative entropies.

Finally, I discussed how the emergence of spontaneous symmetry breaking can be understood from a point of view of dissipative, open system dynamics – providing a novel link between equilibration time-scales, fluctuations in intensive quantities and the breaking of symmetries.

An underlying theme in this thesis is the role of correlations at the microscopic scale in thermodynamics: In chapters 4 and 5, we saw how they have to be taken into account to correctly account for the value of state transitions from a thermodynamic point of view and how correlations to cyclically operating systems can be used to single out the relative entropy and free energy. In chapters 8, we similarly saw that establishing correlations to such "catalysts" allows to circumvent the existence of a "second law of equilibration". In chapters 9, we saw that the inability to control the interaction between particles that constitute a working system leads to irreversible dissipation. Such inability to control interactions in term is another manifestation of an inability to control the correlations between the particles. In chapter 11, we similarly saw that the dissipation that occurs in thermal machines due to strong coupling to heat baths is a measure for the uncontrollable correlations that are build up when coupling the machine to a heat bath. Finally, in chapter 12 we saw how spontaneous symmetry breaking emerges from states with statistical fluctuations in the density of an order parameter. Such fluctuations in turn are simply a statement of the fact that the spins in the lattice are correlated on all scales.

The results in this thesis thus also contribute to the discussion of the role of correlations in thermodynamics [310–314]. The microscopic correlations between the particles that make up a macroscopic body can usually be safely neglected when considering the interaction between macroscopic systems on a macroscopic scale. They hence do not contribute significantly to the thermodynamic bounds on macroscopic quantities. At the microscopic scale, however, we have seen that it is essential to take them into account.

Large parts of this thesis, in particular those that deal with the resource theoretical approach to thermodynamics, are concerned with fundamental bounds, which may require an experimenter to implement almost arbitrary unitary dynamics on a quantum system to saturate the bounds. Fifty years ago, implementing arbitrary unitary dynamics on complex quantum systems was almost unimaginable. Today however, with the possible advent of scalable quantum computers in the foreseeable future, we enter an era where such bounds might indeed be experimentally testable in artificial systems. From the point of view of a quantum computer, different unitary operations can be classified in terms of their quantum

computational complexity [62]. It would therefore be interesting to study thermodynamics under restricted (quantum) computational complexity to establish a notion for how difficult it is to implement thermodynamically optimal operations on quantum computers.

Furthermore, it was recently suggested by Brown and Susskind [315], inspired by ideas from quantum gravity, that quantum computational complexity itself might follow laws similar to those of thermodynamics. It would be fascinating if the resource theoretical approach could be used to understand this relation in more detail.

In the converse direction, we have seen in the example of anomalous heat flow (see chapter 2), that interesting effects can already occur in the presence of "heat-baths" that are made up only of a few atoms. It would be interesting to study in more detailed what can be achieved by microscopic thermal machines having access only to such microscopic heat baths (see Refs. [90] for progress in this direction).

We have made a long way in understanding how statistical mechanics and thermodynamics emerges from basic quantum mechanics. Nevertheless, as the many open problems formulated in the individual chapters of this thesis show, there are still many open problems left. It is my hope that the contributions in this thesis help to motivate more researchers to work on this fascinating and interdisciplinary field.

Acknowledgments

14 Technical proofs und supplementary infor-mation

14.1   Thermodynamics as a resource theory

14.1.1   Proofs for anomalous heat flow

In this section I will prove the formula for the final excitation probability (2.9) in the anomalous heat flow. Consider the algorithm as described in section 2.2. The final probability is given by

$$p' = \text{probability transferred from ground state } + \text{ initial probability that remains in excited state.}$$

I will first calculate the probability that is transferred from the ground state to the excited state. As explained in the description of the algorithm, for each of the energies E_k^0, exactly $\min\{\binom{2n}{k-1}, \binom{2n}{k}\}$ states are swapped with with states corresponding to E_{k-1}^1. Each of them have initial probability given by $(1-p)\frac{e^{-\beta k\Delta}}{Z_\beta^{(n)}}$, with $Z_\beta^{(n)} = \sum_{k=0}^{2n} \binom{2n}{k}e^{-\beta k\Delta}$. The total probability that is transferred from the ground-state to the excited state is then given by

$$
\begin{aligned}
(1-p)\mathcal{P}(1 \leftarrow 0) &:= (1-p) \sum_{k=1}^{2n} \min\{\binom{2n}{k-1}, \binom{2n}{k}\}e^{-\beta k\Delta} \\
&= (1-p)\left[\sum_{k=1}^{n} \binom{2n}{k-1}\frac{e^{-\beta k\Delta}}{Z_\beta^{(n)}} + \sum_{k=n+1}^{2n} \binom{2n}{k}\frac{e^{-\beta k\Delta}}{Z_\beta^{(n)}}\right]. \\
&= (1-p)p_n^*(\beta).
\end{aligned}
\tag{14.1}
$$

Considering now the initial probability that remains in the excited state, we see that in every step of the algorithm exactly $\binom{2n}{k} - \min\{\binom{2n}{k}, \binom{2n}{k+1}\}$ states remain in the subspace associated to E_k^1. Each of them have initial probability $pe^{-\beta k\Delta}/Z_\beta^{(n)}$. We then obtain for the probability that remains in the

Figure 14.1: The function $p_n^*(\beta)$ as a function of n for $\Delta = 1$ and $\beta = 0.1, 0.5, 1$ (from light to dark). The straight, solid lines show the limiting values (and upper bounds) $e^{-\beta\Delta}$.
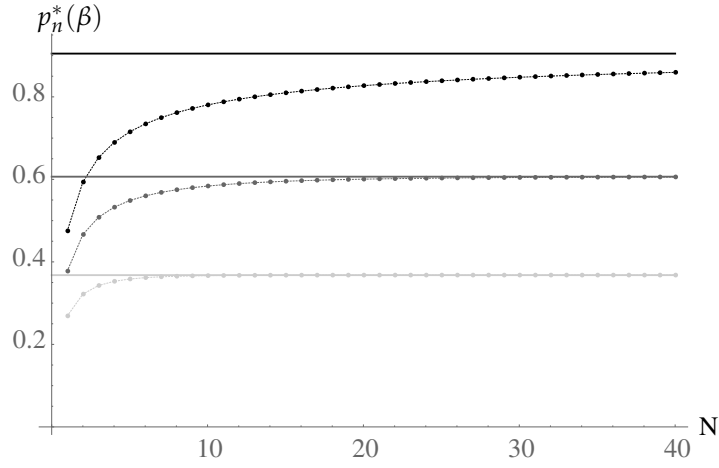
excited state:

$$
pP(1 \leftarrow 1) = p \sum_{k=0}^{2n} \left[\binom{2n}{k} - \min\{\binom{2n}{k}, \binom{2n}{k+1}\} \right] \frac{e^{-\beta k \Delta}}{Z_\beta^{(n)}}
$$

$$
= p \left[1 - \sum_{k=0}^{2n} \min\{\binom{2n}{k}, \binom{2n}{k+1}\} \frac{e^{-\beta k \Delta}}{Z_\beta^{(n)}} \right]
$$

$$
= p \left[1 - e^{\beta\Delta} \sum_{k=1}^{2n} \min\{\binom{2n}{k-1}, \binom{2n}{k}\} \frac{e^{-\beta k \Delta}}{Z_\beta^{(n)}} \right]
$$

$$
= p \left[1 - e^{\beta\Delta} p_n^*(\beta) \right]. \tag{14.2}
$$

Putting everything together, we arrive at the final result

$$
p' = (1 - p) p_n^*(\beta) + p \left(1 - e^{\beta\Delta} p_n^*(\beta) \right)
$$

$$
= p_n^*(\beta) - p \left(\frac{p_n^*(\beta)}{p_\beta} - 1 \right). \tag{14.3}
$$

What is left to show is the achievable upper bound $p_n^*(\beta) \leq e^{-\beta\Delta}$. It can be proven fully rigorously, however the derivation is quite technical and offers no additional insight. I hence simply show a plot for $\Delta = 1$ and different values of n and β in figure 14.1.

14.2   Ground state cooling and the unattainability principle

14.2.1   Proof of theorem 3.2

We will now proof theorem 3.2. Before we go to the details, it is important to understand the basic structure of the proof. It is clear that to obtain a single necessary and sufficient condition for cooling at low temperatures, we have to show that the infinite set of second laws given by theorem 2.1 collapse to a single condition. The first important step in the proof is the following Lemma.

Lemma 14.1 (Concavity at low temperatures). *Let $\beta > 0$ and a Hamiltonian H_S be given. Then there exists a critical inverse temperature β_{cr} such that for all $\beta_S > \beta_{\mathrm{cr}}$ and for all $0 < \alpha < \delta(\beta_S)$ we have*

$$\alpha \mapsto D''_\alpha(\omega_{\beta_S} \| \omega_\beta) \leq 0. \tag{14.4}$$

and

$$D_\infty(\omega_{\beta_S} \| \omega_\beta) \leq \log Z_\beta. \tag{14.5}$$

Here, the critical value $\delta(\beta_S)$ is given by

$$\delta(\beta_S) = \frac{\log(Z_\beta(H_S))}{\mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S)}. \tag{14.6}$$

Proof.   See section 14.2.2.   □

The Lemma tells us that we can *upper bound* the Rényi-divergence in the parameter regime prescribed by the Lemma by the linear approximation at the origin, since a concave function is always upper bounded by the linear approximation at any point of the graph. Since $D'_0(\rho \| \omega_\beta(H)) = D(\omega_\beta(H) \| \rho) = \mathcal{V}_\beta(\rho, H)$, we get

$$D_\alpha(\omega_{\beta_S}(H_S) \| \omega_\beta(H_S)) \leq \mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S)\alpha, \quad \alpha \leq \alpha_c. \tag{14.7}$$

By restricting to small enough target temperatures, we can also make sure that

$$D_\infty(\omega_{\beta_S}(H_S) \| \omega_\beta(H_S)) \leq \mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S)\alpha_c. \tag{14.8}$$

Now, the function $\alpha \mapsto D_\alpha$ is always monotonously increasing. Therefore the sufficient condition

$$D_\alpha(\rho_R \| \omega_\beta(H_R)) > D_\alpha(\omega_{\beta_S}(H_S) \| \omega_\beta(H_S)), \quad \alpha \geq 0, \tag{14.9}$$

are also satisfied if

$$D_\alpha(\rho_R \| \omega_\beta(H_R)) > \mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S)\alpha, \quad \alpha \leq \alpha_c. \tag{14.10}$$

For small temperatures, we can simply restrict the range of α to the interval $[0, \delta(\beta_S))$, where $\delta(\beta_S)$ is given by:

$$\delta(\beta_S) = \frac{D_\infty(\omega_{\beta_S}(H_S) \| \omega_\beta(H_S))}{\mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S)}. \tag{14.11}$$

We now turn to the l.h.s. of the sufficient condition. We will show that the Rényi-divergence of the $D_\alpha(\rho_R \| \omega_\beta(H_R))$ is approximately convex in α for $\alpha < \delta(\beta_S)$ in a sense made more precise below. Note that if it actually was convex we could always lower bound it by $\mathcal{V}_\beta(\rho_R, H_R)\alpha$ and obtain the necessary and sufficient condition (3.10) in full generality.

In general $D_\alpha(\rho_R \| \omega_\beta(H_R))$ is not convex, but we can take advantage of the fact that we only have to check small values of α. We thus Taylor expand $D_\alpha(\rho_R \| \omega_\beta(H_R))$ and use Taylor's theorem to obtain

$$D_\alpha(\rho_R \| \omega_\beta(H_R)) \geq \mathcal{V}_\beta(\rho_R, H_R)\alpha - k(\beta_S, \beta, \rho_R, H_R)\alpha^2. \tag{14.12}$$

This gives us the new *sufficient* condition

$$\mathcal{V}_\beta(\rho_R, H_R)\alpha - k(\beta_S, \beta, \rho_R, H_R)\alpha^2 \geq \mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S)\alpha,$$

for all $0 < \alpha \leq \delta(\beta_S)$. The function $k(\beta_S, \beta, \rho_R, H_R) \geq 0$ is given by

$$k(\beta_S, \beta, \rho_R, H_R) = \max\left\{0, -\min_{\alpha \leq \delta(\beta_S)} D''_\alpha(\rho_R \| \omega_\beta(H_R))\right\}.$$

Dividing by α and replacing α by $\delta(\beta_S)$ we arrive at the final sufficient condition

$$\mathcal{V}_\beta(\rho_R, H_R) - K(\beta_S, \beta, \rho_R, H_R, H_S) \geq \mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S),$$

with $K(\beta_S, \beta, \rho_R, H_R, H_S) = k(\beta_S, \beta, \rho_R, H_R)\delta(\beta_S)$.   □

14.2.2   Proof of concavity of Rényi-divergence for low temperatures

Let us now prove Lemma 14.1. The first statement to prove is that for small enough temperatures the Rényi-divergence $D_\infty(\omega_{\beta_S}(H_S)\|\omega_\beta(H_S))$ is upper bounded by $Z_\beta(H_S)$. For simplicity of notation, we will write Z_β instead of $Z_\beta(H_S)$ in the following. We can then write

$$D_\alpha(\omega_{\beta_S}(H_S)\|\omega_\beta(H_S)) = \frac{1}{\alpha-1}\log\left(\sum_i g_i e^{-\alpha(\beta_S-\beta)E_i}(Z_\beta/Z_{\beta_S})^\alpha \frac{e^{-\beta E_i}}{Z_\beta}\right)$$

$$= \frac{1}{\alpha-1}\log\left(e^{-\alpha(\beta_S-\beta)E_0}(Z_\beta/Z_{\beta_S})^\alpha \sum_i g_i e^{-\alpha(\beta_S-\beta)(E_i-E_0)}\frac{e^{-\beta E_i}}{Z_\beta}\right),$$

where E_i denote the different energies of H_S, with degeneracies g_i. By assuming w.l.o.g. $E_0 = 0$, we can write this as

$$D_\alpha(\omega_{\beta_S}(H_S)\|\omega_\beta(H_S)) = \frac{\alpha}{\alpha-1}\log(Z_\beta/Z_{\beta_S}) + \frac{1}{\alpha-1}\log\left(1 + \sum_{i>0} e^{-\alpha(\beta_S-\beta)E_i}\frac{e^{-\beta E_i}}{Z_\beta}\right).$$

We can now take the limit $\alpha \to \infty$ and obtain ($\beta_S > \beta$)

$$D_\infty(\omega_{\beta_S}(H_S)\|\omega_\beta(H_S)) = \lim_{\alpha\to\infty} D_\alpha(\omega_{\beta_S}(H_S)\|\omega_\beta(H_S)) = \log(Z_\beta) - \log(Z_{\beta_S}) \le \log Z_\beta.$$

Let us now find the condition for

$$\delta(\beta_S) = \frac{\log(Z_\beta(H_S)}{\mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S)} < 1. \tag{14.13}$$

To do that we make use of the fact that for thermal states, the vacancy is a non-equilibrium free energy and can be expressed as

$$\mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S) = \beta_S E_\beta - S_\beta + \log Z_{\beta_S}, \tag{14.14}$$

where $S_\beta := S(\omega_\beta(H_S))$. This gives us the condition

$$\beta_S E_\beta - S_\beta > \log Z_\beta - \log Z_{\beta_S}. \tag{14.15}$$

We can relax this to obtain the sufficient criterion $\beta_S E_\beta - S_\beta > \log Z_\beta = S_\beta - \beta E_\beta$ and obtain

$$\beta_S > \frac{2S_\beta - \beta E_\beta}{E_\beta}. \tag{14.16}$$

We are now in position to prove the concavity. To do that we use the following representation of D_α'', which is proven in section 14.2.3:

$$D_\alpha''(\omega_{\beta_S}\|\omega_\beta) = \frac{2}{(1-\alpha)^3}\left(\log Z_{\beta_S} - \log Z_{\tilde\beta(\alpha)} + (\beta_S - \tilde\beta(\alpha))E_{\tilde\beta(\alpha)} - (\beta_S - \tilde\beta(\alpha))^2 \mathrm{Var}(H)_{\tilde\beta(\alpha)}\right), \tag{14.17}$$

where $\tilde\beta(\alpha) = \beta(1-\alpha) + \alpha\beta_S$. Since we are only interested in $\alpha < \delta(\beta_S) < 1$, we have $\beta \le \tilde\beta(\alpha) < \beta_S$. We thus need to prove that the terms in the parenthesis are negative. To do this, we use that the average energy is monotonic with β and that $Z_{\tilde\beta(\alpha)} > 1$. We can then bound these terms as

$$\text{parenthesis} \le \log Z_{\beta_S} + (\beta_S - \tilde\beta(\alpha))E_{\tilde\beta(\alpha)} - (\beta_S - \tilde\beta(\alpha))^2 \mathrm{Var}(H)_{\tilde\beta(\alpha)}$$

$$\le \log Z_{\beta_S} + (\beta_S - \beta)E_\beta - (\beta_S - \tilde\beta(\alpha))^2 \mathrm{Var}(H)_{\tilde\beta(\alpha)}$$

$$\le \log(d) + (\beta_S - \beta)E_{\max} - (\beta_S - \tilde\beta(\alpha))^2 \min_{x\in[\beta,\tilde\beta(\alpha)]} \mathrm{Var}(H)_x. \tag{14.18}$$

We now use that $\tilde\beta(\alpha) \le \tilde\beta(\delta(\beta_S)) =: \tilde\beta^*(\beta_S)$. If we can bound $\tilde\beta^*(\beta_S)$ by a constant, the terms in the parenthesis become negative for some large enough β_S, because the second order term in β_S dominates. To prove that $\tilde\beta^*(\beta_S)$ is indeed upper bounded by a constant, we again write the vacancy as

$$\mathcal{V}_\beta(\omega_{\beta_S}(H_S), H_S) = -S_\beta + \beta_S E_\beta + \log Z_{\beta_S}. \tag{14.19}$$

From this we get

$$\beta^* := \lim_{\beta_S\to\infty}\tilde\beta^*(\beta_S) = \lim_{\beta_S\to\infty}\beta(1-\delta(\beta_S)) + \delta(\beta_S)\beta_S \tag{14.20}$$

$$= \beta + \lim_{\beta_S\to\infty}\frac{\log Z_\beta}{\beta_S E_\beta + \log Z_{\beta_S} - S(\omega_\beta(H_S))}\beta_S = \beta + \frac{\log Z_\beta}{E_\beta}. \tag{14.21}$$

Thus, there will be some β_{cr} which ensures concavity. Note that $\tilde\beta^*(\beta_S)$ is monotonically decreasing for all β_S such that $\tilde\beta^*(\beta_S) < 1$. This finishes the proof. The critical value β_{cr} can be upper bounded if lower bounds on the energy variance for inverse temperatures in the interval $[\beta, \beta^*]$ are available.

14.2.3   Rényi-divergence between thermal states

In this section, we derive results about the Rényi-divergence between two thermal states of the same Hamiltonian at different temperatures. For concreteness we consider a resource system with Hamiltonian H_R and inverse temperature β_R. By simple rewriting, we first obtain

$$
\begin{aligned}
D_\alpha(\omega_{\beta_R}||\omega_\beta) &= -\frac{\alpha}{\alpha-1}\log Z_{\beta_R} + \log Z_\beta + \frac{1}{\alpha-1}\log \mathrm{Tr}(e^{-\beta_R H\alpha}e^{-\beta H(1-\alpha)}) \\
&= -\frac{\alpha}{\alpha-1}\log Z_{\beta_R} + \log Z_\beta + \frac{1}{\alpha-1}\log Z_{(\beta_R-\beta)\alpha+\beta} \\
&= -\frac{\alpha-1}{\alpha-1}\log Z_{\beta_R} + \log Z_\beta + \frac{1}{\alpha-1}\log(Z_{(\beta_R-\beta)\alpha+\beta}/Z_{\beta_R}) \\
&= \log(Z_\beta/Z_{\beta_R}) + \frac{1}{\alpha-1}\log(Z_{(\beta_R-\beta)\alpha+\beta}/Z_{\beta_R}).
\end{aligned}
\tag{14.22}
$$

We can use this representation to show that the Rényi-divergence is convex in α provided that $\beta_R < \beta$ and that the function $x \mapsto E_{\beta_R+x}$ is convex for $0 \le x \le \beta - \beta_R$. To do that, we calculate the second derivative (with $\tilde{\beta} = (\beta_R-\beta)\alpha + \beta$):

$$
D_\alpha(\omega_{\beta_R}||\omega_\beta)'' = \frac{2}{(1-\alpha)^3}\log Z_{\beta_R} - \frac{2}{(1-\alpha)^3}\log Z_{\tilde{\beta}} - 2\frac{1}{(1-\alpha)^2}\partial_\alpha \log Z_{\tilde{\beta}} + \frac{1}{\alpha-1}\partial_\alpha^2 \log Z_{\tilde{\beta}}
\tag{14.23}
$$

$$
= \frac{2}{(1-\alpha)^3}\log Z_{\beta_R} - \frac{2}{(1-\alpha)^3}\log Z_{\tilde{\beta}} - 2\frac{1}{(1-\alpha)^2}(\beta-\beta_R)E_{\tilde{\beta}} - \frac{1}{1-\alpha}(\beta-\beta_R)^2 \mathrm{Var}(H)_{\tilde{\beta}}
\tag{14.24}
$$

$$
= \frac{2}{(1-\alpha)^3}\left[\log Z_{\beta_R} - \log Z_{\tilde{\beta}} - (1-\alpha)(\beta-\beta_R)E_{\tilde{\beta}} - \frac{(1-\alpha)^2}{2}(\beta-\beta_R)^2 \mathrm{Var}(H)_{\tilde{\beta}}\right].
\tag{14.25}
$$

Utilizing $(1-\alpha)(\beta-\beta_R) = \tilde{\beta} - \beta_R$, we can write this as

$$
D_\alpha(\omega_{\beta_R}||\omega_\beta)'' = \frac{2}{(1-\alpha)^3}\left[\log Z_{\beta_R} - \log Z_{\tilde{\beta}} - (\tilde{\beta}-\beta_R)E_{\tilde{\beta}} - \frac{(\tilde{\beta}-\beta_R)^2}{2}\mathrm{Var}(H)_{\tilde{\beta}}\right].
\tag{14.26}
$$

Here, E_β denotes the energy-expectation value at inverse temperature β and we have introduced the variance $\mathrm{Var}(H)_\beta = \langle H^2 \rangle_\beta - \langle H \rangle_\beta^2$. The two functions are related by $\partial_x E_x = -\mathrm{Var}(H)_x$. We can now show theorem 3.3 and another result about about the convexity of Rényi divergences for a sufficiently large reference temperature β.

Proof of theorem 3.3   We need to show that the r.h.s. of (14.26) is positive under the premise that $x \mapsto E_x$ is convex in $x \in [\beta_R, \beta]$, $\beta_R \le \beta$ and $\alpha < 1$. Thus we have to show that

$$
\log Z_{\beta_R} - \log Z_{\tilde{\beta}} \ge (\tilde{\beta}-\beta_R)E_{\tilde{\beta}} + \frac{(\tilde{\beta}-\beta_R)^2}{2}\mathrm{Var}(H)_{\tilde{\beta}}.
\tag{14.27}
$$

To achieve this, we use the following integral representation of the l.h.s:

$$
\log Z_{\beta_R} - \log Z_{\tilde{\beta}} = -\int_0^{\tilde{\beta}-\beta_R}\frac{\mathrm{d}}{\mathrm{d}x}\log Z_{\beta_R+x}\,\mathrm{d}x = \int_0^{\tilde{\beta}-\beta_R}E_{\beta_R+x}\mathrm{d}x.
\tag{14.28}
$$

We thus need to show

$$
\int_0^{\tilde{\beta}-\beta_R}E_{\beta_R+x}\mathrm{d}x \ge (\tilde{\beta}-\beta_R)E_{\tilde{\beta}} + \frac{(\tilde{\beta}-\beta_R)^2}{2}\mathrm{Var}(H)_{\tilde{\beta}}.
\tag{14.29}
$$

Whether this inequality is satisfied, and therefore, whether $D_\alpha(\rho_R||\omega_\beta(H))$ is convex, is entirely determined by the function $x \mapsto E_x$, because derivative of E_x is given by $-\mathrm{Var}(H)_x$. The right hand side is therefore a linear approximation to the function E_x. Fig. 14.2 provides a geometrical interpretation which makes clear that the inequality is satisfied whenever E_x is convex.

As a side-remark, note that a completely analogous argument implies that if $\alpha < 1$, E_x is convex but in contrast to the previous case $\beta_R \ge \beta$, then we have

$$
\int_0^{\tilde{\beta}-\beta_R}E_{\beta_R+x}\mathrm{d}x \le (\tilde{\beta}-\beta_R)E_{\tilde{\beta}} + \frac{(\tilde{\beta}-\beta_R)^2}{2}\mathrm{Var}(H)_{\tilde{\beta}}.
\tag{14.30}
$$

Thus for systems colder than the bath, the function $D_\alpha(\rho_R||\omega_\beta(H))$ is concave whenever E_x is convex. In particular this is true for qubits as target systems for cooling.
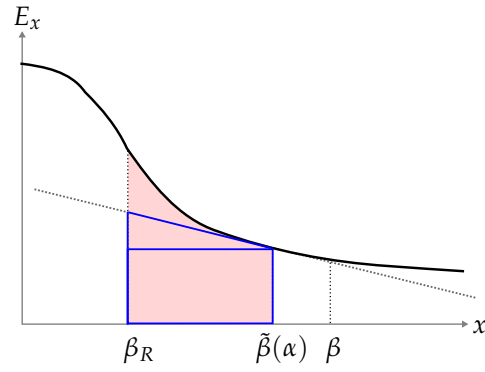
Figure 14.2: The l.h.s. of (14.29) is represented by the light red area under the curve. The r.h.s. corresponds to the area composed of the rectangle and triangle with blue border. The rectangle of side-lengths $\tilde{\beta}(\alpha) - \beta_R$ and $E_{\tilde{\beta}(\alpha)}$ corresponds to the first term of the r.h.s of (14.29). The triangle corresponds to the second term. If the function is E_x convex, the light red area is always larger than the area of the triangle and rectangle combined. (Figure adapted from Ref. [8].)

Very cold heat baths   We can also show that $D_\alpha(\omega_{\beta_R} \| \omega_\beta)$ is convex for very cold environments, i.e., very large β. Thus in this case (3.10) becomes sufficient and necessary as well.

Theorem 14.2. *For any resource of the form $(\omega_{\beta_R}(H_R), H_R)$, given a fixed β_R there exist a sufficiently large value of β so that (3.10) is a sufficient and necessary condition for low temperature cooling.*

Proof-sketch. We show that $D_\alpha(\omega_{\beta_R} \| \omega_\beta)$ is convex for values of $\alpha < \alpha_c$, where $\alpha_c < 1$ is chosen arbitrarily. Recalling Eq. 14.29, we have to show that

$$\int_0^{\tilde{\beta} - \beta_R} E_{\beta_R + x} \mathrm{d}x \geq (\tilde{\beta} - \beta_R) E_{\tilde{\beta}} + \frac{(\tilde{\beta} - \beta_R)^2}{2} \mathrm{Var}(H)_{\tilde{\beta}}. \tag{14.31}$$

In the limit of large β the scaling of the r.h.s. of (14.31) is such that $\tilde{\beta} - \beta_R = (1 - \alpha)(\beta - \beta_R)$ scales proportionally to β, while $E_{\tilde{\beta}}$ and $\mathrm{Var}(H)_{\tilde{\beta}}$ scale as $e^{-k\beta}$. Therefore, the r.h.s. of (14.31) approaches zero as $\beta \to \infty$ whereas the l.h.s. grows monotonically with β. Hence, (14.31) is fulfilled for some large enough β. $\qquad\square$

14.2.4   Equidistant levels

In this section we show that the function E_β is convex for any system of $M + 1$ equidistant levels. We assume that the energy-gap between subsequent levels is Δ and as always we set the ground state energy to zero. The energy E_β is then given by

$$E_\beta = \frac{1}{e^{\beta\Delta} - 1}\Delta - \frac{M + 1}{e^{(M+1)\Delta\beta} - 1}\Delta. \tag{14.32}$$

Note that the limit $M \to \infty$ corresponds to an harmonic oscillator and $M = 1$ corresponds to a two-level system. To show that the energy is a convex function of β, we have to show that the second derivative of the energy is positive. We can express the second derivative as

$$E_\beta'' = \frac{1}{8}\Delta^3 \left[\frac{\sinh(\beta\Delta)}{\sinh(\beta\Delta/2)^4} - \underbrace{(M+1)^3 \frac{\sinh((M+1)\beta\Delta)}{\sinh((M+1)\beta\Delta/2)^4}}_{=:f(\beta,M+1)} \right]. \tag{14.33}$$

For $M = 0$ this expression is certainly positive. To prove positivity for arbitrary M, we set $M + 1 =: \gamma$ and show that $\partial_\gamma f(\beta, \gamma) \leq 0$. The derivative $\partial_\gamma f(\beta, \gamma)$ is given by

$$\partial_\gamma f(\beta, \gamma) = -\gamma^2 \frac{1}{\sinh(\gamma\beta\Delta/2)^4} \left[\gamma\beta\Delta(2 + \cosh(\gamma\beta\Delta)) - 3\sinh(\gamma\beta\Delta) \right]. \tag{14.34}$$

We now set $\gamma\beta\Delta = x$. Then it suffices to show

$$x(2 + \cosh(x)) - 3\sinh(x) \geq 0. \tag{14.35}$$

This can be done using a Taylor-expansion:

$$2x + x\cosh(x) - 3\sinh(x) = 2x + \sum_{n=0}^{\infty} x^{2n+1}\left(\frac{1}{(2n)!} - \frac{3}{(2n+1)!}\right) \tag{14.36}$$

$$= 2x + \sum_{n=0}^{\infty} x^{2n+1}\left(\frac{((2n+1)-3)(2n)!}{(2n)!(2n+1)!}\right) \tag{14.37}$$

$$= 2x - 2x + \sum_{n=1}^{\infty} x^{2n+1}\left(\frac{((2n+1)-3)(2n)!}{(2n)!(2n+1)!}\right) \tag{14.38}$$

$$\geq 0. \tag{14.39}$$

14.2.5 Proof of theorem 3.6

In this section we proof the sufficient condition under exact catalysis presented in theorem 3.6. As explained in section 2.3, a sufficient condition for a state transition using catalytic thermal operations with exact catalysts is given by the decrease of all Rényi-divergences. We thus need to consider the inequalities

$$D_\alpha(\rho_R\|\omega_\beta(H_R)) \geq D_\alpha(\omega_{\beta_S}(H_S)) \tag{14.40}$$

for all $\alpha \in (-\infty, +\infty)$. The case $\alpha \geq 0$ is already treated in theorem 3.2. We thus have to show that by introducing the additional factor $r(\beta, H_S)$, we can also make sure that the inequalities are valid for $\alpha < 0$. To do this, we will prove the following two Lemmas.

Lemma 14.3 (Lower bound to Rényi-divergences with negative α)**.** *Let $[\rho, H]$ for some finite-dimensional Hamiltonian H and let ρ be a quantum state with full rank. Then for any $\beta > 0$ we have*

$$D_{-|\alpha|}(\rho\|\omega_\beta(H)) \geq \frac{|\alpha|}{|\alpha|+1} V_\beta(\rho, H). \tag{14.41}$$

Lemma 14.4 (Upper bound to Rényi-divergences with negative α)**.** *Let $H_S \geq 0$ be any finite dimensional Hamiltonian with ground state energy zero and $\|H_S\| = E_{\max}$ and $\beta_S > \beta > 0$. Then*

$$D_{-|\alpha|}(\omega_{\beta_S}(H_S)\|\omega_\beta(H_S)) \leq \frac{|\alpha|}{|\alpha|+1} V_\beta(\omega_{\beta_S}(H_S), H_S)\left[1 + \frac{E_{\max} - E_\beta}{\Delta F_{\beta_S}(\omega_\beta(H_S), H_S)}\right]. \tag{14.42}$$

The proofs for the two inequalities will be given at the end of this section. Let us now first discuss, how they can be used to prove theorem 3.6. First, we simplify the bound from the last Lemma by choosing β_S small enough so that $\Delta F_{\beta_S}(\omega_\beta(H_S), H_S) \geq E_\beta/2$. This is always possible and in the worst case increases the critical value β_{cr} when compared to the conditions for positive α. That this is possible follows from the fact that

$$\Delta F_{\beta_S}(\omega_\beta(H_S), H_S) = E_\beta - \left[E_{\beta_S} + \frac{1}{\beta_S}(S_\beta - S_{\beta_S})\right], \tag{14.43}$$

where the term in parenthesis can be made arbitrarily small by increasing β_S. We then obtain the bound

$$D_{-|\alpha|}(\omega_{\beta_S}(H_S)\|\omega_\beta(H_S)) \leq \frac{|\alpha|}{|\alpha|+1} V_\beta(\omega_{\beta_S}(H_S), H_S)\left[1 + \frac{E_{\max} - E_\beta}{\Delta F_{\beta_S}(\omega_\beta(H_S), H_S)}\right]$$

$$\leq \frac{|\alpha|}{|\alpha|+1} V_\beta(\omega_{\beta_S}(H_S), H_S)\left[1 + 2\frac{E_{\max} - E_\beta}{E_\beta}\right]$$

$$= \frac{|\alpha|}{|\alpha|+1} V_\beta(\omega_{\beta_S}(H_S), H_S) r(\beta, H_S). \tag{14.44}$$

Combining this with (14.41), we hence obtain as sufficient condition for negative values of α

$$D_{-|\alpha|}(\rho\|\omega_\beta(H)) \geq \frac{|\alpha|}{|\alpha|+1} V_\beta(\rho, H)$$

$$\geq \frac{|\alpha|}{|\alpha|+1} V_\beta(\omega_{\beta_S}(H_S), H_S) r(\beta, H_S) \geq D_{-|\alpha|}(\omega_{\beta_S}(H_S)\|\omega_\beta(H_S)), \tag{14.45}$$

which finishes the proof. Note that no correction term K is necessary for negative values of α.

Proof of Lemma 14.3. Denote the eigenvalues of the states ρ and $\omega_\beta(H)$ by p_i and w_i, respectively. We then have

$$D_{-|\alpha|}(\rho\|\omega_\beta(H)) = \frac{1}{|\alpha|+1}\log\left(\sum_i p_i^{-|\alpha|}w_i^{1+|\alpha|}\right) = \frac{1}{|\alpha|+1}\log\left(\sum_i \left(\frac{w_i}{p_i}\right)w_i\right).$$

From the concavity of the logarithm, we have $\log(\sum_i x_i w_i) \sum_i w_i \log(x_i)$ for any probability distribution w_i. Hence we obtain

$$D_{-|\alpha|}(\rho\|\omega_\beta(H)) \geq \frac{|\alpha|}{|\alpha|+1}\sum_i w_i \log\left(\frac{w_i}{p_i}\right) = \frac{|\alpha|}{|\alpha|+1}\mathcal{V}_\beta(\rho,H), \tag{14.46}$$

which finishes the proof. □

Proof of Lemma 14.4. Let us write out the Rényi-divergence as

$$D_{-|\alpha|}(\omega_{\beta_S}(H_S)\|\omega_\beta(H_S)) = \frac{|\alpha|}{|\alpha|+1}\log(Z_{\beta_S}(H_S)) - \log(Z_\beta(H_S))$$
$$+ \frac{1}{1+|\alpha|}\log\left(\mathrm{Tr}\left(e^{(\beta_S-\beta)|\alpha|H_S}e^{-\beta H_S}\right)\right). \tag{14.47}$$

We will now make use of the *log-sum inequality*. It states that for any sets of d non-negative numbers $\{a_i\}$ and $\{b_i\}$ we have

$$\log\left(\frac{a}{b}\right) \leq \sum_i \frac{a_i}{a}\log\left(\frac{a_i}{b_i}\right), \tag{14.48}$$

where $a = \sum_i a_i$ and $b = \sum_i b_i$. Let us define $\tilde{\beta}(\alpha) := \beta - (\beta_S - \beta)|\alpha|$ and choose

$$a_i := e^{-\tilde{\beta}(\alpha)E_i}, \quad b_i := \frac{e^{-\beta E_i}}{Z_\beta(H_S)}, \tag{14.49}$$

where E_i denote the eigenvalues of H_S. Note that in this case $b = 1$. We can now use the log-sum inequality to get

$$\log(a) = \log\left(\mathrm{Tr}\left(e^{(\beta_S-\beta)|\alpha|H_S}e^{-\beta H_S}\right)\right) \leq \sum_i \frac{a_i}{a}\log\left(\frac{a_i}{b_i}\right)$$
$$= \log(Z_\beta(H_S)) + (\beta_S - \beta)|\alpha|\sum_i \frac{e^{-\tilde{\beta}(\alpha)}}{Z_{\tilde{\beta}(\alpha)}(H_S)}E_i \tag{14.50}$$
$$\leq \log(Z_\beta(H_S)) + (\beta_S - \beta)|\alpha|E_{\max}. \tag{14.51}$$

Combining this bound with (14.47), we then obtain

$$D_{-|\alpha|}(\omega_{\beta_S}(H_S)\|\omega_\beta(H_S)) \leq \frac{|\alpha|}{|\alpha|+1}\left[\log(Z_{\beta_S}(H_S) - \log(Z_\beta(H_S)) + (\beta_S - \beta)E_{\max}\right].$$

We now make use again of the relation between the vacancy and the non-equilibrium free energy:

$$\mathcal{V}_\beta(\omega_{\beta_S}(H_S),H_S) = \beta_S\Delta F_{\beta_S}(\omega_\beta(H_S),H_S) = \beta_S E_\beta - S_\beta + \log(Z_{\beta_S}(H_S)). \tag{14.52}$$

Therefore we can write

$$\log(Z_{\beta_S}(H_S) - \log(Z_\beta(H_S)) + (\beta_S - \beta)E_{\max} = \mathcal{V}_\beta(\omega_{\beta_S}(H_S),H_S) + (\beta_S - \beta)(E_{\max} - E_\beta)$$
$$\leq \mathcal{V}_\beta(\omega_{\beta_S}(H_S),H_S)\left[1 + \frac{\beta_S(E_{\max} - E_\beta)}{\beta_S\Delta F_{\beta_S}(\omega_\beta(H_S),H_S)}\right]$$
$$= \mathcal{V}_\beta(\omega_{\beta_S}(H_S),H_S)\left[1 + \frac{E_{\max} - E_\beta}{\Delta F_{\beta_S}(\omega_\beta(H_S),H_S)}\right]. \tag{14.53}$$

This finishes the proof. □

14.3   What is work?

In the main-text of this chapter, I have focussed on giving intuitive explanations of the framework and discussing its implications on a physical level. I will now provide the formal definitions and proofs of all the statements. In doing so, I will stay very close to the appendix of Ref. [2], but simplify at some points. Apart from proving the results stated in the main-text, the content of this section also serves to develop a more formal notation to deal with resource theories in general and thus complements chapter 2.

14.3.1   Transitions and free transitions

Before coming to the actual proofs we need to set up some terminology. Let us consider a pair of a quantum states and a Hamiltonian $p = (\rho, H)$. In the following, I will call such pairs *objects* and denote the associated Hilbert space by $\mathcal{H}(p)$, which for most of this work is taken to be finite-dimensional. A *transition* is defined by a pair of objects $p^{(i)}, p^{(f)}$ and an ordering between them. I will refer to a transition as $p^{(i)} \to p^{(f)}$. Such transitions are to be interpreted, in the context of the present work, as changes on the system state and Hamiltonian of the battery of Arthur as implemented by Merlin.

If the Hamiltonian in a transition remains constant, that is, if $(\rho^{(i)}, H) \to (\rho^{(f)}, H)$, I will refer to it as a *state transition* and denote it simply by $\rho^{(i)} \to \rho^{(f)}$ if the Hamiltonian is clear from the context. We will often have to consider sequences of transitions, which is simply a set of $n-1$ transitions of the form $\{p^{(k)} \to p^{(k+1)}\}_{k=1}^{n-1}$. We will simply denote it by $p^{(1)} \to p^{(2)} \to \cdots \to p^{(n)}$.

For any initial object there is a set of objects that can be reached by free operations. Here, I will treat such free operations completely abstractly. To do that let us first define what is called the *free image*.

Definition 14.5 (Free image). *A free image is a function $\mathcal{F}$ that maps $p^{(i)}$ and a parameter β into sets of objects $\{p_k\} = \mathcal{F}(p^{(i)}, \beta)$. When F is such that the Hamiltonian remains constant, that is,*

$$\mathcal{F}(\rho^{(i)}, \beta) = \{(\rho_k, H)\}, \tag{14.54}$$

we will refer to it as free state-image.

A *free transition* is now simply defined as any transition $p^{(i)} \to p^{(f)}$, where $p^{(f)} \in \mathcal{F}(p^{(i)}, \beta)$. When the parameter β is clear from the context, I will denote a free transition simply as $p^{(i)} \to \mathcal{F}(p^{(i)})$.

Given two objects $p = (\rho, H)$ and $p' = (\rho', H')$, let us define their tensor product as

$$p \otimes p' := (\rho \otimes \rho', H \otimes \mathbb{I} + \mathbb{I} \otimes H'). \tag{14.55}$$

If an object based on a bipartite system of parts A and B has the form

$$p = (\rho_{AB}, H_A \otimes \mathbb{I}_B + \mathbb{I}_A \otimes H_B) \tag{14.56}$$

it is called an *non-interacting* object, since the Hamiltonian has no interaction between the two subsystems. Non-interacting objects are those objects on which a partial trace is defined in the following way.

Definition 14.6 (Partial traces). *Given any two objects $p_S = (\rho_S, H_S)$ and $p_{|S} = (\rho_{|S}, H_{|S})$, we define the trace $\mathrm{Tr}_{|S}$ as an operator acting on objects p of the form*

$$p = p_S \otimes p_{|S} = (\rho_S \otimes \rho_{|S}, H_S \otimes \mathbb{I}_{|S} + \mathbb{I}_S \otimes H_{|S}), \tag{14.57}$$

such that $\mathrm{Tr}_{|S}(p) = p_S$. This definition extends to all non-interacting *objects by the partial trace on quantum states.*

As is well known, Hamiltonians are only defined up to additive constants and, from a formal point of view, we should speak only of equivalence classes of Hamiltonians. This will be of some importance later, but I will nevertheless not incorporate it into the notation explicitly. As it will turn out, our axioms will imply that any valid work-quantifier is naturally defined on equivalence classes of Hamiltonians.

As discussed in chapter 2, resource theories contain the concept of *free objects*. In the context of thermodynamics, these are usually thermal states at some fixed temperature. In terms of the abstract notation introduced here, they are simply those objects w_β that fulfill $p \otimes w_\beta \in \mathcal{F}(p)$ for all objects p.

We can now define catalytic free transitions in an abstract way, based on the free image.

Definition 14.7 (Catalytic free image). *Given the free image $\mathcal{F}$, we define the* catalytic free image $\mathcal{F}_C$ *as*

$$\mathcal{F}_C(p^{(i)}, \beta) := \{ p \mid \exists\, q;\ p \otimes q \in \mathcal{F}(p^{(i)} \otimes q, \beta) \}. \tag{14.58}$$

A *catalytic free transition* is then defined as any transition $p^{(i)} \to p^{(f)}$ with $p^{(f)} \in \mathcal{F}_C(p^{(i)}, \beta)$. When the parameter β is clear from the context, I will denote a catalytic free transition simply as $p^{(i)} \to p^{(f)} \in \mathcal{F}_c(p^{(i)})$.

Apart from catalytic transitions, we also need a notion that captures the idea of a catalytic sequence, where the catalyst is only returned at the end of the sequence. A useful definition in this direction is that of an assisted sequence.

Definition 14.8 (Assisted transitions and sequences). *Two objects $p^{(1)}$, $p^{(2)}$ form a* transition assisted by (c_1, c_2) *if*

$$p^{(2)} \otimes c_2 \in \mathcal{F}(p^{(1)} \otimes c_1, \beta), \tag{14.59}$$

Now consider a sequence of transitions $p^{(i)} \to p^{(i+1)}$ for $i = 1, \ldots, n-1$. If each transition is a free transition assisted by (c_i, c_{i+1}), respectively, then the sequence is assisted by (c_1, c_n).

In other words, an assisted sequence is a sequence on objects that can be performed by using free operations and an auxiliary system that is at the end uncorrelated with the system but might have changed its state.

Although at the moment the definition is fully abstract, let us anticipate that $\{c_i\}_i$ are going to play the role of the fuel employed by Merlin, which he can use to enable a transition or sequence of transitions (on Arthur's work-storage device) by changing the state of the c_i. Figuratively speaking, Merlin can "burn the fuel" to be able to run a machine.

Definition 14.9 (Free sequence). *A sequence assisted by (c, c) is called a* free sequence.

In the words above, a free sequence is a sequence of transitions that can be implemented while not spending any fuel. It could also be called a catalytic free sequence. In the language of traditional thermodynamics, a free sequence corresponds to an operation which can be done by only acting with a heat bath and leaving everything else in the universe unchanged.

14.3.2 Basic assumptions on the free transitions and work-quantifiers

In the main text I have focused on the resource theory of thermodynamics, where the free operations are, loosely speaking, defined as the energy preserving joint operations on system and bath. These are mathematically characterised by Gibbs preserving maps, or strict subsets of them, such as the thermal operations. However, the results also apply to (potentially very different) resource theories defined by other classes of free operations, which are not motivated by the thermodynamic context. I will therefore continue to present the examples in a way as abstract as possible, making clear which assumptions are fundamentally necessary to derive the results.

Let us now discuss the first assumptions on the free operations that are needed in order to derive the results in the main text of this chapter.

Property 1 (Composability). *If $p^{(3)} \in \mathcal{F}(p^{(2)}, \beta)$ and $p^{(2)} \in \mathcal{F}(p^{(1)}, \beta)$, then $p^{(3)} \in \mathcal{F}(p^{(1)}, \beta)$.*

Property 2 (Swapping products). *Given an object of the form $p^{(1)} \otimes \ldots \otimes p^{(n)}$, then*

$$P(p^{(1)} \otimes \ldots \otimes p^{(n)}) \in \mathcal{F}(p^{(1)} \otimes \ldots \otimes p^{(n)}, \beta), \ \forall \beta, \tag{14.60}$$

where P permutes the labels $(1, \ldots, n)$ into $(\sigma(1), \ldots, \sigma(n))$.

Property 2 in particular implies that the identity is a catalytic free transition, that is, $p \in \mathcal{F}_C(p, \beta)$ for all β. To see this, take as catalyst $q = p$ and perform a swap between the system and the catalyst.

Property 3 (Tracing as free operation). *For any subsystem S of $A_1, \ldots, A_N$ of a product object, tracing out is in the free image. That is,*

$$\mathrm{Tr}_S(p_{A_1} \otimes \ldots \otimes p_{A_N}) \in \mathcal{F}(p_{A_1} \otimes \ldots \otimes p_{A_N}, \beta). \tag{14.61}$$

In the case where the entire system is traced out,

$$S = \cup_{i=1}^{N} A_i, \tag{14.62}$$

let us write $\mathrm{Tr}_S(p) := \varnothing$. The object $\varnothing$ can be seen as the pair $(1, 0)$ on $\mathcal{H} = \mathbb{C}$, representing the vacuum. Note that it therefore fulfills $p \otimes \varnothing = p$ for every object p. It is therefore a free object independent of β.

In the following sections, we will frequently use the following lemma, which allows us to "parallelize" sequences of transitions.

Lemma 14.10 (Mapping time to space). *Suppose $\mathcal{F}$ fulfills properties 1 and 2 and let $p \to p'$ be an assisted transition by (c, c') and $q \to q'$ be an assisted transition by (c', c''). Then the transition $p \otimes q \to p' \otimes q'$ is an assisted transition by (c, c'').*

Proof. By Definition 14.8 of assisted transition and Property 1, the transition $p_1 \otimes c \to p_m \otimes c'$ is free. Therefore, also the transition $p_q \otimes q_1 \otimes c \to p_m \otimes q_1 \otimes c'$ is free. An equivalent argument implies that $p_m \otimes q_1 \otimes c' \to p_m \otimes q_n \otimes c''$ is also free transition. Composing these two transitions yields that $p_1 \otimes q_1 \otimes c \to p_m \otimes q_n \otimes c''$ is also a free transition. $\square$

Once we have specified the transitions and the free transitions by specifying a free image, we define a quantifier of the *value* of a given transitions within the set of allowed work-storage devices $\mathcal{P}$. In the following I will always assume that the empty object $\varnothing$ is an element of $\mathcal{P}$, which simply means that having no work-storage device constitutes a valid, trivial work-storage device. This assumption is purely technical but will simplify some of the arguments later.

Definition 14.11 (Work quantifier). *A work quantifier is a function $\mathcal{W}$ that maps a transition within $\mathcal{P}$ and parameter $(p^{(i)} \to p^{(f)}, \beta)$ into the real numbers. If β is clear from the context, we will simply write $\mathcal{W}(p^{(i)} \to p^{(f)})$.*

14.3.3   General axioms

Using the mathematical definitions from the previous section, we can now present the axioms 1 and 2 of the main text more precisely.

Axiom 1 (Cyclic transitions of the work storage device). *Given a collection of objects of the work-storage device $\{p^{(1)}, \ldots, p^{(n)}\} \subset \mathcal{P}$ such that $p^{(n)} = p^{(1)}$, then*

$$\sum_{i=1}^{n-1} \mathcal{W}\left(p^{(i)} \to p^{(i+1)}, \beta\right) \geq 0. \tag{14.63}$$

Axiom 1 ensures that if a set of states can be arranged in a cyclic sequence, the total work cannot be negative. If this were not true, Arthur could, in principle, repeat the protocol an arbitrarily number of times and obtain an arbitrarily large benefit.

In the main text of this chapter, the second Axiom 2 is formulated in terms of $\mathcal{W}_{\text{trans}}$. It is possible to reformulate Axiom 2 as being directly expressed in terms of $\mathcal{W}$ for transitions of the work-storage device. This will make it a more comfortable to work with in the following proofs.

Axiom 2 (Reformulation of "cyclic transitions of the fuel"). *Let $\{p_A^{(k)} \to q_A^{(k)}\}_{k=1}^{n-1}$ be a collection of assisted transitions of the work-storage device, assisted by (c_k, c_{k+1}) respectively, with $c_n = c_1$. Then*

$$\sum_{k=1}^{n-1} \mathcal{W}(p_A^{(k)} \to q_A^{(k)}, \beta) \leq 0. \tag{14.64}$$

Importantly, the objects $p_A^{(k)}$ and $q_A^{(k)}$ $\forall k$ in this formulation describe the work-storage device, contrary to the formulation of Axiom 2 in the main text of the chapter. A very schematic depiction of the transitions involved in this Axiom is given by Fig. 14.3. This formulation of the axiom might at first seem somewhat unrelated to that of the main text. But it is indeed equivalent. Before coming to the proof of this let me state a Corollary of Axiom 2 that will be useful in further proofs.

Corollary 14.12 (Cyclic free sequences). *Let $p_A^{(1)} \to p_A^{(2)} \to \cdots \to p_A^{(n)}$ a free sequence, then,*

$$\sum_{k=1}^{n-1} \mathcal{W}(p_A^{(k)} \to p_A^{(k+1)}, \beta) \leq 0. \tag{14.65}$$

Corollary 14.12 follows simply by the definition of free sequence, which is a particular case of the conditions of Axiom 2, in the case where $q_A^{(k)} = p_A^{(k+1)}$.

Proof of equivalence between formulations: Let us now see that the version of the Axiom 2 given in the main text is indeed equivalent to the one given above. Let us first assume the version given in the main text. We thus assume that for any sequence of Merlin's system, "the fuel", where $p_M^{(1)} \to \cdots \to p_M^{(n)} = p_M^{(1)}$,

$$\sum_{i=1}^{n-1} W_{\text{trans}}(p_M^{(i)} \to p_M^{(i+1)}, \beta) \leq 0 \tag{14.66}$$
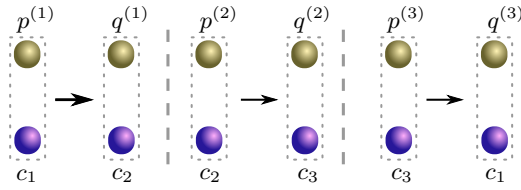
Figure 14.3: A set of transitions to illustrate the conditions of Axiom 2. Merlin holds an auxiliary system described by c_1 and receives the work-storage device in $p^{(1)}$. Both systems are initially uncorrelated. Merlin then performs a free transitions so that the final state is described by $c_2 \otimes q^{(1)}$. After this process, Arthur comes with a new work-storage device, initially uncorrelated so that Merlin holds now $c_2 \otimes p^{(2)}$. Several sequences of such transitions are performed as depicted, so that at the final step, the auxiliary system has returned to its initial state c_1. Merlin therefore did not spend any resource and did not establish any correlations. Axiom 2 then states that Merlin cannot obtain benefit (earn money) when adding up the work value of each transition. (Figure from Ref. [2].)

holds true. Now consider a set of assisted transitions $\{p_A^{(k)} \to q_A^{(k)}\}_{k=1}^{n-1}$ of the work-storage device, assisted by (c_k, c_{k+1}) respectively, with $c_n = c_1$, as in Axiom 2. Using Def. 4.1, we obtain

$$\mathcal{W}(p_A^{(k)} \to q_A^{(k)}) \leq W_{\text{trans}}(c_k \to c_{k+1}), \tag{14.67}$$

for all $k \in \{1, \ldots, n-1\}$. But then by identifying $c_i = p_M^{(i)}$ for all i, we obtain Eq. (14.64).

To show the reverse direction, we have to show that given a sequence $p_M^{(1)} \to \cdots \to p_M^{(n)} = p_M^{(1)}$, Eq. (14.64) implies Eq. (14.66). Each transition $p_M^{(k)} \to p_M^{(k+1)}$ will also induce a transition on the marginal of the work-storage device, given by $p_A^{(k)} \to q_A^{(k)} \in \mathcal{S}_k$, where $\mathcal{S}_k$ is the set of all marginal transitions on the work-storage device that can happen together with $p_M^{(k)} \to p_M^{(k+1)}$ on the fuel, and equivalently for all k. More explicitly,

$$\mathcal{S}_k := \{p_A^{(k)} \to q_A^{(k)} \mid q_A^{(k)} \otimes p_M^{(k+1)} \in \mathcal{F}_C(p_A^{(k)} \otimes p_M^{(k)})\}. \tag{14.68}$$

All the $p_A^{(k)} \to q_A^{(k)} \in \mathcal{S}_k$ are an assisted transition by $p_M^{(k)} \to p_M^{(k+1)}$. By the assumption (14.64), this implies that the total work-value fulfills

$$\sum_{k=1}^{n-1} \mathcal{W}(p_A^{(k)} \to q_A^{(k)}) \leq 0 \tag{14.69}$$

for all $p_A^{(k)} \to q_A^{(k)} \in \mathcal{S}_k$. This in turn implies

$$\sum_{k=1}^{n-1} \sup_{p_A^{(k)} \to q_A^{(k)} \in \mathcal{S}_k} \left(\mathcal{W}(p_A^{(k)} \to q_A^{(k)}) \right) \leq 0. \tag{14.70}$$

But the supremum in Eq. (14.70) is the same as the one in the Def. 4.1 of W_{trans}, which concludes the proof.

14.3.4   Implications for the work definition

We can now discuss the implications of the Axioms for the work quantifiers. Since the two Axioms are now re-formulated in such a way that they refer to objects of the work-storage device only, we can drop the labels M and A. Unless explicitly mentioned, the letters p, q therefore now always refer to the work-storage device.

Lemma 14.13 (Properties of work quantifiers). *Consider a free image $\mathcal{F}$ fulfilling Properties 1-3. In this case, Axioms 1 and 2 are fulfilled if and only if $\mathcal{W}$ satisfies the following properties,*

1. For all $p^{(1)}, \ldots, p^{(m)}$ and $q^{(1)}, \ldots, q^{(m)}$ in $\mathcal{P}$ such that $\bigotimes_{i=1}^m q^{(i)} \in \mathcal{F}_C(\bigotimes_{i=1}^m p^{(i)})$,

$$\sum_{i=1}^m \mathcal{W}(p^{(i)} \to q^{(i)}) \leq 0. \tag{14.71}$$

2. For all $p, q, r \in \mathcal{P}$

$$\mathcal{W}(p \to q) = -\mathcal{W}(q \to p), \tag{14.72}$$

$$\mathcal{W}(p \to q) + \mathcal{W}(q \to r) = \mathcal{W}(p \to r). \tag{14.73}$$

Proof. Let us first see how the properties follow from the Axioms, beginning with properties (14.72) and (14.73). The two properties follow once we have shown that any cyclic sequence $p_1 \to p_2 \to \ldots \to p_n = p_1$ has a total work-value equal to zero. Given Axiom 1, which already implies that the total work-value of a cyclic sequence is larger than zero, this only requires to show that such a sequence has a work-value smaller or equal to zero. We can do this by showing that any cyclic sequence is a free sequence, which is enough to show the claim given Corollary 14.12. Consider the catalyst $c_1 = c_n := c$ given by $c = \otimes_{i=2}^{n-1} p_i$. To see that c assists any cyclic sequence from p_1 to $p_n = p_1$, consider the object $p_1 \otimes c = \otimes_{i=1}^{n-1} p_i$. By swapping, which is a free operation, we arrive at state $p_2 \otimes c'$ with $c' = p_1 \otimes p_3 \otimes p_4 \otimes \cdots \otimes p_{n-1}$. Repeating the swapping sequentially we see that the first system goes through the transitions $p_1 \to p_2 \to \cdots \to p_{n-1}$. Applying a final swap the fuel is returned to c and the system returns to object p_1, proving the claim and thus, Eqs. (14.72) and (14.73).

Let us now show property (14.71) from the axioms. The premise of (14.71) is that there exists a catalytic free transition $\otimes_{i=1}^{n} p_i \to \otimes_{i=1}^{n} q_i$ (taking $m = n$ without loss of generality; the other cases follow by tensoring a suitable number of empty objects $\varnothing$). Then the transition $p_1 \to q_1$ is a transition assisted by

$$\left(c_1 = \bigotimes_{i=2}^{n} p_i, \ c_1' = \bigotimes_{i=2}^{n} q_i \right). \tag{14.74}$$

Secondly, the transition $p_2 \to q_2$ is an transition assisted by $(c_2 = \otimes_{i=2}^{n} q_i, \ p_2 \otimes_{i=3}^{n} q_i)$. This can be seen by just performing a swap between the work-storage system in p_2 and the first element of the fuel in q_2. An equivalent swapping can be used to show that $p_j \to q_j$ is a transition assisted by

$$\left(c_j = \bigotimes_{i=2}^{j-1} p_i \bigotimes_{k=j}^{n} q_k, \ c_j' = \bigotimes_{i=2}^{j} p_i \bigotimes_{k=j+1}^{n} q_k \right) \tag{14.75}$$

for $j = 3, \ldots, n-1$. Lastly, $p_n \to q_n$ is assisted by $(c_n = \otimes_{i=2}^{n-1} p_i \otimes q_n, c_n' = \otimes_{i=2}^{n} p_i)$. Altogether, this implies that the set of sequences $\{p_i \to q_i\}_{i=1}^{n}$ can be each performed with free operations assisted by (c_i, c_i') as described previously. Note that $c_i' = c_{i+1}$ and $c_1 = c_n'$, hence, it meets the conditions of Axiom 2 which by Eq. (14.64) implies

$$\sum_{i=1}^{n} \mathcal{W}(p_i \to q_i) \leq 0.$$

Finally, let us show that the properties (14.71)-(14.73) imply the axioms. Axiom 1 is trivially satisfied since properties (14.72) and (14.73) imply that for any cyclic sequence the total amount of work is zero. Let us move to Axiom 2, which has as a premise that one has $n-1$ assisted transitions $p^{(j)} \to q^{(j)}$, assisted by (c_j, c_{j+1}) with $j = 1, \ldots, n-1$ and $c_n = c_1$. Then, we can use Lemma 14.10 and see that the transition

$$\bigotimes_{j=1}^{n-1} p^{(j)} \to \bigotimes_{j=1}^{n-1} q^{(j)} \tag{14.76}$$

is a transition assisted by $(c_1, c_n = c_1)$. Thus the system c is returned unchanged, hence $\otimes_{j=1}^{n-1} p^{(j)} \to \otimes_{j=1}^{n-1} q^{(j)}$ is indeed a catalytic free transition and (14.71) implies that

$$\sum_{j=1}^{n-1} \mathcal{W}(p^{(j)} \to q^{(j)}) \leq 0, \tag{14.77}$$

proving (14.64) and, as a consequence, Axiom 1. □

With this result at our disposal we can now show that Axioms 1 and 2 (equivalently Eqs. (14.71)-(14.73)) imply very specific properties of the function $\mathcal{W}$.

Theorem 14.14 (theorem 4.2 in the main text). *Given a free image $\mathcal{F}$ that fulfils Properties 1-3, the function $\mathcal{W}$ fulfils Axioms 1 and 2 if and only if it can be written as*

$$\mathcal{W}(p \to q) = M(q) - M(p), \tag{14.78}$$

for a function M such that $M(\varnothing) = 0$ and that fulfils the following property:

Additive monotonicity: *For all $p^{(1)}, \ldots, p^{(n)}$ and $q^{(1)}, \ldots, q^{(n)}$ in $\mathcal{P}$ such that $\otimes_{i=1}^{n} q^{(i)} \in \mathcal{F}_C(\otimes_{i=1}^{n} p^{(i)})$*

$$\sum_{i=1}^{n} M(q^{(i)}) \leq \sum_{i=1}^{n} M(p^{(i)}). \tag{14.79}$$

Proof. We show this by showing that the above conditions are equivalent to the conditions (14.71)-(14.73), which in turn are equivalent with Axioms 1 and 2. Consider the function $M(p) := \mathcal{W}(\varnothing \to p)$. By properties (14.72) and (14.73) we have

$$\mathcal{W}(p \to q) = M(q) - M(p) \tag{14.80}$$

and $M(\varnothing) = 0$ is true by definition. Then (14.71) is fulfilled if and only if additive monotonicity holds for M. $\square$

14.4   Uniqueness of the relative entropy and free energy

14.4.1   Rank-decreasing quantum channels

Here, I provide the proof of Lemma 5.3 as presented in Ref. [7]. We have to show that given a channel $T : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ and a full-rank state σ such that $\mathrm{supp}(T(\sigma)) \subset P$, we also have $\mathrm{supp}(T(\rho)) \subset P$ for all states ρ. Here, P is an arbitrary subspace of the total Hilbert space $\mathcal{H}'$. Let $\sigma = \sum_i q_i |i\rangle\langle i|$ be the eigen-decomposition of σ. Since T maps positive operators to positive operators, and the support of the sum of positive operators is the union of the supports of the operators we conclude that $T(|i\rangle\langle i|)$ is supported in P for all i. We thus only need to show that also operators of the form $T(|i\rangle\langle j|)$ are supported on P. Now consider any density operator $\rho = d + r$ where d is the diagonal part of ρ (in the eigenbasis of σ) and r the rest. We know that $\mathrm{Tr}(T(d)) = 1$ since T is trace-preserving. Hence $\mathrm{Tr}(T(r)) = 0$. Let us now assume (to arrive at a contradiction) that $T(r)$ has support within the subspace $Q = \mathbf{1} - P$. Since T maps positive operators to positive operators,

$$0 \le QT(\rho)Q = QT(r)Q. \tag{14.81}$$

Thus we conclude on the one hand that $QT(r)Q \ge 0$. On the other hand, we know that

$$1 = \mathrm{Tr}(T(\rho)) \ge \mathrm{Tr}(PT(\rho)) = 1 + \mathrm{Tr}(PT(r)). \tag{14.82}$$

Hence, $\mathrm{Tr}(P(T(r))) = 0$. Since T is trace-preserving we also have

$$\mathrm{Tr}(PT(r)) = -\mathrm{Tr}(QT(r)) = 0. \tag{14.83}$$

Hence $QT(r)Q = 0$ and also $QT(\rho)Q = 0$. By positivity and Hermiticity of $T(\rho)$ we also get $PT(\rho)Q = 0$ and $QT(\rho)P = 0$. We thus conclude that $T(\rho) = PT(\rho)P$, which finishes the proof.

14.5 Statistical ensembles from thermodynamic operations

In this section, I provide the proofs for the main results in chapter 6.

14.5.1 Distilling thermal states from the environment

The basic results that will allow us to prove our main result is that if we take many copies of macro-states from the environment, we can distill a thermal microstate to arbitrary accuracy. Furthermore, this can be done using a unitary that exactly preserves energy. In this section we will prove this result. In the following sections we will use this result to establish that the only valid choice for the functio $H \mapsto e_\beta(H)$ is the one that assigns the energy of the Gibbs state and to prove theorem 6.5.

The proof that one can distill thermal microstates from copies of macrostates rests on two steps. First, we show that we can use macrostates from the environment to implement random unitaries . Then we show that we can distill thermal microstates using random unitaries.

Random energy-preserving unitaries   To see how to implement random energy-preserving unitaries, we provide the following simple protocol. The idea is that we can use thermal systems from the environment and apply unitaries conditioned on energy measurements on these systems. To implement this idea, take from the environment M two-level systems with gap Δ and energy $e_\beta(\Delta)$ for the fixed function $e_\beta(\Delta)$. I.e., we assume that the Hamiltonian is given by $H_\Delta = \Delta\,|1\rangle\langle 1|$. For a two-level system, the excitation and ground state probabilities are determined uniquely from the expectation value $e_\beta(\Delta)$:

$$p_1 := \mathrm{Tr}(\,|1\rangle\langle 1|\,\rho) = \mathrm{Tr}(H_\Delta\rho)/\Delta, \tag{14.84}$$
$$p_0 := \mathrm{Tr}(\,|0\rangle\langle 0|\,\rho) = 1 - \mathrm{Tr}(H_\Delta\rho)/\Delta. \tag{14.85}$$

Therefore, we have $p_1 = e_\beta(\Delta)/\Delta$ and $p_0 = 1 - e_\beta(\Delta)/\Delta$. If we take many copies M, the probabilities for the energy-levels on the joint-system (labelled by bit-strings 0110011 etc.) will become arbitrarily dense on the interval $[0, 1]$. For any probability distribution q_j over K elements and any error ϵ, there is therefore a number M such that we can then find a partition $\mathcal{P}$ of the bit-strings (basis states of the M copies) into K subsets $\mathcal{P}_j$ such that

$$\left|\mathrm{Tr}\left(P_j(e_\beta(\Delta), H_\Delta)^{\otimes M}\right) - q_j\right| < \epsilon, \tag{14.86}$$

where P_j is the projector onto the bit-strings in $\mathcal{P}_j$:

$$P_j = \sum_{\substack{s \in \{0,1\}^{\times M}, \\ s \in \mathcal{P}_j}} |s\rangle\langle s|. \tag{14.87}$$

Now suppose we want to implement the unitaries U_j with probability q_j on a *different* system H_E. Then we can do this (to arbitrary accuracy) by implementing the unitary

$$V = \sum_j P_j \otimes U_j \tag{14.88}$$

on the system $(e_\beta(\Delta), H_\Delta)^{\otimes M} \otimes (e_\beta(H_E), H_E)$. Finally, provided that the unitaries U_j commute with the Hamiltonian H_E the unitary V commutes with the total Hamiltonian and hence preserves the energy. In particular, the M two-level systems as well as the system with Hamiltonian H_E will remain in the same macrostate.

Distilling thermal states   Having shown how to implement random energy-preserving unitaries we can now proceed to discuss how to distill thermal states from the macrostate environment. Suppose we want to distill a single thermal microstate of the Hamiltonian H_E. We will do this using the following steps:

1. Take a large number N of copies of $(e_\beta(H_E), H_E)$.
2. Use a random energy-preserving unitary to dephase the density matrix in the energy-eigenbasis.
3. In each (highly degenerate) energy-eigenspace of the N copies of H_E perform a mixture of unitaries that changes the global density matrix to the maximally mixed state in this energy-eigenspace. If we label the energy-eigenspaces by α, with being $P_\alpha^{(N)}$ being the projector onto the energy-eigenspace with energy $E_\alpha^{(N)}$, this brings the density matrix to a state of the form

$$\Omega^{(N)} = \sum_\alpha q_\alpha^{(N)} \frac{P_\alpha^{(N)}}{d_\alpha^{(N)}}, \tag{14.89}$$

where $d_\alpha^{(N)}$ is the dimension of the subspace $P_\alpha^{(N)}$.

4. We prove that the reduced state of Ω on any of the copies is close to the Gibbs state of H_E with energy $e_\beta(H_E)$.

The only point left to prove is the last point. To see this point, first note that since we assumed that the many copies are uncorrelated, the global energy distribution of the N copies is highly peaked around the mean $Ne_\beta(H_E)$. This is true despite the fact that the systems might all be in different microstates. More precisely, from Hoeffding's inequality it follows that the total probability to find the system with an energy that deviates from $Ne_\beta(H_E)$ by more than $O(\sqrt{N})$ is exponentially small in N. Let us collect thus collect the energies that differ from $Ne_\beta(H_E)$ by less than $c\sqrt{N}$ in the subspace $\mathcal{M}^{(N)}$, where $c > 0$ is some constant. Then we have

$$\Omega^{(N)} = \sum_{\alpha \in \mathcal{M}^{(N)}} q_\alpha^{(N)} \frac{P_\alpha^{(N)}}{d_\alpha^{(N)}} + \epsilon^{(N)}\sigma, \tag{14.90}$$

where σ is some density matrix and $\epsilon^{(N)}$ decreases exponentially with N. Note that $E_{\alpha_N}^{(N)}/N \to e_\beta(H_E)$ as $N \to \infty$ for any sequence $\alpha_N \in \mathcal{M}^{(N)}$.

Each of the states $P_\alpha^{(N)}/d_\alpha^{(N)}$ is a microcanonical state on N copies of the same system, with the energy-density $e_\beta(H_E)$ as N *diverges*. Therefore, the standard arguments from statistical mechanics that show how the canonical state follows from the microcanonical ensemble suffice to show its local reduction on any of the identical N copies is arbitrary close to the Gibbs-state at energy $E_\alpha^{(N)}/N$ (see Ref. [11] for a more detailed discussion of the error in the approximation).

Since $E_\alpha^{(N)}/N \to e_\beta(H_E)$ as $N \to \infty$ the reduced states of the $P_\alpha^{(N)}/d_\alpha^{(N)}$ converge to the Gibbs state at energy $e_\beta(H_E)$, i.e., at temperature β. Note that, while the local marginals $\mathrm{Tr}_{N-1}(\Omega^{(N)})$ are close to Gibbs states, the global state $\Omega^{(N)}$ is highly correlated and hence we cannot expect it to be close to a Gibbs state. In particular, in principle we have to repeat this process for any uncorrelated copy of the Gibbs state $\omega_\beta(H_E)$ that is required in the protocol.

A similar argument can be made if a set of commuting conserved quantities is fixed instead of only the Hamiltonian, but requires more care in the details of the derivation (see Ref. [11] for details). In this case one obtains that the reduced state of the corresponding state $\Omega^{(N)}$ is a Generalized Gibbs ensemble.

14.5.2   Proof of theorem 6.5

The main-result theorem 6.5 can be shown very easily using the distillation protocol above. Note that to prove the theorem, it is sufficient to show that we can map the macrostate (e, H) to the microstate $\omega_{\beta(e)}(H)$. To do this, distill from the environment a thermal state of the re-scaled Hamiltonian

$$H_E = \frac{\beta(e)}{\beta}H. \tag{14.91}$$

Then, we simply swap the state on the system with the one that we distilled from the environment. The final state on the system is then given by

$$\omega_\beta(\beta(e)/\beta\, H) = \omega_{\beta(e)}(H). \tag{14.92}$$

The energy before the swap is given by:

$$E_{\text{before}} = e + \mathrm{Tr}\left(\omega_\beta(\beta(e)/\beta\, H)H\right)\beta(e)/\beta = e + \mathrm{Tr}\left(\omega_{\beta(e)}(H)H\right)\beta(e)/\beta = e(1 + \beta(e)/\beta). \tag{14.93}$$

The energy after the swap is given by

$$E_{\text{after}} = \mathrm{Tr}\left(\omega_\beta(\beta(e)/\beta\, H)H\right) + e\beta(e)/\beta = \mathrm{Tr}\left(\omega_{\beta(e)}(H)H\right) + e\beta(e)/\beta = E_{\text{before}}. \tag{14.94}$$

The operations is thus a valid macrostate operation and maps the macrostate (e, H) to the microstate $\omega_{\beta(e)}(H)$. Further note that this operation would in general be impossible using unitary operations that preserve the energy on the level of probability distributions: While we could still distill the thermal state $\omega_\beta(\beta(e)/\beta H)$, swapping it with the microstate on the system would in general not preserve the probability distribution of energy, but only its expectation value.

14.5.3   Non-Gibbsian energies trivialize the resource theory

Let us now use the distillation protocol to sketch the proof-idea for the following result:

Result 14.15. *Assume that $e_\beta(H_E)$ neither assigns the ground state energy nor the energy of the highest excited state to the system. Then if the function $H_E \mapsto e_\beta(H_E)$ cannot be expressed as $H_E \mapsto \mathrm{Tr}(\omega_{\beta'}(H_E)H_E)$ for some β', any macrostate (e, H) can be transformed to any other valid macrostate (e', H) by a macrostate-to-macrostate operation.*

Any function $e_\beta(H_E)$ that fulfills the assumption of result 14.15 can be expressed as

$$e_\beta(H_E) = \mathrm{Tr}\left(\omega_{\tilde{\beta}(H_E)}(H_E)\right), \tag{14.95}$$

for some function $\tilde{\beta}(H_E)$ that assigns inverse temperatures to Hamiltonians. Suppose now, that $e_\beta(H_E)$ is not of the form $\mathrm{Tr}\left(\omega_{\beta'}(H_E)\right)$ for some fixed β'. Then there are at least two Hamiltonians $H_E^{(1)}$ and $H_E^{(1)}$ such that $\tilde{\beta}(H_E^{(1)}) \neq \tilde{\beta}(H_E^{(2)})$. Let us introduce the short-hands $\beta_1 = \tilde{\beta}(H_E^{(1)})$ and $\beta_2 = \tilde{\beta}(H_E^{(2)})$. Using the distillation protocol form the last subsection, we can then distill for free states of the form

$$\left(\omega_{\beta_1}(H_E^{(1)}) \otimes \omega_{\beta_2}(H_E^{(2)})\right)^{\otimes N}, \tag{14.96}$$

where N is arbitrary. What we have at our hand then, are two thermal heat baths of arbitrary size at different temperatures from which we can extract as much work as we wish. We can then use this work to bring any system into any macrostate. In our framework, this works by introduce a further system $\omega_{\tilde{\beta}(H_E)}(H_E)$ from the environment and applying a global, energy-preserving unitary operation to bring this system to any macrostate that we want. Using the results of Ref. [80], we can even prepare any *microstate*.

14.6   Thermal machines under control restrictions

14.6.1   Work in isothermal process

In this section, we proof that the work in an optimal isothermal process between two Hamiltonians H_0 and H_1 is given by $F_\beta(\omega_\beta(H_0), H_0) - F_\beta(\omega_\beta(H_1), H_1)$. It has already been shown in section 9.1.3 that the work in an isothermal process of N steps can be written as

$$
\begin{aligned}
W_N(0 \to 1) &= \sum_{i=1}^{N} W(t_{i-1} \to t_i) \\
&= \sum_{i=1}^{N} \mathrm{Tr} \left(\omega_\beta(H_{t_{i-1}}) \left(H_{t_{i-1}} - H_{t_i} \right) \right),
\end{aligned} \tag{14.97}
$$

where H_{t_i} with $t_i = i/N$ are the Hamiltonians along the discretization of the smooth path H_t that connects H_0 with H_1. We will now follow the strategy in Ref. [235]: In the limit $N \to \infty$, we can replace the sum by the integral

$$
\begin{aligned}
W(0 \to 1) &= \lim_{N \to \infty} W_N(0 \to 1) \\
&= - \int_0^1 \mathrm{Tr} \left(\omega_\beta(H_t) \frac{\partial H_t}{\partial t} \right) \, \mathrm{d}t.
\end{aligned} \tag{14.98}
$$

To find this integral, consider the derivative of the free energy:

$$
\frac{\partial}{\partial t} F_\beta(\omega_\beta(H_t), H_t) = -\frac{\partial}{\partial t} \frac{1}{\beta} \log \left(\mathrm{Tr} \left(\mathrm{e}^{-\beta H_t} \right) \right) = -\frac{1}{\beta} \frac{\mathrm{Tr} \left(\frac{\partial}{\partial t} \mathrm{e}^{-\beta H_t} \right)}{\mathrm{Tr} \left(\mathrm{e}^{-\beta H_t} \right)} \tag{14.99}
$$

$$
= \mathrm{Tr} \left(\omega_\beta(H_t) \frac{\partial H_t}{\partial t} \right). \tag{14.100}
$$

We thus find

$$
W(0 \to 1) = F_\beta(\omega_\beta(H_0), H_0) - F_\beta(\omega_\beta(H_1), H_1). \tag{14.101}
$$

14.6.2   Proof of theorem 9.1

In this section, I prove the upper bound in theorem 9.1. Consider any cyclic protocol $\mathcal{P}$ starting with Hamiltonian $H = H_{t_0}$ and state $\rho = \rho_{t_0}$. Let the protocol consist of N steps and denote the quantum state before the i-th thermal contact by ρ_{t_i} and the state after the i-th thermal contact by $\sigma_{t_i} = \mathcal{G}_{t_i}(\rho_{t_i})$. Since the protocol is cyclic, $H_{t_N} = H$. The states ρ_{t_i} fulfill

$$
\rho_{t_i} = U_i \sigma_{t_{i-1}} U_i^\dagger, \ \sigma_{t_0} = \rho_{t_0} = \rho.
$$

The total extracted work in the protocol is then given by

$$
\begin{aligned}
W(\mathcal{P}, \rho, H) &= \sum_{i=1}^{N} \left[\mathrm{Tr} \left(H_{t_{i-1}} \sigma_{t_{i-1}} \right) - \mathrm{Tr} \left(H_{t_i} \rho_{t_i} \right) \right] = \sum_{i=1}^{N} F_\beta(\sigma_{t_{i-1}}, H_{t_{i-1}}) - F_\beta(\rho_{t_i}, H_{t_i}) \\
&= F_\beta(\sigma_{t_0}, H_{t_0}) - F_\beta(\rho_{t_N}, H_{t_N}) - \sum_{i=1}^{N-1} \left(F_\beta(\rho_{t_i}, H_{t_i}) - F_\beta(\sigma_{t_i}, H_{t_i}) \right).
\end{aligned} \tag{14.102}
$$

Since $\sigma_{t_i} = \mathcal{G}_{t_i}(\rho_{t_i})$ and $\mathcal{G}_{t_i}$ is Gibbs-preserving with respect to H_{t_i} we obtain from the monotonicity of the free energy under Gibbs-preserving maps that

$$
F_\beta(\rho_{t_i}, H_{t_i}) \geq F_\beta(\sigma_{t_i}, H_{t_i}). \tag{14.103}
$$

Furthermore, by the extremality of the Gibbs state, we have $F_\beta(\rho_{t_N}, H) \geq F_\beta(\omega_\beta(H), H)$. Therefore we get the final upper bound

$$
\begin{aligned}
W(\mathcal{P}, \rho, H) &\leq F_\beta(\sigma_{t_0}, H_{t_0}) - F_\beta(\rho_{t_N}, H_{t_N}) \leq F_\beta(\rho, H) - F_\beta(\omega_\beta(H), H) \\
&= \Delta F_\beta(\rho, H).
\end{aligned} \tag{14.104}
$$

Figure 14.4: The curves g (in blue, see Eq. (14.114)) and f (in orange, see Eq. (14.115)) illustrating the thermo-majorization condition for the example of breakdown of universality of thermalizing contacts in the case of two spins and local control. Since the orange curve (corresponding to the thermal state of H_s for $s < s_c$) lies below the blue curve (corresponding to the maximally mixed state) the transition from the maximally mixed state to this state is possible using a thermal operation. The green curve (identity function) corresponds to the thermal state of H_0. (Figure adapted from Ref. [1].)
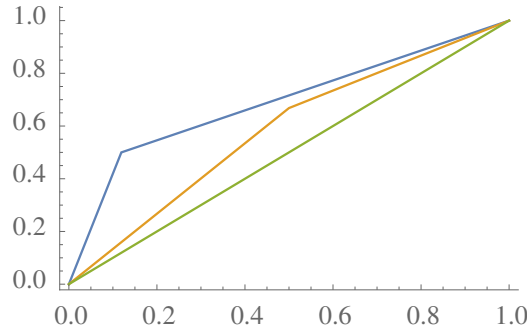
14.6.3   Proof of Lemma 9.2

In this section, I prove the upper bound in Lemma 9.2. Consider any protocol $\mathcal{P}$ that employs thermalizing maps as thermal contacts only. First note that it is always sub-optimal to start the protocol with a thermal contact, since then no work can be extracted anymore. Similarly, we can always assume that the protocol ends in the thermal state. Then subdivide the protocol into two parts: The first adiabatic evolution (with unitary U and ending with Hamiltonian H_t) and the rest. The total work is then given by

$$W(\mathcal{P}, \rho, H) = \mathrm{Tr}(\rho H) - \mathrm{Tr}(\sigma H_t) + W((\sigma, H_t) \to (\omega_\beta(H), H)), \qquad (14.105)$$

where $U\rho U^\dagger = \sigma$ and $W((\sigma, H_t) \to (\omega_\beta(H), H))$ denotes the work after the first adiabatic evolution. By essentially the same calculation as in section 9.1.3, the latter part of the work is upper bounded as

$$W((\sigma, H_t) \to (\omega_\beta(H), H)) \le F_\beta(\sigma, H_t) - F_\beta(\omega_\beta(H), H). \qquad (14.106)$$

Using that the von Neumann entropy is unitarily invariant, using extremality of the Gibbs state and plugging the above inequality into (14.105), we then get

$$\begin{aligned} W(\mathcal{P}, \rho, H) &= F_\beta(\rho, H) - F_\beta(\sigma, H_t) + W((\sigma, H_t) \to (\omega_\beta(H), H)) \\ &\le F_\beta(\rho, H) - F_\beta(\omega_\beta(H_t), H_t) + W((\sigma, H_t) \to (\omega_\beta(H), H)) \\ &\le F_\beta(\rho, H) - F_\beta(\omega_\beta(H), H) - \left[F_\beta(\sigma, H_t) - F_\beta(\omega_\beta(H_t), H_t) \right]. \end{aligned} \qquad (14.107)$$

Optimizing over H_t and $\sigma \in \mathcal{U}_\mathcal{H}[\rho]$ yields the final upper bound.

14.6.4   Proof of Lemma 9.3

Consider any two Hamiltonians H_1 and H_2. Then

$$\mathrm{Tr}(\omega_\beta(H_1) \log \omega_\beta(H_2)) = -\beta \mathrm{Tr}(\omega_\beta(H_1) H_2) - \log Z_\beta(H_B).$$

The relative entropy between the thermal states of H_1 and $H_1 + H_2$ is always positive and (using the above equation) given by

$$\begin{aligned} 0 &\le D(\omega_\beta(H_1) \| \omega_\beta(H_1 + H_2)) \\ &= -\beta \mathrm{Tr}(\omega_\beta(H_1) H_1) - \log Z_\beta(H_1) + \beta \mathrm{Tr}(\omega_\beta(H_1)(H_1 + H_2)) + \log Z_\beta(H_1 + H_2). \end{aligned}$$

Cancelling terms and recognising that $F_\beta(\omega_\beta(H), H) = -\frac{1}{\beta} \log Z_\beta(H)$ finishes the proof:

$$F_\beta(H_1 + H_2) \le F_\beta(H_1) + \mathrm{Tr}(\omega_\beta(H_1) H_2). \qquad (14.108)$$

14.6.5   Thermo-majorization argument for the example of two spins

In this section I show that the maximally mixed state can be mapped to the thermal state of the Hamiltonian $H_s = H_0 + s\mathbf{1} \otimes \sigma_z$ (see section 9.3.3) using a thermal operation. To show this, I will use the technique of thermo-majorization. To simplify the notation discussion, let us assume $\beta = 1$ without loss of generality

Denote the (ordered) vectors of eigenvalues of the Gibbs state of H_0 and H_s by

$$\omega_0 = \frac{1}{2(1 + e^2)}(e^2, e^2, 1, 1), \omega_s = \frac{1}{2(1 + e^2)}(e^2 f^+(s), e^2 f^-(s)), f^+(t), f^-(s)), \qquad (14.109)$$

respectively, with $f^{\pm}(s) = 1 \pm \tanh(s)$. Furthermore, denote the vector of eigenvalues of the maximally mixed state by $p = (1, 1, 1, 1)/4$. To evaluate the thermo-majorization condition we have to order the vectors with entries $r_i = (\omega_s)_i/(\omega_0)_i$ and $r'_i = p_i/(\omega_0)_i$ in non-increasing order. Let σ, σ' be the permutations that do this, i.e.,

$$r_{\sigma(1)} \geq \cdots \geq r_{\sigma(4)} \tag{14.110}$$

and similarly for p and σ'. The vectors r, r' are given by

$$r = (1 + \tanh(s), 1 - \tanh(s), 1 + \tanh(s), 1 - \tanh(s)), \tag{14.111}$$

$$r' = \frac{1 + e^2}{2}(e^{-2}, e^{-2}, 1, 1). \tag{14.112}$$

Thus, one possible choice of permutations is given by

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}, \quad \sigma' = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}. \tag{14.113}$$

Now construct the curve of straight lines g that connects the points with coordinates

$$\left(\sum_{j=1}^{k} (\omega_0)_{\sigma'(j)}, \sum_{j=1}^{k} p_{\sigma'(j)} \right), \quad k = 1, \ldots, 4. \tag{14.114}$$

Similarly, construct the curve of straight lines f that connects the points with coordinates

$$\left(\sum_{j=1}^{k} (\omega_0)_{\sigma(j)}, \sum_{j=1}^{k} (\omega_s)_{\sigma(j)} \right), \quad k = 1, \ldots, 4. \tag{14.115}$$

The thermo-majorization condition says that the maximally mixed state (represented by the curve g) can by mapped by a thermal operation to the thermal state of the Hamiltonian H_s if and only if $g(x) \geq f(x)$ for all x.

The condition clearly holds for the first points. For the second point, a simple calculation shows that

$$g(1/2) = \frac{1}{2}\left(1 + \frac{e^2 - 1}{2e^2}\right) \tag{14.116}$$

while

$$f(1/2) = \frac{1}{2}(1 + \tanh(s)), \tag{14.117}$$

which yields the condition

$$s \leq s_c := \tanh^{-1}\left(\frac{e^2 - 1}{2e^2}\right) \simeq 0.46. \tag{14.118}$$

There are no further constraints (see Fig. 14.4) and the proof is finished.

14.6.6   Proof of theorem 9.5

In this section I provide the proof of theorem 9.5. To do that, let us first define more precisely a "Carnot-like" protocol. In the following, a Carnot-like protocol will be any protocol that consists of i) a series of n_h consecutive thermal contacts with the hot bath and possible adiabatic evolution between the contacts; ii) after that n_c consecutive thermal contacts with the cold bath, again with possible adiabatic evolutions in-between; iii) a final adiabatic evolution to return the Hamiltonian back to its initial value. Both a Carnot-like protocol and a protocol that is not Carnot-like are schematically depicted in Figure 14.5.

From Figure 14.5, it seems intuitively clear that Carnot-like protocols have higher efficiency than other protocols since they minimize the number of dissipation-events. Similarly, to every protocol $\mathcal{P}$ we can associate Carnot-like protocols by considering a subset of the involved Hamiltonians. Every such constructed Carnot-like protocol has higher efficiency than $\mathcal{P}$.

We will prove all of this below. Before that, it is instructive to calculate the efficiency of the Carnot-like protocol. Let us assume (without loss of generality) that we start thermalized with the cold bath in Hamiltonian $H^{(4)}$. We then do an adiabatic evolution to the Hamiltonian $H^{(1)}$. After that we do the first thermal contact to the hot heat bath and then alternate between adiabatic evolutions and thermal contacts until we arrive at Hamiltonian $H^{(2)}$ in the state $\omega_{\beta_h}(H^{(2)})$ in $n_h - 1$ steps.
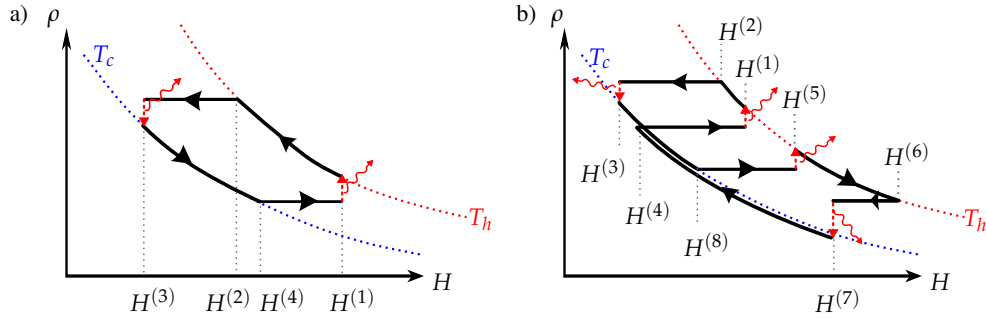
Figure 14.5: (a) Carnot-like protocol; (b) protocol that is not Carnot-like. Red, dashed arrows symbolize thermalization-events and the wiggly arrows symbolize corresponding dissipation. (Figure adapted from Ref. [6].)

From essentially the same calculation as in the proof of Lemma 9.2, we obtain that the work in this process is bounded as

$$W(4 \to 1) \le F_{\beta_h}(\omega_{\beta_c}(H^{(4)}), H^{(4)}) - F_{\beta_h}(\omega_{\beta_h}(H^{(2)}), H^{(2)}) - T_h D(U\omega_{\beta_c}(H^{(4)})U^\dagger \| \omega_{\beta_h}(H^{(1)}))$$
(14.119)

$$=: F_{\beta_h}(\omega_{\beta_c}(H^{(4)}), H^{(4)}) - F_{\beta_h}(\omega_{\beta_h}(H^{(2)}), H^{(2)}) - T_h C_h^{(1)}$$
(14.120)

The bound is achieved in an adiabatic evolution followed by an isothermal process. This protocol also lower bounds the absorbed heat from the hot bath. From the first law of thermodynamics we find that this heat is upper bounded by

$$Q_h(4 \to 1) \le T_h \left(S(\omega_{\beta_h}(H^{(2)})) - S(\omega_{\beta_c}(H^{(4)})) \right) - T_h C_h^{(1)}$$
(14.121)

$$=: T_h \left(S_h^{(2)} - S_c^{(4)} \right) - T_h C_h^{(1)}.$$
(14.122)

After the contacts with the hot bath, we then do an adiabatic evolution to $H^{(3)}$, followed by n_c thermal contacts with the cold bath, with adiabatic evolutions in-between, back to the initial Hamiltonian $H^{(4)}$. Completely analogous to the work in the first part, the work done in this part is upper bounded as

$$W(1 \to 4) \le F_{\beta_c}(\omega_{\beta_h}(H^{(2)}), H^{(2)}) - F_{\beta_c}(\omega_{\beta_c}(H^{(4)}), H^{(4)}) - T_c D(V\omega_{\beta_h}(H^{(2)})V^\dagger \| \omega_{\beta_c}(H^{(3)}))$$
(14.123)

$$=: F_{\beta_c}(\omega_{\beta_h}(H^{(2)}), H^{(2)}) - F_{\beta_c}(\omega_{\beta_c}(H^{(4)}), H^{(4)}) - T_c C_c^{(3)}$$
(14.124)

and the heat from the cold bath is bounded as

$$Q_c(1 \to 4) \le -T_c \left(S_h^{(2)} - S_c^{(4)} \right) - T_c C_c^{(3)}.$$
(14.125)

The total work is then given by

$$Q_h(4 \to 1) + Q_c(1 \to 4) = W(\mathcal{P}) \le (T_h - T_c) \left(S_h^{(2)} - S_c^{(4)} \right) - T_h C_h^{(1)} - T_c C_c^{(3)}.$$
(14.126)

If the work $W(\mathcal{P})$ is positive, then the heat from the hot bath is positive and the heat from the cold bath is negative. Hence the efficiency fulfills

$$\eta(\mathcal{P}) = 1 + \frac{Q_c}{Q_h} = 1 - \frac{|Q_c|}{Q_h}$$

$$\le 1 - \frac{T_c}{T_h} \left(\frac{\Delta S^{(2,4)} + C_c^{(3)}}{\Delta S^{(2,4)} - C_h^{(1)}} \right),$$
(14.127)

with $\Delta S = S_h^{(2)} - S_c^{(4)}$. This proves the theorem for Carnot-like protocols.

For non-Carnot like protocols, let us first discuss an example to understand what happens. Consider the protocol in Figure 14.5 b). From the picture we can see that we can understand this protocol as two Carnot-like protocols, connected by an isothermal at the cold bath: First the lower Carnot-like cycle is done (with work $W^{(1)}$ and heat $Q_h^{(1)}$), then an isothermal from $H^{(8)}$ to $H^{(7)}$ is implemented, followed by a Carnot-like cycle (with work $W^{(2)}$ and heat $Q_h^{(2)}$) and finally an isothermal back from $H^{(7)}$ to $H^{(8)}$. The total work done by the isothermal process connecting the two cycles is at most

zero (it is a cyclic process with a single heat bath and initial state in thermal equilibrium) and no heat is exchanged with the hot bath during these processes. Therefore, the total work and heat fulfill

$$W(\mathcal{P}) \leq W^{(1)} + W^{(2)}, \quad Q(\mathcal{P}) = Q_h^{(1)} + Q_h^{(2)}. \tag{14.128}$$

This yields an efficiency bounded as

$$\eta(\mathcal{P}) \leq \frac{W^{(1)} + W^{(2)}}{Q_h^{(1)} + Q_h^{(2)}} \leq \max\left\{\eta^{(1)}, \eta^{(2)}\right\}, \tag{14.129}$$

with $\eta^{(i)} = W^{(i)}/Q_h^{(i)}$.

In the same way, since we demand that any protocol be cyclic and isothermals are reversible, it follows that the efficiency of any protocol can be upper bounded by one which can be subdivided into m Carnot-like cycles connected by isothermal processes at the cold bath – just like in the above example for $m = 2$. From the point of efficiencies, these then act like m Carnot-like cycles running in parallel and the total efficiency is bounded by the maximum efficiency of the m different Carnot-like cycles.

We thus find that the efficiency of any protocol can be upper bounded

$$\eta(\mathcal{P}) \leq \max_i \left\{\eta^{(i)}\right\}. \tag{14.130}$$

The optimal cycle $\eta^{(i)}$ then determines the Hamiltonians $H^{(j)}$ appearing in the theorem. This finishes the proof.

14.6.7 Ratios of entropies in the strong coupling limit

In this section, we calculate how the ratio of entropies at different temperatures

$$\frac{S(\omega_{\beta_c}(H))}{S(\omega_{\beta_h}(H))} \tag{14.131}$$

behaves in the limit of strong interactions for a ferromagnetic Hamiltonian. To keep the discussion as general as possible, we will only use very general properties of the Ising model, namely the fact that it has a gap and a two ground states that are couple to an external magnetic field in a macroscopic way (this will become clearer below).

Let us consider a Hamiltonian of the form

$$JH(J) = J\left(H_0 + \frac{B}{J}V\right) = JH_0 + BV \tag{14.132}$$

on a system of N spins. In the following N is fixed. Note that as $J \to \infty$, the Hamiltonian $H(J)$ approaches H_0. We make the following assumptions, which are all fulfilled for the classical Ising model if $V = \sum_j \sigma_j^z$ denotes the external field.

1. H_0 is a local Hamiltonian with spectral gap Δ.

2. H_0 has a two-fold degenerate ground state.

3. V is a local Hamiltonian.

4. For large values of J, the effect of the perturbation $\frac{B}{J}V$ on H_0 is to split the ground state degeneracy by an amount $cN\frac{B}{J}$, for some constant $c > 0$. The spectral gap above the two ground states remains of order unity as $J \to \infty$.

We can then prove the following Lemma.

Lemma 14.16 (Ratios of entropies). *Let $H(J)$ fulfill the above assumptions. Define*

$$S_{c/h}(J) := S(\omega_{\beta_{c/h}}(JH(J))) = S(\omega_{\beta_{c/h}}(JH_0 + BV)). \tag{14.133}$$

Then

$$\lim_{J\to\infty} \frac{S_c(J)}{S_h(J)} \begin{cases} \leq 2\frac{\beta_c}{\beta_h}e^{-(\beta_c - \beta_h)2cBN} + \frac{\log(1+e^{-\beta_c 2cBN})}{\log(1+e^{-\beta_h 2cBN})}, & B \neq 0, \\ = 1, & B = 0. \end{cases} \tag{14.134}$$

Proof. Let $E_i(J)$ denote the eigen-energies of $JH(J)$ in non-decreasing order, i.e., $E_{i+1}(J) \geq E_i(J)$. Since we are interested in the entropy, we can set the ground state energy of $JH(J)$ to $E_0(J) = 0$ for every J separately. By the assumptions above, we then have $E_1(J) = 2cNB$ and $E_2(J) = J(\Delta + O(1))$. Let us also write $Z_{c/h}(J) = \text{Tr}(e^{-\beta_{c/h}JH(J)})$ for the partition function. Using the above conventions, it fulfills $Z_{c/h}(J) \geq 1$.

We will use the formula

$$S_{c/h}(J) = \beta_{c/h} E_{\beta_{c/h}}(J) + \log(Z_{c/h}(J))$$

with $E_{\beta_{c/h}}(J)$ the thermal energies of $JH(J)$ at inverse temperature $\beta_{c/h}$. Since the gap of $JH(J)$ increases linearly with J, only the levels $E_0(J)$ and $E_1(J)$ contribute in the limit $J \to \infty$. We hence obtain

$$\lim_{J\to\infty} S_{c/h}(J) = \beta_{c/h} \frac{2cBN}{1+e^{-\beta_{c/h}2cBN}} + \log(1+e^{-\beta_{c/h}2cBN}).$$

We can then bound the ratio of entropies as

$$\lim_{J\to\infty} \frac{S_c(J)}{S_h(J)} = \frac{\beta_c \frac{2cBN}{1+e^{-\beta_c 2cBN}} + \log(1+e^{-\beta_c 2cBN})}{\beta_h \frac{2cBN}{1+e^{-\beta_h 2cBN}} + \log(1+e^{-\beta_h 2cBN})},$$

$$\leq 2\frac{\beta_c}{\beta_h}e^{-(\beta_c-\beta_h)2cBN} + \frac{\log(1+e^{-\beta_c 2cBN})}{\log(1+e^{-\beta_h 2cBN})}. \qquad (14.135)$$

where he have used in the last line that all the terms appearing are positive. For $B = 0$ we get

$$\lim_{J\to\infty} \frac{S_c(J)}{S_h(J)} = 1. \qquad (14.136)$$

$\square$

14.7 Thermodynamics in closed quantum systems using GGEs

14.7.1 Example of a non-passive state with passive correlation matrix

Consider a three-fermion system with Hamiltonian

$$H = \epsilon_1 \eta_1^\dagger \eta_1 + \epsilon_2 \eta_2^\dagger \eta_2 + \epsilon_3 \eta_3^\dagger \eta_3 \qquad (14.137)$$

and a state ρ diagonal in the basis of H with $\mathrm{Tr}(\eta_i^\dagger \eta_i \rho) = p_i$ with $i = 1, 2, 3$. The correlation matrix of the state ρ is simply $C(\rho) = \mathrm{diag}\{p_1, p_2, p_3\}$ and is hence passive whenever $p_1 \geq p_2 \geq p_3$. Let us now choose $\epsilon_1 = 1$, $\epsilon_2 = 2$, $\epsilon_3 = 2.5$ and $p_1 = 0.4$, $p_2 = 0.3$, and $p_3 = 0.1$. Then the energy level with energy $\epsilon_3 = 2.5$ has probability $(1 - p_1)(1 - p_2)p_3 = 0.042$. On the other hand, the level with energy $\epsilon_1 + \epsilon_2 = 3$ has probability $p_1 p_2 (1 - p_3) = 0.108$. Hence, the state is not passive.

14.8 Corrections to work and efficiency under strong coupling

14.8.1 Derivation of the correction terms to the work

In this section, I provide the explicit calculation leading to the correction terms $\Delta F^{(\text{irrev})}$ and $\Delta F^{(\text{res})}$. As described in the main-text, the work that is extracted in the given protocol is given by

$$W(\rho(0)_S, H(0)_S) = W_1 + W_2 + W_3, \tag{14.138}$$
$$\tag{14.139}$$

with

$$W_1 = \text{Tr}\left(\rho(0) \otimes \omega_\beta(H_B)\left(H(0)_S - H_S^{(1)} - V\right)\right), \tag{14.140}$$

$$W_2 = F_\beta(\omega_\beta(H^{(1)}), H^{(1)}) - F_\beta(\omega_\beta(H^{(f)}), H^{(f)}), \tag{14.141}$$

$$W_3 = \text{Tr}\left(\omega_\beta(H^{(f)})\left(H_S^{(f)} + V - H(0)_S\right)\right). \tag{14.142}$$

Here, the Hamiltonians $H^{(1)/(f)} = H_S^{(1)/(f)} + V + H_B$ still need to be optimized to maximize the extracted work. We want to arrive at an expression of the form

$$W(\rho(0)_S, H(0)_S) = W^{(\text{weak})}(\rho(0), H(0)_S) - \Delta F^{(\text{irrev})}(H_S^{(1)}) - \Delta F^{(\text{res})}(H_S^{(f)}), \tag{14.143}$$

with $W^{(\text{weak})}(\rho(0), H(0)_S) = \Delta F_\beta(\rho(0), H(0)_S)$. To arrive at this expression, first note that

$$F_\beta\left(\omega_\beta(H^{(f)}), H^{(f)}\right) = F_\beta\left(\omega_\beta(H^{(f)}), H_S^{(f)} + H_B\right) + \text{Tr}\left(\omega_\beta(H^{(f)})V\right)$$
$$= F_\beta\left(\omega_\beta(H^{(f)}), H(0)_S + H_B\right) + \text{Tr}\left(\omega_\beta(H^{(f)})(V + H_S^{(f)} - H(0)_S)\right).$$

Therefore

$$W_2 + W_3 = F_\beta\left(\omega_\beta(H^{(1)}), H^{(1)}\right) - F_\beta\left(\omega_\beta(H^{(f)}), H(0)_S + H_B\right) \tag{14.144}$$
$$= F_\beta\left(\omega_\beta(H^{(1)}), H^{(1)}\right) - F_\beta\left(\omega_\beta(H(0)_S + H_B), H(0)_S + H_B\right)$$
$$- \Delta F_\beta\left(\omega_\beta(H^{(f)}), H(0)_S + H_B\right) \tag{14.145}$$

We need two further relations. First we have

$$F_\beta\left(\omega_\beta(H(0)_S + H_B), H(0)_S + H_B\right) = F_\beta\left(\omega_\beta(H(0)_S) \otimes \omega_\beta(H_B), H^{(1)}\right)$$
$$+ \text{Tr}\left(\omega_\beta(H(0)_S) \otimes \omega_\beta(H_B)(H(0)_S - H_S^{(1)} - V)\right).$$

Second,

$$F_\beta\left(\omega_\beta(H(0)_S) \otimes \omega_\beta(H_B), H^{(1)}\right) = F_\beta\left(\rho(0) \otimes \omega_\beta(H_B), H^{(1)}\right) + \text{Tr}\left(H^{(1)}(\omega_\beta(H(0)_S) - \rho(0))\right)$$
$$+ \frac{1}{\beta}\left(S(\rho(0)) - S(\omega_\beta(H(0)_S))\right)$$
$$= F_\beta\left(\rho(0) \otimes \omega_\beta(H_B), H^{(1)}\right) - \Delta F_\beta\left(\rho(0), H(0)_S\right)$$
$$+ \text{Tr}\left((H_S^{(1)} + V - H(0)_S)(\omega_\beta(H(0)_S) - \rho(0)) \otimes \omega_\beta(H_B)\right),$$

where we have used $\text{Tr}(A(\rho - \sigma) \otimes \chi) = \text{Tr}(A(\rho - \sigma))$ for any states ρ, σ, χ and any operator A. Combining these two relations, we obtain

$$F_\beta\left(\omega_\beta(H(0)_S + H_B), H(0)_S + H_B\right) = F_\beta\left(\rho(0) \otimes \omega_\beta(H_B), H^{(1)}\right) - \Delta F_\beta(\rho(0), H(0)_S)$$
$$+ \text{Tr}\left(\rho(0) \otimes \omega_\beta(H_B)(H(0)_S - H_S^{(1)} - V)\right)$$
$$= F_\beta\left(\rho(0) \otimes \omega_\beta(H_B), H^{(1)}\right) - \Delta F_\beta(\rho(0), H(0)_S) + W_1.$$

Inserting this last result into (14.145), we finally obtain the desired result

$$W_1 + W_2 + W_3 = \Delta F_\beta\left(\rho(0), H(0)_S\right)$$
$$- \Delta F_\beta\left(\rho(0) \otimes \omega_\beta(H_B), H^{(1)}\right) - \Delta F_\beta\left(\omega_\beta(H^{(f)}), H(0)_S + H_B\right). \tag{14.146}$$

14.8.2   Proof of Lemma 11.3: Minimizing residual free energy

In this section we prove Lemma 11.3. The strategy is the same as for the proof of Lemma 11.2: we simply take the derivative of $\Delta F^{(\mathrm{res})}$ and set it to zero. To that end define the operators

$$R_S(t) = R_S + tY_S, \quad R(t) = R_S(t) + V + H_B, \tag{14.147}$$

where R_S is one choice of local Hamiltonian that gives a (local) minimum of the residual free energy. The derivative of $\Delta F^{(\mathrm{res})}(R_S(t))$ with respect to t is given by

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \Delta F^{(\mathrm{res})} = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \frac{1}{\beta} D\left(\omega_\beta(R(t))\|\omega_\beta(H(0)_S + H_B)\right), \tag{14.148}$$

which we require to vanish, since R_S is a local minimum. To compute the derivative of the relative entropy, we use a similar strategy as in the proof of Lemma 11.4. First, the derivative of the entropy of $\omega_\beta(R(t))$ is given by

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} S(\omega_\beta(R(t)) = -\mathrm{Tr}\left(\log(\omega_\beta(R(0)))\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0}\omega_\beta(R(t))\right). \tag{14.149}$$

Then the derivative of the relative entropy can be expressed as

$$
\begin{aligned}
&\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} D\left(\omega_\beta(R(t))\|\omega_\beta(H(0)_S + H_B)\right)\\
&= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0}\left(-S(\omega_\beta(R(t))) - \mathrm{Tr}\left(\omega_\beta(R(t))\log(\omega_\beta(H(0)_S + H_B))\right)\right)\\
&= -\mathrm{Tr}\left(\left(\log(\omega_\beta(H(0)_S + H_B)) - \log(\omega_\beta(R(0)))\right)\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0}\omega_\beta(R(t))\right)\\
&= \beta\mathrm{Tr}\left((H(0)_S + H_B - R(0))\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0}\omega_\beta(R(t))\right),
\end{aligned}
\tag{14.150}
$$

where we have used that the trace of the derivative of $\omega_\beta(R(t))$ vanishes. Let us now compute the derivative of the state $\omega_\beta(R(t))$. This can be done using Duhamel's formula (11.31) and yields

$$
\begin{aligned}
\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0}\omega_\beta(R(t)) &= -\beta(Y_S)_{R(0)}\omega_\beta(R(0)) + \beta\omega_\beta(R(0))\mathrm{Tr}\left(Y_S\omega_\beta(R(0))\right)\\
&= \beta\left(\mathrm{Tr}\left(Y_S\omega_\beta(R(0))\right) - (Y_S)_{R(0)}\right)\omega_\beta(R(0)),
\end{aligned}
\tag{14.151}
$$

where we have again employed the notation

$$(Y_S)_{R(0)} = \int_0^1 \mathrm{e}^{-\beta\tau R(0)}Y_S\,\mathrm{e}^{\beta\tau R(0)}\mathrm{d}\tau. \tag{14.152}$$

Using this result and the cyclic nature of the trace, the derivative of the relative entropy then takes the form

$$
\begin{aligned}
&\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} D\left(\omega_\beta(R(t))\|\omega_\beta(H(0)_S + H_B)\right)\\
&= \beta^2\mathrm{Tr}\left((H(0)_S + H_B - R(0))\,\omega_\beta(R(0))\right)\mathrm{Tr}\left(Y_S\omega_\beta(R(0))\right)\\
&\quad - \beta^2\mathrm{Tr}\left(\omega_\beta(R(0))\,(H(0)_S + H_B - R(0))_{-R(0)}\,Y_S\right).
\end{aligned}
\tag{14.153}
$$

This expression has to vanish for all operators Y_S. Identifying $R(0)$ with the operator R in the statement of the Lemma, we can then reformulate the condition that the derivative vanishes as

$$\mathrm{Tr}_B(\omega_\beta(R))\mathrm{Tr}\left(\omega_\beta(R)\,(R_S + V - H(0)_S)\right) = \mathrm{Tr}_B\left(\omega_\beta(R)\,(R_S + V - H(0)_S)_{-R}\right),$$

which is the desired result.

14.9 On spontaneous symmetry breaking in dissipative systems

14.9.1 General proof for the case of a discrete symmetry

In the main-text of chapter 12, I only provided the proof of theorem 12.4 for the case of exactly local Liouvillians. In this section, I will generalize the result to the case of Liouvillians that are only approximately local. This is more difficult from a purely technical point of view, but the basic steps of the proof will be exactly the same.

First we need to formalize, what we mean by an approximately local Liouvillian. Intuitively, this means that each term $\mathcal{L}_x^\Lambda$ can be well approximated by a term $\tilde{\mathcal{L}}_x^\Lambda$ that is strictly local and supported on a ball $B_l(x)$ of radius l and center x. The error of approximation should decrease with increasing l according to some function f, which determines *how* local the Liouvillian $\mathcal{L}^\Lambda$ is.

Definition 14.17 (f-local Liouvillian). *Let* $f : \mathbb{R} \to \mathbb{R}$ *be a given positive semi-definite function with* $f(0) = 1$. *A sequence of Liouvillians* $\mathcal{L}^\Lambda = \sum_x \mathcal{L}_x^\Lambda$ *is called f-local if there exists a sequence of Liouvillians* $\tilde{\mathcal{L}}^\Lambda = \sum_x \tilde{\mathcal{L}}_x^\Lambda$ *such that*

1. *Each term* $\tilde{\mathcal{L}}_x^\Lambda$ *is supported within* $B_l(x)$.

2. $\tilde{\mathcal{L}}^\Lambda$ *approximates* $\mathcal{L}^\Lambda$*: For any local operator A, we have*

$$\left\| \tilde{\mathcal{L}}_x^\Lambda[A] - \mathcal{L}_x^\Lambda[A] \right\| \le bf(l)\|A\|, \tag{14.154}$$

for some constant $b > 0$.

In this section, we can choose f to be any function that decays at least like a short-ranged power-law.

Definition 14.18 (Approximately local Liouvillian). *We call a sequence of f-local Liouvillians* $\mathcal{L}^\Lambda$ approximately local *if f decays at least as fast as*

$$g_\beta(l) := \frac{1}{1 + l^\beta}, \tag{14.155}$$

for some $\beta > 2D$.

Similarly, I will call an operator A approximately local, if it can be approximated by a sequence of operators $\tilde{A}_l$, each supported in a ball of radius l and such that $\|A - \tilde{A}_l\| \le \|A\| g_\beta(l)$. I will call it *quasi-local* if the same holds for a function that decays faster than any power instead of a power-law.

In chapter 12, I formally only considered the magnetization as an order parameter. Here, the order parameter O can be any extensive quantity

$$O_\Lambda = \sum_{x \in \Lambda} O_{\{x\}}, \tag{14.156}$$

where $O_{\{x\}}$ is supported within the ball $B_{l_O}(x)$ for some constant l_O and is uniformly bounded, i.e., there exists a constant o such that $\left\| O_{\{x\}} \right\| \le o$ for all $x \in \mathbb{Z}^D$. As before, given any sequence of states ω_Λ which fulfill $\omega_\Lambda(O_\Lambda) = 0$, we can formally define the sequence of states

$$\omega_\Lambda^\pm(A) := \omega_\Lambda^\pm \left(\tilde{O}_\Lambda^\pm A \tilde{O}_\Lambda^\pm \right), \quad \tilde{O}_\Lambda^\pm := \frac{1}{2} \left(\mathbf{1} \pm \frac{O_\Lambda}{\sqrt{\omega_\Lambda(O_\Lambda^2)}} \right). \tag{14.157}$$

These states are well defined as long as $\omega_\Lambda(O_\Lambda^2) > 0$, which is true as long as the associated density-matrix ρ^ω is not completely supported in the kernel of O_Λ. We can now state the precise theorem.

Theorem 14.19. *Let* $\mathcal{L}^\Lambda$ *be a sequence of purely dissipative, approximately local Liouvillians which are in detailed balance with a sequence of states* ω_Λ *that fulfill* $\omega_\Lambda(O_\Lambda) = 0$ *and* $\omega_\Lambda(O_\Lambda^2) \ge \mu^2 o^2 |\Lambda|^2$. *Then the states* $\omega_\Lambda^\pm$ *are asymptotically-reversible.*

By the same calculation as in chapter 12, it follows that the states $\omega_\Lambda^\pm$ are symmetry-breaking if there is a sequence of unitaries U_Λ such

$$U_\Lambda O_\Lambda U_\Lambda^{-1} = -O_\Lambda, \tag{14.158}$$

ω_Λ is symmetric with respect to U_Λ, and $\omega_\Lambda(O_\Lambda^2) \ge \mu^2 o^2 |\Lambda|^2$ for some constant $\mu > 0$.

Let us now prove the theorem. To simplify the notation, we will again drop the Λ from all the operators and states and set $N := |\Lambda|$. Since the argument again is based on an approximate Leibniz-property, it will turn out to be useful to define the function

$$\Gamma_\mathcal{L}(A, B) := \mathcal{L}[AB] - \mathcal{L}[A]B - A\mathcal{L}[B]. \tag{14.159}$$

It measures how far the Liouvillian $\mathcal{L}$ deviates from fulfilling the Leibniz rule on the operator AB. Similarly, let us introduce a measure for how far the states $\omega^\pm$ deviate from being reversible:

$$\Delta^\pm(A, B) := \omega^\pm(A\mathcal{L}[B]) - \omega^\pm(\mathcal{L}[A]B). \tag{14.160}$$

We need to show that $\Delta^\pm(A, B) \to 0$ as for $N \to \infty$ for any two local operators A and B. To achieve this, we again first use that

$$\lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\omega\left(O(A\mathcal{L}[B] - \mathcal{L}[A]B)O\right)}{N^2} = \lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\omega\left(O^2(A\mathcal{L}[B] - \mathcal{L}[A]B)\right)}{N^2},$$

which can be seen by writing $OXO = O[X, O] + O^2 X$ and using that $O[O, X]$ scales like N for any approximately local operator X. We then need to prove

$$\lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\omega\left(O^2(A\mathcal{L}[B] - \mathcal{L}[A]B)\right)}{N^2} = 0, \tag{14.161}$$

$$\lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\omega\left(O(A\mathcal{L}[B] - \mathcal{L}[A]B)\right)}{N} = 0, \tag{14.162}$$

$$\lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\omega\left((A\mathcal{L}[B] - \mathcal{L}[A]B)O\right)}{N} = 0. \tag{14.163}$$

To do this, we will use the following Lemma, whose proof is given further below.

Lemma 14.20 (Approximate Leibniz rule). *Let $\mathcal{L}$ be an approximately local Liouvillian, let A, B be local operators and O an extensive quantity (order parameter). Then*

$$\lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\|\Gamma_\mathcal{L}(O^2, A)B\|}{N^2} = 0, \quad \lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\|\Gamma_\mathcal{L}(O, A)B\|}{N} = 0, \quad \lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\|A\Gamma_\mathcal{L}(B, O)\|}{N} = 0. \tag{14.164}$$

Given the statement of the Lemma, we can use reversibility to write

$$\frac{\omega\left(O^2(A\mathcal{L}[B] - \mathcal{L}[A]B)\right)}{N^2} = \frac{\omega\left((\mathcal{L}[O^2 A] - O^2\mathcal{L}[A])B\right)}{N^2} = \frac{\omega\left(\Gamma_\mathcal{L}(O^2, A)B\right)}{N^2} + \frac{\omega\left(\mathcal{L}[O^2]AB\right)}{N^2}.$$

The first term on the right-hand side vanishes in the thermodynamic limit. For the second term, we observe that

$$\omega\left(\mathcal{L}[O^2]AB\right) = -\omega\left(\Gamma_\mathcal{L}(O^2, AB)\right) + \omega\left(\mathcal{L}[O^2 AB]\right) - \omega\left(O^2\mathcal{L}[AB]\right) \tag{14.165}$$

$$= -\omega\left(\Gamma_\mathcal{L}(O^2, AB)\right) - \omega\left(O^2\mathcal{L}[AB]\right) \tag{14.166}$$

$$= -\omega\left(\Gamma_\mathcal{L}(O^2, AB)\right) - \omega\left(\mathcal{L}[O^2]AB\right). \tag{14.167}$$

But AB is also a local operator. Then we can use the Lemma again with $A' = AB$ and $B' = \mathbf{1}$ and conclude that

$$\lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\omega\left(O^2(A\mathcal{L}[B] - \mathcal{L}[A]B)\right)}{N^2} = \lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\omega\left(\mathcal{L}[O^2]AB\right)}{N^2} = 0. \tag{14.168}$$

A completely analogous argument can be used to deduce that

$$\lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\omega\left(O(A\mathcal{L}[B] - \mathcal{L}[A]B)\right)}{N} = \lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\omega\left((A\mathcal{L}[B] - \mathcal{L}[A]B)O\right)}{N} = 0.$$

Proof of Lemma 14.20. To prove the Lemma, we introduce an approximate Liouvillian $\tilde{\mathcal{L}}$ whose terms $\tilde{\mathcal{L}}_x$ are supported within balls $B_{L^\alpha}(x)$, where the constant $0 < \alpha < 1$ is chosen later. For any operator X, each such term introduces an error of

$$\left\|\mathcal{L}_x[X] - \tilde{\mathcal{L}}_x[X]\right\| \le \|X\| \, b \frac{1}{1 + L^{\alpha\beta}}. \tag{14.169}$$

Let us write $\mathcal{R} := \mathcal{L} - \tilde{\mathcal{L}}$. Note that $\mathcal{R}$ has the property $\mathcal{R}[\mathbf{1}] = 0$. For any operator A let us further write $\tilde{\mathcal{L}}_{\tilde{A}}$ for the Liouvillian that contains all the terms that have overlap with the support of A. The support $\tilde{A}$ of $\tilde{\mathcal{L}}_{\tilde{A}}$ then can be bounded by $k|A|L^{\alpha D}$ for some constant k in the limit of large L. Using this notation, we can observe that

$$\Gamma_{\tilde{\mathcal{L}}}(X, A) = \Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}(X, A), \tag{14.170}$$

since $\tilde{\mathcal{L}} - \tilde{\mathcal{L}}_{\tilde{A}}[XA] = \tilde{\mathcal{L}} - \tilde{\mathcal{L}}_{\tilde{A}}[X]\,A$. This implies, since Γ is linear in the Liouvillian, that we can write $\Gamma_{\mathcal{L}}(X, A) = \Gamma_{\mathcal{R}}(X, A) + \Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}(X, A)$. In particular, we find that to prove the Lemma, it is sufficient to show

$$\lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\left\|\Gamma_{\mathcal{R}}(O^2, A)\right\|}{N^2} = \lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\left\|\Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}(O^2, A)\right\|}{N^2} = 0, \tag{14.171}$$

$$\lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\left\|\Gamma_{\mathcal{R}}(O, A)\right\|}{N} = \lim_{\Lambda \nearrow \mathbb{Z}^D} \frac{\left\|\Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}(O, A)\right\|}{N} = 0. \tag{14.172}$$

I will only prove the first line, since the latter follows from completely analogous reasoning. For the first term, we can use the definition of Γ together with the triangle inequality and sub-multiplicativity of the norm to obtain

$$
\begin{aligned}
\frac{\left\|\Gamma_{\mathcal{R}}(O^2, A)\right\|}{N^2} &= \frac{\left\|\mathcal{R}\left[O^2 A\right] - \mathcal{R}\left[O^2\right]A - O^2 \mathcal{R}\left[O^2\right]\right\|}{L^{2D}} \\
&\le \frac{3|\Lambda|\,\|O^2\|\,\|A\|}{L^{2D}} b g_\beta(L^\alpha) \le \frac{3|\Lambda|^3 o^2\,\|A\|}{L^{2D}} b g_\beta(L^\alpha) \tag{14.173} \\
&= \frac{3L^{3D} o^2\,\|A\|}{L^{2D}} b g_\beta(L^\alpha) = 3 o^2\,\|A\|\,b\frac{L^D}{1 + L^{\alpha\beta}}. \tag{14.174}
\end{aligned}
$$

This term therefore vanishes in the thermodynamic limit provided that $\alpha\beta > D$. For the second term we decompose O into $O = Q + R$, such that R is supported on $\tilde{A}$, while Q is supported on the complement of $\tilde{A}$. In particular, this implies $\Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}(Q, X) = Q\Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}(\mathbf{1}, X) = 0$, since $\Gamma_{\mathcal{L}}(\mathbf{1}, X)$ for any Liouvillian $\mathcal{L}$ and operator X. We can use this property to obtain

$$\Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}(O^2, A) = 2Q\Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}(R, A) + \Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}(R^2, A). \tag{14.175}$$

Using the triangle inequality and the estimate $|\tilde{A}| \le k|A|L^{\alpha D}$, we then obtain

$$
\begin{aligned}
\frac{\left\|\Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}(O^2, A)\right\|}{L^{2D}} &\le \frac{K' o^2 |\Lambda|\,\|A\|\,|\tilde{A}|^2}{L^{2D}} \le \frac{K o^2\,\|A\|\,|A| L^D L^{2\alpha D}}{L^{2D}} \\
&\le K o^2\,\|A\|\,|A| L^{(2\alpha - 1)D}, \tag{14.176}
\end{aligned}
$$

where K' and K are constants independent of the system-size. We thus find that the term vanishes in the thermodynamic limit if $\alpha < 1/2$. Combining this bound with the previous bound $\alpha\beta > D$, we arrive at the final bound $\beta > 2D$. The corresponding estimates for the term $\left\|\Gamma_{\tilde{\mathcal{L}}}(O, A)\right\|/N$ yields the same restriction. This finishes the proof of the Lemma. □

14.9.2 General proof for the case of a continuous symmetry

In this section, I provide the general proof for theorem 12.6. Similarly as in the previous section, I will show it for Liouvillians that are not strictly local. Slightly stronger locality properties are needed in the case of continuous symmetries, though.

Definition 14.21 (Short-ranged Liouvillian). *A sequence of f-local Liouvillians $\mathcal{L}^\Lambda$ is called* short-ranged *if f decays at least as fast as*

$$h_\beta(l) = e^{-l^\beta/\xi}, \tag{14.177}$$

for some $\beta > 0$ and $\xi > 0$.

In the following, we will assume that $\mathcal{L}^\Lambda$ is short-ranged.

As discussed in the main-text of chapter 12, the symmetry-breaking states $\omega_\Lambda^{(M)}$ are linear combinations of the functionals $\chi_\Lambda^{m,m'}$ defined as

$$\chi_\Lambda^{(m,m')}(A) := \frac{\omega_\Lambda\left((O_\Lambda^-)^{m'} A (O_\Lambda^+)^m\right)}{Z(m)Z(m')}, \tag{14.178}$$

which are defined for states that fulfill the conditions

$$[\rho_\Lambda^\omega, C_\Lambda] = 0, \quad \omega_\Lambda\left((O_\Lambda^{(i)})^2\right) \ge \mu^2 o^2 |\Lambda|^2, \quad i = 1, 2. \tag{14.179}$$

Let us fix some arbitrary $M > 0$ and assume $|m|, |m'| \le M$ in what follows. As before, we will also drop any explicit Λ-labels on operators and states and set $|\Lambda| = N$. Let us define the quantity

$$\Delta^{(m,m')}(A, B) := \chi^{(m,m')}(B\mathcal{L}[A]) - \chi^{(m,m')}(\mathcal{L}[B]\,A). \tag{14.180}$$

It measures how reversible the functional $\chi^{(m,m')}$ is with respect to the Liouvillian $\mathcal{L}$.

Theorem 14.22 (Reversibility from fluctuations for continuous symmetries)**.** *Suppose $\mathcal{L}$ is a short-ranged Liouvillian that satisfies detailed balance with respect to ω. Suppose that ω fulfills the conditions* (14.179)*. Then*

$$\lim_{\Lambda \nearrow \mathbb{Z}^D} \Delta^{(m,m')}(A,B) = 0. \tag{14.181}$$

Consequently, the symmetry-breaking states $\omega^{(M)}$ are asymptotically reversible.

I will split the proof of this theorem into several Lemmas. The basic proof idea will, however, be the same as in the case of discrete symmetry-breaking, namely the use of an approximate Leibniz-rule. The first key Lemma in the proof of the theorem was proven by Koma and Tasaki. The result of this Lemma is used in the second Lemma, which allows us to make use of detailed balance.

Lemma 14.23 (Koma and Tasaki [287])**.** *Let the conditions in* (14.179) *be fulfilled for ω. Now let A be some finite region and decompose O^+ as $O^+ = Q_A + R_A$, where Q_A is supported on the complement of A and R_A is supported on A. Then we have the inequalities*

$$\frac{\mathrm{Tr}(Q_A^{m-k}\rho^\omega(Q_A^\dagger)^{m-k})}{\mathrm{Tr}(Q_A^m\rho^\omega(Q_A^\dagger)^m)} \le (\mu o N)^{-2k} \tag{14.182}$$

and

$$r_A^{(M)} = \left| \frac{\mathrm{Tr}\left((O^+)^M\rho^\omega(O^-)^M\right)}{\mathrm{Tr}(Q_A^M\rho^\omega(Q_A^\dagger)M)} \right| \ge 2 - \exp\left(\frac{2|A|M}{\mu N}\right) \ge 2 - e^{\mu/8}. \tag{14.183}$$

for $N \ge \frac{16|A|^2}{\mu^2}$ and $|\frac{M}{N}| \le \frac{\mu^2}{16|A|}$.

Proof. A proof can be found at the end of this section. □

Lemma 14.24 (Local observables)**.** *Let A be any local observable, ω be represented by the density-matrix ρ^ω and fulfill* (14.179)*. Then*

$$\left| \mathrm{Tr}\left(\chi^{(m,m')}A\right) \right| \le O\left(\frac{M|A|\,\|A\|}{N}\right) + \left| \frac{\mathrm{Tr}\left(\rho^\omega(O^-)^{m'}(O^+)^m A\right)}{\mathrm{Tr}\left((O^+)^m\rho^\omega(O^-)^m\right)^{1/2}\mathrm{Tr}\left((O^+)^{m'}\rho^\omega(O^-)^{m'}\right)^{1/2}} \right|.$$

Proof. Let us first again rewrite a product of operators in terms of the commutator and the reversed product:

$$\mathrm{Tr}\left(\rho^\omega(O^-)^{m'}A(O^+)^m\right) = \mathrm{Tr}\left(\rho^\omega(O^-)^{m'}[A,(O^+)^m]\right) + \mathrm{Tr}\left(\rho^\omega(O^-)^{m'}(O^+)^m A\right). \tag{14.184}$$

We have to show that the first term is small enough so that it is of order N^{-1} when divided by the denumerator $Z(m)Z(m')$ in the definition of $\chi^{(m,m')}$. We will do this using the preceeding Lemma and therefore split up O^+ as $O^+ = Q_A + R_A$. Note that Q_A and R_A are not self-adjoint. We can then use a binomial expansion and obtain

$$\mathrm{Tr}\left(\rho^\omega(O^-)^{m'}[A,(O^+)^m]\right) = \sum_{k=0}^{m'}\sum_{l=0}^{m}\binom{m'}{k}\binom{m}{l}\mathrm{Tr}\left(\rho^\omega(Q_A^\dagger)^{m'-k}(R_A^\dagger)^k[A,Q_A^{m-l}R_A^l]\right)$$

$$= \sum_{k=0}^{m'}\sum_{l=1}^{m}\binom{m'}{k}\binom{m}{l}\mathrm{Tr}\left(\rho^\omega(Q_A^\dagger)^{m'-k}(R_A^\dagger)^k[A,R_A^l]Q_A^{m-l}\right),$$

where we used that $[Q_A,A] = 0$ and $[Q_A,R_A] = 0$. We can now use the Schwartz inequality in the form

$$|\mathrm{Tr}(\rho^\omega A^\dagger BC)| \le \left[\mathrm{Tr}(\rho^\omega A^\dagger A)\mathrm{Tr}(\rho^\omega C^\dagger B^\dagger BC)\right]^{1/2}$$

$$\le \|B\|\left[\mathrm{Tr}(\rho^\omega A^\dagger A)\mathrm{Tr}(\rho^\omega C^\dagger C)\right]^{1/2},$$

and the inequality (14.182). This yields

$$\left| \frac{\mathrm{Tr}\left(\rho^\omega(O^-)^{m'}[A,(O^+)^m]\right)}{\mathrm{Tr}(Q_A^m\rho^\omega(Q_A^\dagger)^m)^{1/2}\mathrm{Tr}(Q_A^{m'}\rho^\omega(Q_A^\dagger)^{m'})^{1/2}} \right| \le 2\,\|A\|\sum_{k=0}^{m'}\sum_{l=1}^{m}\binom{m'}{k}\binom{m}{l}\left(\frac{|A|}{\mu N}\right)^{k+l}$$

$$\le 2\,\|A\|\exp\left(\frac{|A|m'}{\mu N}\right)\left(\exp\left(\frac{|A|m}{\mu N}\right)-1\right)$$

$$\le 2\,\|A\|\exp\left(\frac{|A|M}{\mu N}\right)\left(\exp\left(\frac{|A|M}{\mu N}\right)-1\right)$$

$$\le 2\,\|A\|\frac{16|A|}{\mu^2}e^{\mu/16}(e^{\mu/16}-1)\frac{M}{N},$$

where we assumed $N \geq \frac{16|A|^2}{\mu^2}$ and $|\frac{M}{N}| \leq \frac{\mu^2}{16|A|}$. Multiplying with the ratio (14.183) from Lemma 14.23 yields the desired bound. $\qquad\square$

With this Lemma at our disposal, we can now start with the proof of the theorem. Let us assume, without loss of generality that $m, m' \geq 0$ in the following. For the other cases, we merely have to replace some O^+ with O^- and some operators with their adjoints in the following derivations.

To simplify the notation and reduce the number of appearances of $\lim_{\Lambda \nearrow \mathbb{Z}^D}$, let us also write $\approx$ in the following for any equation that becomes true in the thermodynamic limit.

As in the case of discrete symmetry breaking, let us now introduce a strictly local Liouvillian $\tilde{\mathcal{L}}$ that approximate $\mathcal{L}$. As before, we assume that each term $\tilde{\mathcal{L}}_x$ is supported within a Ball of radius l centered at x. For each term, this introduces an error at most $bh_\beta(l)$: $\|\tilde{\mathcal{L}}_x[A] - \mathcal{L}_x[A]\| \leq bh_\beta(l)$. In the following we will assume that l scales with the linear size of the system L as L^α for some $0 < \alpha < 1/2$. We will collect the error in a Liouvillian $\mathcal{R}$ again: $\mathcal{R} = \mathcal{L} - \tilde{\mathcal{L}}$. With this notation at hand, note that

$$\left| \omega(X\mathcal{L}[Y]Z) - \omega(X\tilde{\mathcal{L}}[Y]Z) \right| \leq |\Lambda| \, \|X\| \, \|Y\| \, \|Z\| \, bh_\beta(l). \qquad (14.185)$$

As the system size-increases, the last error term becomes arbitrarily small provided that $\|Z\| \, \|Y\| \, \|Z\|$ grows at most like a polynomial in the system-size. This will be crucial in what follows. In particular, it shows that if $\mathcal{L}$ is in detailed balance with ω, then $\tilde{\mathcal{L}}$ is approximately in detailed balance with ω:

$$\left| \omega(A\tilde{\mathcal{L}}[B]) - \omega(\tilde{\mathcal{L}}[A]B) \right| \leq 2|\Lambda| \, \|A\| \, \|B\| \, bh_\beta(l). \qquad (14.186)$$

From the Schwarz-inequality (14.185) we have $|\chi^{(m,m')}(A)| \leq \|A\|$. Let us now write $\tilde{\Delta}^{(m,m')}$ for the same quantity as $\Delta^{(m,m')}$, but with $\mathcal{L}$ replaced by $\tilde{\mathcal{L}}$. Then the above relation also implies

$$\left| \Delta^{(m,m')}(A,B) - \tilde{\Delta}^{(m,m')}(A,B) \right| \leq 3|\Lambda| \, \|A\| \, \|B\| \, bh_\beta(l). \qquad (14.187)$$

To prove the theorem, it therefore suffices to show that $\tilde{\Delta}^{(m,m')}$ vanishes in the thermodynamic limit. From Lemma 14.24 we see that it suffices to show that

$$\omega \left(\Omega^{(m,m')} \left(\tilde{\mathcal{L}}[A]B - A\tilde{\mathcal{L}}[B] \right) \right)$$

$$:= \omega \left(\frac{(O^-)^{m'}(O^+)^m}{\omega\left((O^+)^m(O^-)^m\right)^{1/2} \omega\left((O^+)^{m'}(O^-)^{m'}\right)^{1/2}} \left(\tilde{\mathcal{L}}[A]B - A\tilde{\mathcal{L}}[B] \right) \right)$$

$$= \omega \left(\frac{(O^-)^{m'}(O^+)^m}{Z(m)Z(m')} \left(\tilde{\mathcal{L}}[A]B - A\tilde{\mathcal{L}}[B] \right) \right)$$

vanishes in the thermodynamic limit to achieve this, since $\left(\tilde{\mathcal{L}}[A]B - A\tilde{\mathcal{L}}[B] \right)$ is an observable whose support grows slower than linearly with the system size.

We will now use that ω is asymptotically reversible with respect to $\tilde{\mathcal{L}}$. For any local observable A, we again denote by $\tilde{\mathcal{L}}_{\tilde{A}}$ the Liouvillian composed of all those terms $\tilde{\mathcal{L}}_x$ that have overlap with A. The number of these terms is upper bounded by $k|A|l^D = k|A|L^{\alpha D}$ for some constant k and each of them is a bounded Liouvillian. The following Lemma is then the key to the proof.

Lemma 14.25. *Let A be a local observable. If ω fulfills condition (14.179) and $\mathcal{L}$ is a short-ranged Liouvillian in detailed balance with ω, we have*

$$\omega \left(\Gamma_{\tilde{\mathcal{L}}}(\Omega^{(m,m')}, A) \right) = \omega \left(\Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}(\Omega^{(m,m')}, A) \right) \approx 0, \qquad (14.188)$$

if the approximation $\tilde{\mathcal{L}}$ to $\mathcal{L}$ is chosen such that $l \propto L^\alpha$ with $0 < \alpha < 1/2$.

Let us first use the Lemma to see that it implies that $\tilde{\Delta}^{(m,m')}$ vanishes and then give the proof. To see that, let A, B be local operators. Then we first use Lemma 14.24 and approximate detailed balance to write

$$\tilde{\Delta}^{(m,m')}(A,B) \approx \omega \left(\Omega^{(m,m')} \left(A\tilde{\mathcal{L}}[B] - \tilde{\mathcal{L}}[A]B \right) \right) \qquad (14.189)$$

$$\approx \omega \left(\left(\tilde{\mathcal{L}}\left[\Omega^{(m,m')}A \right] - \Omega^{(m,m')}\tilde{\mathcal{L}}[A] \right) B \right). \qquad (14.190)$$

We can now use Lemma 14.25 and apply approximate detailed balance again to obtain

$$\tilde{\Delta}^{(m,m')}(A,B) \approx \omega \left(\tilde{\mathcal{L}}\left[\Omega^{(m,m')} \right] AB \right) \approx \omega \left(\Omega^{(m,m')}\tilde{\mathcal{L}}[AB] \right). \qquad (14.191)$$

On the other hand, AB is also a local observable. We can hence also (instead of doing the last step) use Lemma 14.24 to write

$$\omega\left(\tilde{\mathcal{L}}\left[\Omega^{(m,m')}\right]AB\right) \approx \omega\left(\tilde{\mathcal{L}}\left[\Omega^{(m,m')}AB\right]\right) - \omega\left(\Omega^{(m,m')}\tilde{\mathcal{L}}\left[AB\right]\right) \tag{14.192}$$

$$\approx -\omega\left(\Omega^{(m,m')}\tilde{\mathcal{L}}\left[AB\right]\right). \tag{14.193}$$

Combining these two estimates, we conclude

$$-\omega\left(\Omega^{(m,m')}\tilde{\mathcal{L}}\left[AB\right]\right) \approx \tilde{\Delta}^{(m,m')}(A,B) \approx \omega\left(\Omega^{(m,m')}\tilde{\mathcal{L}}\left[AB\right]\right). \tag{14.194}$$

In other words, we have our desired result

$$\lim_{\Lambda\nearrow\mathbb{Z}^D}\tilde{\Delta}^{(m,m')}(A,B) = \lim_{\Lambda\nearrow\mathbb{Z}^D}\Delta^{(m,m')}(A,B) = 0. \tag{14.195}$$

Proof of Lemma 14.25. The proof of the Lemma proceeds similarly to the proof of Lemma 14.24. Split up $O^+ = Q + R$, where Q is supported on the complement of $\tilde{A}$ and R collects the remaining terms inside of $\tilde{A}$. As by now usual, we then have $\tilde{\mathcal{L}}_{\tilde{A}}[QX] = Q\tilde{\mathcal{L}}_{\tilde{A}}[X]$ for any operator X.

Using a binomial expansion, we then obtain

$$\left|\omega\left(\Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}(\Omega^{(m,m')},A)\right)\right| \leq \sum_{k=0}^{m}\sum_{k'=0}^{m'}\binom{m}{k}\binom{m'}{k'}\left|\frac{\omega\left(\Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}((Q^\dagger)^{m'-k}Q^{m-k}(R^\dagger)^{k'}R^k,A)\right)}{Z(m)Z(m')}\right|$$

$$= \sum_{k=0}^{m}\sum_{k'=0}^{m'}\binom{m}{k}\binom{m'}{k'}\left|\frac{\omega\left((Q^\dagger)^{m'-k}Q^{m-k}\Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}((R^\dagger)^{k'}R^k,A)\right)}{Z(m)Z(m')}\right|.$$

Since $\Gamma_{\tilde{L}}(\mathbf{1},X) = 0$ for any X, we can neglect the term with $k = k' = 0$. Together with the Schwartz inequality we then get

$$\left|\omega\left(\Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}(\Omega^{(m,m')},A)\right)\right| \leq \sum_{k,k'}'\binom{m}{k}\binom{m'}{k'}\frac{\omega\left((Q^\dagger)^{m'-k}Q^{m'-k}\right)^{1/2}\omega\left((Q^\dagger)^{m-k}Q^{m-k}\right)^{1/2}}{Z(m)Z(m')}$$

$$\times\left\|\Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}((R^\dagger)^{k'}R^k,A)\right\|, \tag{14.196}$$

where the primed sum omits the term $k = k' = 0$. We now use again Lemma 14.23 to get

$$\frac{\omega\left((Q^\dagger)^{m'-k}Q^{m'-k}\right)^{1/2}\omega\left((Q^\dagger)^{m-k}Q^{m-k}\right)^{1/2}}{Z(m)Z(m')} \leq \frac{(\mu o L^D)^{-2(k+k')}}{2-e^{\mu/8}}, \tag{14.197}$$

which is true as long as $L^D \geq \frac{16|\tilde{A}|^2}{\mu^2}$ and $\left|\frac{M}{L^D}\right| \leq \frac{\mu^2}{16|\tilde{A}|}$, where $M \geq |m|, |m'|$. Since we can bound $|\tilde{A}| \leq k|A|L^{\alpha D}$, these conditions are fulfilled on large systems provided that $\alpha < 1/2$, as we assumed from the beginning. From the locality of the Liouvillian $\tilde{\mathcal{L}}$ we can also bound the norm-factor as

$$\left\|\Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}((R^\dagger)^{\tilde{k}'}R^{\tilde{k}},A)\right\| \leq k_3|\tilde{A}|\,\|R\|^{\tilde{k}+\tilde{k}'}\,\|A\| \leq k_4\,\|A\|\,|A|L^{\alpha D}(koL^{\alpha D}|A|)^{\tilde{k}+\tilde{k}'}, \tag{14.198}$$

for some constants k_3, k_4. Combining the two results, we then get

$$\left|\omega\left(\Gamma_{\tilde{\mathcal{L}}_{\tilde{A}}}(\Omega^{(m,m')},A)\right)\right| \leq \frac{k_4\,\|A\|\,|A|L^{\alpha D}}{2-e^{\mu/8}}\sum_{\tilde{k},\tilde{k}'}'\binom{m}{\tilde{k}}\binom{m'}{\tilde{k}'}\left(\frac{k|A|}{\mu^2 o}L^{(\alpha-2)D}\right)^{\tilde{k}+\tilde{k}'} \tag{14.199}$$

$$\leq \frac{k_4\,\|A\|\,|A|L^{\alpha D}}{2-e^{\mu/8}}\left(\exp\left(\frac{kM|A|}{\mu^2 o}L^{(\alpha-2)D}\right)-1\right). \tag{14.200}$$

Since we assumed $\alpha < 1/2$, the expression converges to zero as $L \to \infty$ for any fixed M and A. □

Proof of Lemma 14.23. Let $a_m := \text{Tr}(Q_A^m\rho^\omega(Q_A^\dagger)^m)$. In this notation, we have to prove

$$\frac{a_m}{a_{m-1}} \geq (\mu o N)^2. \tag{14.201}$$

Let us first bound a_1:

$$
\begin{aligned}
a_1 &= \mathrm{Tr}((O^+ - R_A)\rho^\omega(O^- - R_A^\dagger)) \\
&\geq \mathrm{Tr}(\rho^\omega O^- O^+) - 2\left\| O^+ R_A^\dagger \right\| \leq 2No^2|A| \\
&= \frac{1}{2}\left[\mathrm{Tr}(\rho^\omega O^+ O^-)\mathrm{Tr}(\rho^\omega O^- O^+) + \mathrm{Tr}(\rho[O^+, O^-]) \right] - 2o^2N|A| \\
&\geq \mathrm{Tr}(\rho^\omega O^{(1)2}) + \mathrm{Tr}(\rho^\omega O^{(2)2}) - 2o^2(1 + |A|)N \\
&\geq 2o^2\mu^2N^2\left[1 - \frac{1 + |A|}{\mu^2 N} \right].
\end{aligned}
\tag{14.202}
$$

We can now use the bound $N \geq \frac{16|A|^2}{\mu^2}$ to obtain

$$
1 - \frac{1 + |A|}{\mu^2 N} \geq 1 - \frac{1 + |A|}{16|A|^2} \geq 1 - \frac{1}{8} > 0,
\tag{14.203}
$$

since $|A| \geq 1$. Therefore $a_1 > 0$. Let us now use the Schwartz inequality to get

$$
\begin{aligned}
(a_{m-1})^2 &\leq \mathrm{Tr}(\rho^\omega (Q_A^\dagger)^{m-2}Q_A^{m-2})\mathrm{Tr}(\rho^\omega(Q_A^\dagger)^{m-1}Q_AQ_A^\dagger Q_A^{m-1}) \\
&= a_{m-2}\left\{ \mathrm{Tr}(\rho^\omega(Q_A^\dagger)^m Q_A^m) + \mathrm{Tr}(\rho^\omega(Q_A^\dagger)^{m-1}[Q_A, Q_A^\dagger]Q_A^{m-1}) \right\} \\
&\leq a_{m-2}\left\{ a_m + 4o^2 N a_{m-1} \right\}.
\end{aligned}
\tag{14.204}
$$

Let us now assume $a_{m-2} \neq 0, a_{m-1} \neq 0$, which is true for $m = 2$. We then get

$$
\frac{a_m}{a_{m-1}} \geq \frac{a_{m-1}}{a_{m-2}} - 4o^2 N.
\tag{14.205}
$$

We can now use this relation iteratively to obtain

$$
\begin{aligned}
\frac{a_m}{a_{m-1}} &\geq a_1 - 4o^2 N(m - 2) \\
&\geq 2(\mu o N)^2\left[1 - \frac{1 + |A|}{\mu^2 N} - \frac{2(m-2)}{\mu^2 N} \right] \\
&\geq 2(\mu o N)^2\left[1 - \frac{1 + |A|}{\mu^2 N} - \frac{2M}{\mu^2 N} \right] \\
&\geq 2(\mu o N)^2\left[1 - \frac{1 + |A|}{16|A|^2} - \frac{1}{8|A|} \right] \\
&\geq 2(\mu o N)^2\left[\frac{16 - 2 - 2}{16} \right] = (\mu o N)^2 \frac{3}{2} > (\mu o N)^2,
\end{aligned}
\tag{14.206}
$$

where the inequalities $N \geq \frac{16|A|^2}{\mu^2}$, $|\frac{M}{N}| \leq \frac{\mu^2}{16|A|}$ and $|A| \geq 1$ were used. This proves the first statement. We can now lower bound the ratio

$$
r_A^{(M)} = \left| \frac{\mathrm{Tr}\left((O^+)^M\rho^\omega(O^-)^M\right)}{\mathrm{Tr}(Q_A^M\rho^\omega(Q_A^\dagger)^M)} \right|.
\tag{14.207}
$$

To do this, let us use again a binomial expansion:

$$
\begin{aligned}
\left| \mathrm{Tr}\left((O^+)^M\rho^\omega(O^-)^M\right) \right| &\geq |\mathrm{Tr}(Q_A^M\rho^\omega(Q_A^\dagger)^M)| \\
&- \sideset{}{'}\sum_{k,l}\left| \binom{M}{k}\binom{M}{l}\mathrm{Tr}\left(\rho^\omega(Q_A^\dagger)^{M-k}(R_A^*)^k Q_A^{M-l}R_A^{M-l}\right) \right|.
\end{aligned}
$$

Here, the primed sum goes again over all $k, l = 0, \ldots, M$ except for $k = l = 0$. A further application of the Schwartz inequality together with (14.182) then yields

$$
\begin{aligned}
r_A^{(M)} &\geq 1 - \sideset{}{'}\sum_{k,l}(|A|o)^{k+l}(\mu o N)^{-(k+l)} \geq 1 - \left[\left(1 + \frac{|A|}{\mu N}\right)^{2M} - 1 \right] \\
&\geq 2 - \exp\left(\frac{2|A|M}{\mu N}\right) \geq 2 - e^{\mu/8}.
\end{aligned}
$$

Note that $r_A^{(M)} > 0$ since $0 < \mu \leq 1$.

$\square$

15 Bibliography

[1] H. Wilming, R. Gallego, and J. Eisert. "Second law of thermodynamics under control restrictions". *Phys. Rev. E* 93.4 (2016). DOI: 10.1103/physreve.93.042126.

[2] R. Gallego, J. Eisert, and H. Wilming. "Thermodynamic work from operational principles". *New J. Phys.* 18.10 (2016), p. 103017. DOI: 10.1088/1367-2630/18/10/103017.

[3] M. Perarnau-Llobet, A. Riera, R. Gallego, H. Wilming, and J. Eisert. "Work and entropy production in generalised Gibbs ensembles". *New J. Phys.* 18.12 (2016), p. 123035. DOI: 10.1088/1367-2630/aa4fa6.

[4] H. Wilming, M. J. Kastoryano, A. H. Werner, and J. Eisert. "Emergence of spontaneous symmetry breaking in dissipative lattice systems". *J. Math. Phys.* 58.3 (2017), p. 033302. DOI: 10.1063/1.4978328.

[5] F. Pastawski, J. Eisert, and H. Wilming. "Towards Holography via Quantum Source-Channel Codes". *Phys. Rev. Lett.* 119.2 (2017), p. 020501. DOI: 10.1103/PhysRevLett.119.020501.

[6] J Lekscha, H Wilming, J Eisert, and R Gallego. "Quantum thermodynamics with local control". *Phys. Rev. E* 97.2 (2018), p. 022142. DOI: 10.1103/PhysRevE.97.022142. arXiv: 10.1103/PhysRevD.97.086015.

[7] H. Wilming, R. Gallego, and J. Eisert. "Axiomatic Characterization of the Quantum Relative Entropy and Free Energy". *Entropy* 19.6 (2017), p. 241. DOI: 10.3390/e19060241.

[8] H. Wilming and R. Gallego. "Third Law of Thermodynamics as a Single Inequality". *Phys. Rev. X* 7 (4 2017), p. 041033. DOI: 10.1103/PhysRevX.7.041033.

[9] M. Perarnau-Llobet, H. Wilming, A. Riera, R. Gallego, and J. Eisert. "Strong Coupling Corrections in Quantum Thermodynamics". *Phys. Rev. Lett.* 120.12 (2018). DOI: 10.1103/physrevlett.120.120602. arXiv: 1704.05864.

[10] H. Wilming, M. Goihl, C. Krumnow, and J. Eisert. "Towards local equilibration in closed interacting quantum many-body systems" (2017). arXiv: 1704.06291 [quant-ph].

[11] P. Boes, H. Wilming, J. Eisert, and R. Gallego. "Statistical ensembles without typicality". *Nature Comm.* 9.1 (2018), p. 1022. DOI: 10.1038/s41467-018-03230-y. arXiv: 1707.08218.

[12] R. Gallego, H. Wilming, J. Eisert, and C. Gogolin. "What it takes to avoid equilibration". *Phys. Rev. A* 98.2 (2018). DOI: 10.1103/physreva.98.022135. arXiv: 1711.09832.

[13] J. D. Bekenstein. "Black holes and entropy". *Physical Review D* 7.8 (1973), p. 2333. DOI: 10.1103/PhysRevD.7.2333.

[14] J. M. Bardeen, B. Carter, and S. W. Hawking. "The four laws of black hole mechanics". *Commun. Math. Phys.* 31.2 (1973), pp. 161–170. DOI: 10.1007/BF01645742.

[15] S. W. Hawking. "Particle creation by black holes". *Commun. Math. Phys.* 43.3 (1975), pp. 199–220. DOI: 10.1007/BF02345020.

[16] J. D. Bekenstein. "Universal upper bound on the entropy-to-energy ratio for bounded systems". *Physical Review D* 23.2 (1981), p. 287. DOI: 10.1007/BF01645742.

[17] G. Hooft. "Dimensional reduction in quantum gravity" (1993). arXiv: gr-qc/9310026.

[18] L. Susskind. "The world as a hologram". *J. Math. Phys.* 36.11 (1995), pp. 6377–6396. DOI: 10.1063/1.59653.

[19] J. M. Maldacena. "The Large N Limit of Superconformal Field Theories and Supergravity". *Adv. Theor. Math. Phys* 2 (1998), pp. 253–291.

[20] R. Bousso. "The holographic principle". *Rev. Mod. Phys.* 74.3 (2002), p. 825. DOI: 10.1103/RevModPhys.74.825.

[21] C. Carathéodory. "Untersuchungen über die Grundlagen der Thermodynamik". *Mathematische Annalen* 67.3 (1909), pp. 355–386. DOI: 10.1007/BF01450409.

[22] H. B. Callen. *Thermodynamics and an Introduction to Thermostatistics, 2nd Edition*. Wiley-VCH, 1985.

[23]   E. H. Lieb and J. Yngvason. "The physics and mathematics of the second law of thermo-
        dynamics". *Phys. Rep.* 310.1 (1999), pp. 1–96. DOI: 10.1016/S0370-1573(98)
        00082-9.

[24]   E. H. Lieb and J. Yngvason. "A fresh look at entropy and the second law of thermodynam-
        ics". *Statistical Mechanics*. Springer, 2000, pp. 365–370. DOI: 10.1007/978-3-662-
        10018-9_20.

[25]   L. Boltzmann. "Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen: vorgelegt
        in der Sitzung am 10. October 1872". *Sitzungsberichte Akademie der Wissenschaften* (1872).

[26]   J. Gibbs. *Elementary Principles in Statistical Mechanics: Developed with Especial Refer-
        ence to the Rational Foundations of Thermodynamics*. Charles Scribner's sons, New York,
        1902.

[27]   E. Schrödinger. "Energieaustausch nach der Wellenmechanik". *Ann. Phys.* 388.15 (1927),
        pp. 956–968. DOI: 10.1002/andp.19273881504.

[28]   L. Szilard. "über die Entropieverminderung in einem thermodynamischen System bei Ein-
        griffen intelligenter Wesen". *Zeitschrift für Physik* 53.11 (1929), pp. 840–856. DOI: 10.
        1007/BF01341281.

[29]   J. v. Neumann. "Beweis des Ergodensatzes und des H-Theorems in der neuen Mechanik".
        Zeitschrift für Physik A Hadrons and Nuclei 57 (1 1929), pp. 30–70. DOI: 10.1007/
        BF01339852.

[30]   C. E. Shannon. "A mathematical theory of communication". *The Bell System Technical
        Journal* 27.3 (1948), pp. 379–423.

[31]   E. Jaynes. "Information theory and statistical mechanics". *Phys. Rev.* 106 (1957), pp. 620–
        630. DOI: 10.1103/PhysRev.106.620.

[32]   E. Jaynes. "Information theory and statistical mechanics. II". *Phys. Rev.* 108 (1957), pp. 171–
        190. DOI: 10.1103/PhysRev.108.171.

[33]   R. Landauer. "Irreversibility and heat generation in the computing process". *IBM J. Res.
        Dev.* 5.3 (1961), pp. 183–191. DOI: 10.1147/rd.53.0183.

[34]   D. Ruelle. *Statistical mechanics: Rigorous results*. World Scientific, 1999.

[35]   C. H. Bennett. "The thermodynamics of computation – a review". *Int. J. Theor. Phys.* 21.12
        (1982), pp. 905–940. DOI: 10.1007/BF02084158.

[36]   C. H. Bennett and G. Grinstein. "Role of irreversibility in stabilizing complex and noner-
        godic behavior in locally interacting discrete systems". *Phys. Rev. Lett.* 55.7 (1985), p. 657.
        DOI: 10.1103/PhysRevLett.55.657.

[37]   C. Jarzynski. "Nonequilibrium Equality for Free Energy Differences". *Phys. Rev. Lett.* 78
        (1997), p. 2690. DOI: 10.1103/PhysRevLett.78.2690.

[38]   D Ter Haar. "Foundations of statistical mechanics". *Rev. Mod. Phys.* 27.3 (1955), p. 289.
        DOI: 10.1103/RevModPhys.27.289.

[39]   J Uffink. "Compendium of the Foundations of Classical Statistical Physics, Handbook for
        Philosophy of Physics, eds". *Butterfield, J. and Earman, J* (2006).

[40]   A. Polkovnikov, K. Sengupta, A. Silva, and M. Vengalattore. "Nonequilibrium dynamics
        of closed interacting quantum systems". *Rev. Mod. Phys.* 83 (2011), pp. 863–883. DOI:
        10.1103/RevModPhys.83.863.

[41]   L. D'Alessio, Y. Kafri, A. Polkovnikov, and M. Rigol. "From quantum chaos and eigenstate
        thermalization to statistical mechanics and thermodynamics". *Adv. Phys.* 65 (2016), p. 239.
        DOI: 10.1080/00018732.2016.1198134.

[42]   C. Gogolin and J. Eisert. "Equilibration, thermalisation, and the emergence of statistical
        mechanics in closed quantum systems". *Rep. Prog. Phys.* 79.5 (2016), p. 56001. DOI: 10.
        1088/0034-4885/79/5/056001. arXiv: 1503.07538.

[43]   A. Zee. *Quantum field theory in a nutshell*. Princeton university press, 2010.

[44]   G. E. Crooks. "Entropy production fluctuation theorem and the nonequilibrium work re-
        lation for free energy differences". *Phys. Rev. E* 60 (1999), p. 2721. DOI: 10.1103/
        PhysRevE.60.2721.

[45]   U. Seifert. "Stochastic thermodynamics, fluctuation theorems and molecular machines".
        Rep. Prog. Phys. 75.12 (2012), p. 126001. DOI: 10.1088/0034-4885/75/12/
        126001.

[46]   D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and T. Beth. "Thermodynamic Cost of Reliability
        and Low Temperatures: Tightening Landauer's Principle and the Second Law". *Int. J. Th.
        Phys.* 39 (2000), p. 2717. DOI: 10.1023/A:1026422630734.

[47]   M. Horodecki and J. Oppenheim. "Fundamental limitations for quantum and nanoscale ther-
        modynamics". *Nature Comm.* 4 (2013), p. 2059. DOI: 10.1038/ncomms3059.

[48]   F. G.S. L. Brandao, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens. "The resource theory of quantum states out of thermal equilibrium". *Phys. Rev. Lett.* 111 (2013), p. 250404. DOI: `10.1103/PhysRevLett.111.250404`.

[49]   F. G.S. L. Brandao, M. Horodecki, N. H. Y. Ng, J. Oppenheim, and S. Wehner. "The second laws of quantum thermodynamics". *PNAS* 112 (2015), pp. 3275–3279. DOI: `10.1073/pnas.1411728112`.

[50]   J. Åberg. "Truly work-like work extraction via a single-shot analysis". *Nature Comm.* 4 (2013), p. 1925. DOI: `10.1038/ncomms2712`.

[51]   J. M. Renes. "Work cost of thermal operations in quantum thermodynamics". *The European Physical Journal Plus* 129.7 (2014), p. 153. DOI: `10.1140/epjp/i2014-14153-8`.

[52]   P. Faist, F. Dupuis, J. Oppenheim, and R. Renner. "The Minimal Work Cost of Information Processing". *Nature Comm.* 6 (2015), p. 7669. DOI: `10.1038/ncomms8669`.

[53]   N. Yunger Halpern and J. M. Renes. "Beyond heat baths: Generalized resource theories for small-scale thermodynamics". *Phys. Rev. E* 93.2 (2016). DOI: `10.1103/physreve.93.022126`.

[54]   J. Åberg. "Catalytic coherence". *Phys. Rev. Lett.* 113 (2014), p. 150402. DOI: `10.1103/PhysRevLett.113.150402`.

[55]   P. Ćwikliński, M. Studziński, M. Horodecki, and J. Oppenheim. "Limitations on the Evolution of Quantum Coherences: Towards Fully Quantum Second Laws of Thermodynamics". *Phys. Rev. Lett.* 115 (2015). DOI: `10.1103/PhysRevLett.115.210403`.

[56]   M. Lostaglio, D. Jennings, and T. Rudolph. "Description of quantum coherence in thermodynamic processes requires constraints beyond free energy". *Nature Comm.* 6 (2015). DOI: `10.1038/ncomms7383`.

[57]   M. Lostaglio, K. Korzekwa, D. Jennings, and T. Rudolph. "Quantum coherence, time-translation symmetry, and thermodynamics". *Phys. Rev. X* 5.2 (2015), p. 021001. DOI: `10.1103/PhysRevX.5.021001`.

[58]   K. Korzekwa, M. Lostaglio, J. Oppenheim, and D. Jennings. "The extraction of work from quantum coherence". *New J. Phys.* 18.2 (2016), p. 023045. DOI: `10.1088/1367-2630/18/2/023045`.

[59]   J. Åberg. "Fully Quantum Fluctuation Theorems". *Phys. Rev. X* 8.1 (2018). DOI: `10.1103/physrevx.8.011019`. arXiv: `1601.01302`.

[60]   Á. M. Alhambra, L. Masanes, J. Oppenheim, and C. Perry. "Fluctuating Work: From Quantum Thermodynamical Identities to a Second Law Equality". *Phys. Rev. X* 6.4 (2016). DOI: `10.1103/physrevx.6.041017`.

[61]   L. Masanes and J. Oppenheim. "A general derivation and quantification of the third law of thermodynamics". *Nature Comm.* 8 (2017). DOI: `10.1038/ncomms14538`.

[62]   M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information.* Cambridge University Press, 2000.

[63]   J.-S. Caux and J. Mossel. "Remarks on the notion of quantum integrability". *J. Stat. Mech.* (2011). DOI: `10.1088/1742-5468/2011/02/P02023`.

[64]   P Ullersma. "An exactly solvable model for Brownian motion: II. derivation of the Fokker-Planck equation and the master equation". *Physica* 32.1 (1966), pp. 56–73. DOI: `10.1016/0031-8914(66)90103-0`.

[65]   A. O. Caldeira and A. J. Leggett. "Path integral approach to quantum Brownian motion". *Physica A* 121.3 (1983), pp. 587–616. DOI: `10.1016/0378-4371(83)90013-4`.

[66]   F. Martinelli. *Lectures on probability theory and statistics.* Vol. 1717. Lecture Notes in Math. Berlin: Springer, 1997, pp. 93–191.

[67]   D. Jennings and T. Rudolph. "Entanglement and the thermodynamic arrow of time". *Physical Review E* 81.6 (2010), p. 061130. DOI: `10.1103/PhysRevE.81.061130`.

[68]   N. H. Y. Ng, L Mančinska, C Cirstoiu, J Eisert, and S Wehner. "Limits to catalysis in quantum thermodynamics". *New J. Phys.* 17.8 (2015), p. 085004. DOI: `10.1088/1367-2630/17/8/085004`.

[69]   M. Tomamichel. "Quantum Information Processing with Finite Resources". *SpringerBriefs in Mathematical Physics* (2016). DOI: `10.1007/978-3-319-21891-5`.

[70]   L. del Rio, J. Åberg, R. Renner, O. Dahlsten, and V. Vedral. "The thermodynamic meaning of negative entropy". *Nature* 474 (2011), p. 61. DOI: `10.1038/nature10123`.

[71]   M. Esposito, U. Harbola, and S. Mukamel. "Nonequilibrium fluctuations, fluctuation theorems, and counting statistics in quantum systems". *Rev. Mod. Phys.* 81.4 (2009), p. 1665. DOI: `10.1103/RevModPhys.81.1665`.

[72]   M. Campisi, P. Hänggi, and P. Talkner. "Colloquium: Quantum fluctuation relations: Foundations and applications". *Rev. Mod. Phys.* 83.3 (2011), p. 771. DOI: 10.1103/RevModPhys.83.771.

[73]   H. Tasaki. "Jarzynski relations for quantum systems and some applications" (2000). arXiv: cond-mat/0009244.

[74]   M. Esposito and S. Mukamel. "Fluctuation theorems for quantum master equations". *Physical Review E* 73.4 (2006), p. 046129. DOI: 10.1103/PhysRevE.73.046129.

[75]   G. E. Crooks. *On the fluctuations of dissipation: an annotated bibliography*. 2008.

[76]   M. Campisi, P. Talkner, and P. Hänggi. "Thermodynamics and fluctuation theorems for a strongly coupled open quantum system: an exactly solvable case". *J. Phys. A: Math. Theor.* 42.39 (2009), p. 392002. DOI: 10.1088/1751-8113/42/39/392002.

[77]   M. Campisi, P. Talkner, and P. Hänggi. "Fluctuation theorem for arbitrary open quantum systems". *Phys. Rev. Lett.* 102.21 (2009), p. 210401. DOI: 10.1103/PhysRevLett.102.210401.

[78]   N. Yunger Halpern, A. J. P. Garner, O. C. O. Dahlsten, and V. Vedral. "Introducing one-shot work into fluctuation relations". *New J. Phys.* 17.9 (2015), p. 095003. DOI: 10.1088/1367-2630/17/9/095003.

[79]   M. Perarnau-Llobet, E. Bäumer, K. V. Hovhannisyan, M. Huber, and A. Acin. "No-Go Theorem for the Characterization of Work Fluctuations in Coherent Quantum Systems". *Phys. Rev. Lett.* 118.7 (2017), p. 070601. DOI: 10.1103/PhysRevLett.118.070601.

[80]   P. Skrzypczyk, A. J. Short, and S. Popescu. "Work extraction and thermodynamics for individual quantum systems". *Nature Comm.* 5 (2014), p. 4185. DOI: 10.1038/ncomms5185.

[81]   C. Gardiner and P. Zoller. *Quantum noise: a handbook of Markovian and non-Markovian quantum stochastic methods with applications to quantum optics*. Vol. 56. Springer Science & Business Media, 2004.

[82]   R. Alicki and K. Lendi. *Quantum Dynamical Semigroups and Applications*. Vol. 717. Lecture Notes in Physics. Springer Berlin Heidelberg, 2007. DOI: 10.1007/b11976790.

[83]   G. Lindblad. "On the generators of quantum dynamical semigroups." *Commun. Math. Phys.* 48 (1976), pp. 119–130. DOI: 10.1007/BF01608499.

[84]   E. B. Davies. "Markovian master equations". *Commun. Math. Phys.* 39.2 (1974), pp. 91–110. DOI: 10.1007/BF01608389.

[85]   F. Schlögl. "Zur statistischen Theorie der Entropieproduktion in nicht abgeschlossenen Systemen". *Zeitschrift für Physik* 191.1 (1966), pp. 81–90. DOI: 10.1007/BF01362471.

[86]   H. Spohn. "Entropy production for quantum dynamical semigroups". *J. Math. Phys.* 19.5 (1978), p. 1227. DOI: 10.1063/1.523789.

[87]   H. Spohn and J. L. Lebowitz. "Irreversible thermodynamics for quantum systems weakly coupled to thermal reservoirs". *Adv. Chem. Phys* 38 (1978), pp. 109–142. DOI: 10.1002/9780470142578.ch2.

[88]   R. Alicki. "The quantum open system as a model of the heat engine". *J. Phys. A: Math. Gen.* 12.5 (1979), p. L103. DOI: 10.1088/0305-4470/12/5/007.

[89]   M. Esposito, K. Lindenberg, and C. Van den Broeck. "Entropy production as correlation between system and reservoir". *New J. Phys.* 12.1 (2010), p. 013013. DOI: 10.1088/1367-2630/12/1/013013.

[90]   J. G. Richens, Á. M. Alhambra, and L. Masanes. "Finite-bath corrections to the second law of thermodynamics". *Phys. Rev. E* 97.6 (2018). DOI: 10.1103/physreve.97.062132. arXiv: 1702.03357.

[91]   M. Lostaglio, Á. M. Alhambra, and C. Perry. "Elementary Thermal Operations". *Quantum* 2 (2018), p. 52. DOI: 10.22331/q-2018-02-08-52. arXiv: 1607.00394v2.

[92]   N. Yunger Halpern. "Beyond heat baths II: framework for generalized thermodynamic resource theories". *J. Phys. A: Math. Theor.* 51.9 (2018), p. 094001. DOI: 10.1088/1751-8121/aaa62f. arXiv: 1409.7845.

[93]   N. Yunger Halpern, P. Faist, J. Oppenheim, and A. Winter. "Microcanonical and resource-theoretic derivations of the thermal state of a quantum system with noncommuting charges". *Nature Comm.* 7 (2016). DOI: 10.1038/ncomms12051.

[94]   M. Horodecki and J. Oppenheim. "(Quantumness in the context of) Resource Theories". *Int. J. Mod. Phys. B* 27 (2013). DOI: 10.1142/S0217979213450197.

[95]   T. Fritz. "Resource convertibility and ordered commutative monoids". *Math. Struct. Comp. Sci.* (2015), pp. 850–938. DOI: 10.1017/S0960129515000444.

[96]   L. del Rio. "Resource theories of knowledge". PhD thesis. ETH Zurich, 2015.

[97] L. del Rio, L. Krämer, and R. Renner. "Resource theories of knowledge" (2015). arXiv: 1511.08818.

[98] F. G.S. L. Brandao and G. Gour. "The general structure of quantum resource theories". *Phys. Rev. Lett.* 115 (2015), p. 070503. DOI: 10.1103/PhysRevLett.115.070503.

[99] B. Coecke, T. Fritz, and R. W. Spekkens. "A mathematical theory of resources". *Information and Computation* 250 (2016), pp. 59–86. DOI: 10.1016/j.ic.2016.02.008.

[100] L. Krämer and L. del Rio. "Currencies in resource theories" (2016). arXiv: 1605.01064.

[101] J. Oppenheim, M. Horodecki, P. Horodecki, and R. Horodecki. "Thermodynamical approach to quantifying quantum correlations". *Phys. Rev. Lett.* 89.18 (2002), p. 180402. DOI: 10.1103/PhysRevLett.89.180402.

[102] M. Horodecki, J. Oppenheim, and R. Horodecki. "Are the laws of entanglement theory thermodynamical?" *Phys. Rev. Lett.* 89.24 (2002), p. 240403. DOI: 10.1103/PhysRevLett.89.240403.

[103] M. Weilenmann, L. Krämer, P. Faist, and R. Renner. "Axiomatic relation between thermodynamic and information-theoretic entropies". *Phys. Rev. Lett.* 117 (2015), p. 260601. DOI: 10.1103/PhysRevLett.117.260601. arXiv: 1501.06920v2 [quant-ph].

[104] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Yunger Halpern. "The resource theory of informational nonequilibrium in thermodynamics". *Phys. Rep.* 583 (2015), pp. 1–58. DOI: 10.1016/j.physrep.2015.04.003.

[105] M. Lostaglio, M. P. Müller, and M. Pastena. "Stochastic Independence as a Resource in Small-Scale Thermodynamics". *Phys. Rev. Lett.* 115 (15 2015), p. 150402. DOI: 10.1103/PhysRevLett.115.150402.

[106] Á. M. Alhambra, L. Masanes, J. Oppenheim, and C. Perry. "Entanglement fluctuation theorems" (2017). arXiv: 1709.06139.

[107] S. D. Bartlett, T. Rudolph, and R. W. Spekkens. "Reference frames, superselection rules, and quantum information". *Rev. Mod. Phys.* 79 (2 2007), pp. 555–609. DOI: 10.1103/RevModPhys.79.555.

[108] J. A. Vaccaro, F Anselmi, H. M. Wiseman, and K. Jacobs. "Tradeoff between extractable mechanical work, accessible entanglement, and ability to act as a reference system, under arbitrary superselection rules". *Phys. Rev. A* 77.3 (2008), p. 032114. DOI: 10.1103/PhysRevA.77.032114.

[109] G. Gour and R. W. Spekkens. "The resource theory of quantum reference frames: manipulations and monotones". *New J. Phys.* 10.3 (2008), p. 033023. DOI: 10.1088/1367-2630/10/3/033023.

[110] T Baumgratz, M Cramer, and M. Plenio. "Quantifying coherence". *Phys. Rev. Lett.* 113.14 (2014), p. 140401. DOI: 10.1103/PhysRevLett.113.140401.

[111] I. Marvian and R. W. Spekkens. "Modes of asymmetry: the application of harmonic analysis to symmetric quantum dynamics and quantum reference frames". *Phys. Rev. A* 90.6 (2014), p. 062110. DOI: 10.1103/PhysRevA.90.062110.

[112] I. Marvian and R. W. Spekkens. "Extending Noether's theorem by quantifying the asymmetry of quantum states". *Nature Comm.* 5.3821 (2014). DOI: 10.1038/ncomms4821.

[113] I. Marvian and R. W. Spekkens. "How to quantify coherence: Distinguishing speakable and unspeakable notions". *Phys. Rev. A* 94 (5 2016), p. 052324. DOI: 10.1103/PhysRevA.94.052324.

[114] A. Streltsov, G. Adesso, and M. B. Plenio. "Colloquium : Quantum coherence as a resource". *Reviews of Modern Physics* 89.4 (2017). DOI: 10.1103/revmodphys.89.041003. arXiv: 1609.02439.

[115] A. Winter and D. Yang. "Operational Resource Theory of Coherence". *Phys. Rev. Lett.* 116 (12 2016), p. 120404. DOI: 10.1103/PhysRevLett.116.120404.

[116] R. Gallego and L. Aolita. "Resource theory of steering". *Phys. Rev. X* 5.4 (2015), p. 041008. DOI: 10.1103/PhysRevX.5.041008.

[117] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson. "The resource theory of stabilizer quantum computation". *New J. Phys.* 16.1 (2014), p. 013009. DOI: 10.1088/1367-2630/16/1/013009.

[118] C. Matt, U. Maurer, C. Portmann, R. Renner, and B. Tackmann. "Toward an algebraic theory of systems". *Theoretical Computer Science* 747 (2018), pp. 1–25. DOI: 10.1016/j.tcs.2018.06.001. arXiv: 1609.04293.

[119] W. Nernst. "Über die Berechnung chemischer Gleichgewichte aus thermischen Messungen". *Nachr. Kgl. Ges. Wiss. Gött.* 1 (1906), pp. 1–40.

[120] W. Nernst. "Über die Beziehung zwischen Wärmeentwicklung und maximaler Arbeit bei kondensierten Systemen". *Ber. Kgl. Pr. Akad. Wiss.* 52 (1906), pp. 933–940.

[121] J. Scharlau and M. P. Mueller. "Quantum Horn's lemma, finite heat baths, and the third law of thermodynamics". *Quantum* 2 (2018), p. 54. DOI: 10.22331/q-2018-02-22-54. arXiv: 1605.06092.

[122] R. Silva, G. Manzano, P. Skrzypczyk, and N. Brunner. "Performance of autonomous quantum thermal machines: Hilbert space dimension as a thermodynamical resource". *Phys. Rev. E* 94.3 (2016). DOI: 10.1103/physreve.94.032120. arXiv: 1604.04098.

[123] C. Di Franco and M. Paternostro. "A no-go result on the purification of quantum states". *Sci. Rep.* 3 (2013). DOI: 10.1038/srep01387.

[124] L.-A. Wu, D. Segal, and P. Brumer. "No-go theorem for ground state cooling given initial system-thermal bath factorization". *Sci. Rep.* 3 (2013). DOI: 10.1038/srep01824.

[125] F. Ticozzi and L. Viola. "Quantum resources for purification and cooling: fundamental limits and opportunities". *Sci. Rep.* 4 (2014). DOI: 10.1038/srep05192.

[126] L. J. Schulman and U. V. Vazirani. "Molecular scale heat engines and scalable quantum computation". *Proceedings of the thirty-first annual ACM symposium on Theory of computing - STOC '99* (1999). DOI: 10.1145/301250.301332.

[127] P. O. Boykin, T. Mor, V. Roychowdhury, F. Vatan, and R. Vrijen. "Algorithmic cooling and scalable NMR quantum computers". *PNAS* 99.6 (2002), pp. 3388–3393. DOI: 10.1073/pnas.241641898.

[128] L. J. Schulman, T. Mor, and Y. Weinstein. "Physical Limits of Heat-Bath Algorithmic Cooling". *Phys. Rev. Lett.* 94.12 (2005). DOI: 10.1103/physrevlett.94.120501.

[129] A. E. Allahverdyan, K. V. Hovhannisyan, D. Janzing, and G. Mahler. "Thermodynamic limits of dynamic cooling". *Phys. Rev. E* 84.4 (2011). DOI: 10.1103/physreve.84.041109.

[130] S. Raeisi and M. Mosca. "Asymptotic Bound for Heat-Bath Algorithmic Cooling". *Phys. Rev. Lett.* 114.10 (2015). DOI: 10.1103/physrevlett.114.100404.

[131] A. E. Allahverdyan, K. Hovhannisyan, and G. Mahler. "Optimal refrigerator". *Phys. Rev. E* 81.5 (2010). DOI: 10.1103/physreve.81.051129.

[132] N. Linden, S. Popescu, and P. Skrzypczyk. "How Small Can Thermal Machines Be? The Smallest Possible Refrigerator". *Phys. Rev. Lett.* 105 (13 2010), p. 130401. DOI: 10.1103/PhysRevLett.105.130401.

[133] P. Skrzypczyk, N. Brunner, N. Linden, and S. Popescu. "The smallest refrigerators can reach maximal efficiency". *J. Phys. A* 44.49 (2011), p. 492002. DOI: 10.1088/1751-8113/44/49/492002.

[134] A. Levy, R. Alicki, and R. Kosloff. "Quantum refrigerators and the third law of thermodynamics". *Phys. Rev. E* 85.6 (2012). DOI: 10.1103/physreve.85.061126.

[135] A. Mari and J. Eisert. "Cooling by Heating: Very Hot Thermal Light Can Significantly Cool Quantum Systems". *Phys. Rev. Lett.* 108.12 (2012). DOI: 10.1103/physrevlett.108.120602.

[136] R. Silva, P. Skrzypczyk, and N. Brunner. "Small quantum absorption refrigerator with reversed couplings". *Phys. Rev. E* 92.1 (2015). DOI: 10.1103/physreve.92.012136.

[137] P. P. Hofer, M. Perarnau-Llobet, J. B. Brask, R. Silva, M. Huber, and N. Brunner. "Autonomous quantum refrigerator in a circuit QED architecture based on a Josephson junction". *Physical Review B* 94.23 (2016). DOI: 10.1103/physrevb.94.235420. arXiv: 1607.05218.

[138] B. Cleuren, B. Rutten, and C. Van den Broeck. "Cooling by Heating: Refrigeration Powered by Photons". *Phys. Rev. Lett.* 108.12 (2012). DOI: 10.1103/physrevlett.108.120603.

[139] A. E. Allahverdyan, K. V. Hovhannisyan, and G. Mahler. "Comment on "Cooling by Heating: Refrigeration Powered by Photons"". *Phys. Rev. Lett.* 109.24 (2012). DOI: 10.1103/physrevlett.109.248903.

[140] A. Levy, R. Alicki, and R. Kosloff. "Comment on "Cooling by Heating: Refrigeration Powered by Photons"". *Phys. Rev. Lett.* 109.24 (2012). DOI: 10.1103/physrevlett.109.248901.

[141] O. Entin-Wohlman and Y. Imry. "Comment on "Cooling by Heating: Refrigeration Powered by Photons"". *Phys. Rev. Lett.* 112.4 (2014). DOI: 10.1103/physrevlett.112.048901.

[142] B. Cleuren, B. Rutten, and C. Van den Broeck. "Cleuren et al. Reply:" *Phys. Rev. Lett.* 109.24 (2012). DOI: 10.1103/physrevlett.109.248902.

[143] M. Kolář, D. Gelbwaser-Klimovsky, R. Alicki, and G. Kurizki. "Quantum Bath Refrigeration towards Absolute Zero: Challenging the Unattainability Principle". *Phys. Rev. Lett.* 109.9 (2012). DOI: 10.1103/physrevlett.109.090601.

[144]   N. Freitas and J. P. Paz. "Fundamental limits for cooling of linear quantum refrigerators". *Physical Review E* 95.1 (2017), p. 012146. DOI: 10.1103/PhysRevE.95.012146.

[145]   N. Freitas and J. P. Paz. "How much can we cool a quantum oscillator? A useful analogy to understand laser cooling as a thermodynamical process" (2017). arXiv: 1710.11554.

[146]   S. Ryu, A. Schnyder, A. Furusaki, and A. Ludwig. "Topological insulators and superconductors: ten-fold way and dimensional hierarchy". *New J. Phys.* 12 (2009), p. 065010. DOI: 10.1088/1367-2630/12/6/065010.

[147]   M. C. Fernando G. S. L. Brandao. "Equivalence of statistical mechanical ensembles for non-critical quantum systems" (2015). arXiv: 1502.03263.

[148]   R. P. Feynman, J. Hey, and R. W. Allen. *Feynman lectures on computation*. Addison-Wesley Longman Publishing Co., Inc., 1998.

[149]   C. H. Bennett. "Notes on Landauer's principle, reversible computation, and Maxwell's Demon". *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics* 34.3 (2003), pp. 501–510. DOI: 10.1016/S1355-2198(03)00039-X.

[150]   O. C. O. Dahlsten, R. Renner, E. Rieper, and V. Vedral. "Inadequacy of von Neumann entropy for characterizing extractable work". *New J. Phys.* 13.5 (2011), p. 053015. DOI: 10.1088/1367-2630/13/5/053015.

[151]   D Egloff, O. C. O. Dahlsten, R Renner, and V Vedral. "A measure of majorization emerging from single-shot statistical mechanics". *New J. Phys.* 17.7 (2015), p. 073001. DOI: 10.1088/1367-2630/17/7/073001.

[152]   M. P. Woods, N. Ng, and S. Wehner. "The maximum efficiency of nano heat engines depends on more than temperature" (2015). arXiv: 1506.02322.

[153]   N. H. Y. Ng, M. Woods, and S. Wehner. "Surpassing the Carnot Efficiency by extracting imperfect work". *New J. Phys.* (2017). DOI: 10.1088/1367-2630/aa8ced.

[154]   J. Gemmer and J. Anders. "From single-shot towards general work extraction in a quantum thermodynamic framework". *New J. Phys.* 17.8 (2015), p. 085006. DOI: 10.1088/1367-2630/17/8/085006.

[155]   K. Matsumoto. "Reverse test and characterization of quantum relative entropy" (2010). arXiv: 1010.1030.

[156]   F. Hiai and D. Petz. "The Proper Formula for Relative Entropy and its Asymptotics in Quantum Probability". *Asymptotic Theory of Quantum Statistical Inference*. World Scientific Pub Co Pte Lt, 2005, pp. 43–63. DOI: 10.1142/9789812563071_0004.

[157]   T. Ogawa and H. Nagaoka. "Strong Converse and Stein's Lemma in Quantum Hypothesis Testing". *Asymptotic Theory of Quantum Statistical Inference*. World Scientific Pub Co Pte Lt, 2005, pp. 28–42. DOI: 10.1142/9789812563071_0003.

[158]   P. Faist, J. Oppenheim, and R. Renner. "Gibbs-Preserving maps outperform thermal operations in the quantum regime". *New J. Phys.* 17 (2015), p. 043003. DOI: 10.1088/1367-2630/17/4/043003.

[159]   V. Vedral. "The role of relative entropy in quantum information theory". *Rev. Mod. Phys.* 74.1 (2002), pp. 197–234. DOI: 10.1103/revmodphys.74.197.

[160]   F. G. Brandao and M. B. Plenio. "A generalization of quantum stein's Lemma". *Commun. Math. Phys.* 295 (2010), p. 791.

[161]   V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight. "Quantifying Entanglement". *Phys. Rev. Lett.* 78.12 (1997), p. 2275. DOI: 10.1103/physrevlett.78.2275.

[162]   M. Junge, R. Renner, D. Sutter, M. M. Wilde, and A. Winter. "Universal recoverability in quantum information". *2016 IEEE International Symposium on Information Theory (ISIT)*. Institute of Electrical and Electronics Engineers (IEEE), 2016, pp. 2494–2498. DOI: 10.1109/isit.2016.7541748.

[163]   N. Lashkari and M. Van Raamsdonk. "Canonical energy is quantum Fisher information". *JHEP* 2016.4 (2016). DOI: 10.1007/JHEP04(2016)153.

[164]   M. P. Müller. "Correlating thermal machines and the second law at the nanoscale" (2017). arXiv: 1707.03451.

[165]   S. Goldstein, J. L. Lebowitz, R. Tumulka, and N. Zanghí. "Canonical typicality". *Phys. Rev. Lett.* 96.5 (2006), p. 050403. DOI: 10.1103/PhysRevLett.96.050403.

[166]   S. Popescu, A. J. Short, and A. Winter. "Entanglement and the foundations of statistical mechanics". *Nature Phys.* 2.11 (2006), pp. 754–758. DOI: 10.1038/nphys444.

[167]   P. Reimann. "Foundation of Statistical Mechanics under Experimentally Realistic Conditions". *Phys. Rev. Lett.* 101.19 (2008), p. 190403. DOI: 10.1103/PhysRevLett.101.190403. arXiv: 0810.3092.

[168]   S. Goldstein, J. L. Lebowitz, C. Mastrodonato, R. Tumulka, and N. Zanghì. "Normal typicality and von Neumann's quantum ergodic theorem". *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. Vol. 466. 2123. The Royal Society. 2010, pp. 3203–3224. DOI: `10.1098/rspa.2009.0635`.

[169]   I. Bloch, J. Dalibard, and W. Zwerger. "Many-body physics with ultracold gases". *Rev. Mod. Phys.* 80 (3 2008), pp. 885–964. DOI: `10.1103/RevModPhys.80.885`.

[170]   I. Bloch, J. Dalibard, and S. Nascimbene. "Quantum simulations with ultracold quantum gases". *Nature Phys.* 8 (2012), pp. 267–76. DOI: `10.1038/nphys2259`.

[171]   S. Trotzky, Y.-A. Chen, A. Flesch, I. P. McCulloch, U. Schollwöck, J. Eisert, and I. Bloch. "Probing the relaxation towards equilibrium in an isolated strongly correlated one-dimensional Bose gas". *Nature Phys.* 8 (2012), pp. 325–330. DOI: `10.1038/nphys2232`.

[172]   M. Gring, M. Kuhnert, T. Langen, T. Kitagawa, B. Rauer, M. Schreitl, I. Mazets, D. A. Smith, E. Demler, and J. Schmiedmayer. "Relaxation and pre-thermalization in an isolated quantum system". *Science* 337 (2012), p. 1318. DOI: `10.1126/science.1224953`.

[173]   U. Schneider, L. Hackermuller, J. P. Ronzheimer, S. Will, S. Braun, T. Best, I. Bloch, E. Demler, S. Mandt, D. Rasch, and A. Rosch. "Fermionic transport and out-of-equilibrium dynamics in a homogeneous Hubbard model with ultracold atoms". *Nature Phys.* 8 (2012), p. 213. DOI: `10.1038/nphys2205`.

[174]   S. Braun, M. Friesdorf, S. S. Hodgman, M. Schreiber, J. P. p Ronzheimer, A. Riera, M. del Rey, I. Bloch, J. Eisert, and U. Schneider. "Emergence of coherence and the dynamics of quantum phase transitions". *Proc. Natl. Ac. Sc.* 112 (2015), p. 3641. DOI: `10.1073/pnas.1408861112`.

[175]   P Bocchieri and A Loinger. "Ergodic foundation of quantum statistical mechanics". *Phys. Rev.* 114.4 (1959), p. 948. DOI: `10.1103/PhysRev.114.948`.

[176]   H. Tasaki. "From Quantum Dynamics to the Canonical Distribution: General Picture and a Rigorous Example". *Phys. Rev. Lett.* 80 (7 1998), pp. 1373–1376. DOI: `10.1103/PhysRevLett.80.1373`.

[177]   N. Linden, S. Popescu, A. Short, and A. Winter. "Quantum mechanical evolution towards thermal equilibrium". *Physical Review E* 79.6 (2009), p. 61103. DOI: `10.1103/PhysRevE.79.061103`. arXiv: `0812.2385`.

[178]   M. Cramer and J. Eisert. "A quantum central limit theorem for non-equilibrium systems: Exact local relaxation of correlated states". *New J. Phys.* 12 (2010), p. 055020. DOI: `10.1088/1367-2630/12/5/055020`.

[179]   N. Linden, S. Popescu, A. J. Short, and A. Winter. "On the speed of fluctuations around thermodynamic equilibrium". *New J. Phys.* 12 (2010). DOI: `10.1088/1367-2630/12/5/055021`. arXiv: `0907.1267`.

[180]   P. Reimann. "Equilibration of isolated macroscopic quantum systems under experimentally realistic conditions". *Phys. Scr.* 86.5 (2012), p. 58512. DOI: `10.1088/0031-8949/86/05/058512`. arXiv: `1210.5821`.

[181]   P. Reimann and M. Kastner. "Equilibration of isolated macroscopic quantum systems". *New J. Phys.* 14.4 (2012), p. 43020. DOI: `10.1088/1367-2630/14/4/043020`. arXiv: `1202.2768`.

[182]   A. J. Short and T. C. Farrelly. "Quantum equilibration in finite time". *New J. Phys.* 14.1 (2012), p. 013063. DOI: `10.1088/1367-2630/14/1/013063`. arXiv: `1110.5759`.

[183]   S. Goldstein, T. Hara, and H. Tasaki. "Time scales in the approach to equilibrium of macroscopic quantum systems". *Phys. Rev. Lett.* 111.14 (2013), p. 140401. DOI: `10.1103/PhysRevLett.111.140401`.

[184]   L. Masanes, A. J. Roncaglia, and A. Acín. "Complexity of energy eigenstates as a mechanism for equilibration". *Physical Review E* 87.3 (2013), p. 32137. DOI: `10.1103/PhysRevE.87.032137`. arXiv: `1108.0374`.

[185]   A. S. L. Malabarba, L. P. García-Pintos, N. Linden, T. C. Farrelly, and A. J. Short. "Quantum systems equilibrate rapidly for most observables". *Physical Review E* 90.1 (2014), p. 012121. DOI: `10.1103/PhysRevE.90.012121`. arXiv: `arXiv:1402.1093v1`.

[186]   L. P. García-Pintos, N. Linden, A. S. Malabarba, A. J. Short, and A. Winter. "Equilibration time scales of physically relevant observables". *Phys. Rev. X* 7.3 (2017), p. 031027. DOI: `10.1103/PhysRevX.7.031027`.

[187]   S. Goldstein, T. Hara, and H. Tasaki. "Extremely quick thermalization in a macroscopic quantum system for a typical nonequilibrium subspace". *New J. Phys.* 17.4 (2015), p. 045002. DOI: `10.1088/1367-2630/17/4/045002`.

[188]   B. N. Balz and P. Reimann. "Equilibration of isolated many-body quantum systems with respect to general distinguishability measures". *Phys. Rev. E* 93 (2016), p. 062107. DOI: `10.1103/PhysRevE.93.062107`.

[189] M. Gluza, C. Krumnow, M. Friesdorf, C. Gogolin, and J. Eisert. "Equilibration via Gaussification in fermionic lattice systems". *Phys. Rev. Lett.* 117.19 (2016), p. 190602. DOI: 10.1103/PhysRevLett.117.190602.

[190] T. Farrelly, F. G. Brandão, and M. Cramer. "Thermalization and Return to Equilibrium on Finite Quantum Lattice Systems". *Phys. Rev. Lett.* 118.14 (2017), p. 140601. DOI: 10.1103/PhysRevLett.118.140601.

[191] T. Farrelly. "Equilibration of quantum gases". *New J. Phys.* 18.7 (2016), p. 073014. DOI: 10.1088/1367-2630/18/7/073014.

[192] P. Calabrese and J. Cardy. "Time dependence of correlation functions following a quantum quench". *Phys. Rev. Lett.* (2006), p. 136801. DOI: 10.1103/PhysRevLett.96.136801.

[193] P. Calabrese and J.-S. Caux. "Dynamics of the attractive 1D Bose gas: Analytical treatment from integrability". *J. Stat. Mech.* (2007), P08032. DOI: 10.1088/1742-5468/2007/08/P08032.

[194] V. Eisler and I. Peschel. "Evolution of entanglement after a local quench". *J. Stat. Mech: Theory Exp.* 2007.06 (2007), P06005. DOI: 10.1088/1742-5468/2007/06/P06005.

[195] M. Rigol, V. Dunjko, V. Yurovsky, and M. Olshanii. "Relaxation in a completely integrable many-body quantum system: An ab initio study of the dynamics of the highly excited states of 1D lattice hard-core bosons". *Phys. Rev. Lett.* 98 (2007), p. 050405. DOI: 10.1103/PhysRevLett.98.050405.

[196] M. Cramer, C. M. Dawson, J. Eisert, and T. J. Osborne. "Exact relaxation in a class of non-equilibrium quantum lattice systems". *Phys. Rev. Lett.* 100 (2008), p. 030602. DOI: 10.1103/PhysRevLett.100.030602.

[197] M. Cramer, A. Flesch, I. P. McCulloch, U. Schollwöck, and J. Eisert. "Exploring local quantum many-body relaxation by atoms in optical superlattices". *Phys. Rev. Lett.* 101 (2008), p. 063001. DOI: 10.1103/PhysRevLett.101.063001.

[198] T. Barthel and U. Schollwöck. "Dephasing and the steady state in quantum many-particle systems". *Phys. Rev. Lett.* 100.10 (2008), p. 100601. DOI: 10.1103/PhysRevLett.100.100601.

[199] A. Flesch, M. Cramer, I. P. McCulloch, U. Schollwoeck, and J. Eisert. "Probing local relaxation of cold atoms in optical superlattices". *Phys. Rev. A* 78 (2008), p. 033608. DOI: 10.1103/PhysRevA.78.033608.

[200] P. Calabrese, F. H. L. Essler, and M. Fagotti. "Quantum quench in the transverse-field Ising chain". *Phys. Rev. Lett.* 106 (2011), p. 227203. DOI: 10.1103/PhysRevLett.106.227203.

[201] J.-S. Caux and F. H. L. Essler. "Time evolution of local observables after quenching to an integrable model". *Phys. Rev. Lett.* 110 (2013), p. 257203. DOI: 10.1103/PhysRevLett.110.257203.

[202] E. Ilievski, M. Medenjak, T. Prosen, and L. Zadnik. "Quasilocal charges in integrable lattice systems". *J. Stat. Mech: Theory Exp.* 2016.6 (2016), p. 064008. DOI: 10.1088/1742-5468/2016/06/064008.

[203] D. Wallace. "Recurrence theorems: a unified account". *J. Math. Phys.* 56 (2015), p. 022105. DOI: 10.1063/1.4907384.

[204] A. J. Short. "Equilibration of quantum systems and subsystems". *New J. Phys.* 13.5 (2011), p. 053009. DOI: 10.1088/1367-2630/13/5/053009.

[205] M. P. Müller, E. Adlam, L. Masanes, and N. Wiebe. "Thermalization and canonical typicality in translation-invariant quantum lattice systems". *Commun. Math. Phys.* 340.2 (2015), pp. 499–561. DOI: 10.1007/s00220-015-2473-y.

[206] M. Kliesch, C. Gogolin, M. J. Kastoryano, A. Riera, and J. Eisert. "Locality of temperature". *Phys. Rev. X* 4 (2014), p. 031019. DOI: 10.1103/PhysRevX.4.031019.

[207] L. P. Kadanoff and P. C. Martin. "Hydrodynamic equations and correlation functions". *Ann. Phys.* 24 (1963), pp. 419–469. DOI: 10.1142/9789812798763_0003.

[208] H. Spohn. *Large scale dynamics of interacting particles*. Berlin: Springer, 1991.

[209] M. Schreiber, S. S. Hodgman, P. Bordia, H. P. Lüschen, M. H. Fischer, R. Vosk, E. Altman, U. Schneider, and I. Bloch. "Observation of many-body localization of interacting fermions in a quasirandom optical lattice". *Science* 349 (2015), pp. 842–845. DOI: 10.1126/science.aaa7432.

[210] M. Serbyn, Z. Papić, and D. A. Abanin. "Quantum quenches in the many-body localized phase". *Phys. Rev. B* 90 (2014), p. 174302. DOI: 10.1103/PhysRevB.90.174302.

[211]   E. Leviatan, F. Pollmann, J. H. Bardarson, and E. Altman. "Quantum thermalization dynamics with Matrix-Product States" (2017). arXiv: 1702.08894.

[212]   M. C. Bañuls, J. I. Cirac, and M. B. Hastings. "Strong and weak thermalization of in finite non-integrable quantum systems". *Phys. Rev. Lett.* 106 (2011), p. 050405. DOI: 10.1103/PhysRevLett.106.050405.

[213]   T. R. de Oliveira, C. Charalambous, D. Jonathan, M. Lewenstein, and A. Riera. "Equilibration time scales in closed many-body quantum systems". *New Journal of Physics* 20.3 (2018), p. 033032. DOI: 10.1088/1367-2630/aab03b. arXiv: 1704.06646.

[214]   H. Kim, M. C. Bañuls, J. I. Cirac, M. B. Hastings, and D. A. Huse. "Slowest local operators in quantum spin chains". *Phys. Rev. E* 92 (2015), p. 012128. DOI: 10.1103/PhysRevE.92.012128.

[215]   C.-J. Lin and O. I. Motrunich. "Quasiparticle explanation of the weak-thermalization regime under quench in a nonintegrable quantum spin chain". *Phys. Rev. A* 95 (2017). DOI: 10.1103/PhysRevA.95.023621.

[216]   I. Arad, T. Kuwahara, and Z. Landau. "Connecting global and local energy distributions in quantum spin models on a lattice". *Journal of Statistical Mechanics: Theory and Experiment* 2016.3 (2016), p. 033301. DOI: 10.1088/1742-5468/2016/03/033301.

[217]   S. Bochner and K. Chandrasekharan. *Fourier transforms*. Princeton: Princeton University Press, 1949.

[218]   J. P. Keating, N. Linden, and H. J. Wells. "Spectra and eigenstates of spin chain Hamiltonians". *Commun. Math. Phys.* 338.1 (2015), pp. 81–102. DOI: 10.1007/s00220-015-2366-0.

[219]   A. Anshu. "Concentration bounds for quantum states with finite correlation length on quantum spin lattice systems". *New J. Phys.* 18 (2016), p. 083011. DOI: 10.1088/1367-2630/18/8/083011.

[220]   J. M. Deutsch. "Quantum statistical mechanics in a closed system". *Phys. Rev. A* 43 (1991), pp. 2046–2049. DOI: 10.1103/PhysRevA.43.2046.

[221]   M. Srednicki. "Chaos and quantum thermalization". *Phys. Rev. E* 50 (1994), pp. 888–901. DOI: 10.1103/PhysRevE.50.888.

[222]   M. Rigol, V. Dunjko, and M. Olshanii. "Thermalization and its mechanism for generic isolated quantum systems". *Nature* 452 (2008), pp. 854–858. DOI: 10.1038/nature06838.

[223]   E. H. Lieb and D. W. Robinson. "The finite group velocity of quantum spin systems". *Commun. Math. Phys.* 28 (1972), pp. 251–257. DOI: 10.1007/BF01645779.

[224]   A. Nanduri, H. Kim, and D. A. Huse. "Entanglement spreading in a many-body localized system". *Phys. Rev. B* 90 (2014), p. 064201. DOI: 10.1103/PhysRevB.90.064201.

[225]   I. H. Kim, A. Chandran, and D. A. Abanin. "Local integrals of motion and the logarithmic light cone in many-body localised systems" (2014). arXiv: 1412.3073.

[226]   M. Friesdorf, A. H. Werner, M. Goihl, J. Eisert, and W. Brown. "Local constants of motion imply information propagation". *New J. Phys.* 17 (2015), p. 113054. DOI: 10.1088/1367-2630/17/11/113054.

[227]   P. W. Anderson. "Absence of diffusion in certain random lattices". *Phys. Rev.* 109.5 (1958), p. 1492. DOI: 10.1103/PhysRev.109.1492.

[228]   C. K. Burrell and T. J. Osborne. "Bounds on the speed of information propagation in disordered quantum spin chains". *Phys. Rev. Lett.* 99 (2007), p. 167201. DOI: 10.1103/PhysRevLett.99.167201.

[229]   E. Hamza, R. Sims, and G. Stolz. "Dynamical Localization in Disordered Quantum Spin Systems". *Commun. Math. Phys.* 315.1 (2012), p. 215. DOI: 10.1007/s00220-012-1544-6.

[230]   P. Reimann. "Typical fast thermalization processes in closed many-body systems". *Nature Comm.* 7 (2016), p. 10821. DOI: 10.1038/ncomms10821. arXiv: 1603.00669.

[231]   C. Gogolin. "Environment-induced super selection without pointer states". *Physical Review E* 81.5 (2010), p. 51127. DOI: 10.1103/PhysRevE.81.051127. arXiv: 0908.2921.

[232]   A. S. L. Malabarba, A. J. Short, and P. Kammerlander. "Clock-driven quantum thermal engines". *New J. Phys.* 17.4 (2015), p. 045027. DOI: 10.1088/1367-2630/17/4/045027.

[233]   R. Alicki, M. Horodecki, P. Horodecki, and R. Horodecki. "Thermodynamics of quantum informational systems - Hamiltonian description". *Open Syst. Inf. Dyn.* 11, (2004), p. 205. DOI: 10.1023/B:OPSY.0000047566.72717.71. eprint: quant-ph/0402012.

[234]   J. Anders and V. Giovannetti. "Thermodynamics of discrete quantum processes". *New J. Phys.* 15.3 (2013), p. 033022.

[235] R Gallego, A Riera, and J Eisert. "Thermal machines beyond the weak coupling regime". *New J. Phys.* 16.12 (2014), p. 125009. DOI: 10.1088/1367-2630/16/12/125009.

[236] W Pusz and S. Woronowicz. "Passive states and KMS states for general quantum systems". *Commun. Math. Phys.* 58.3 (1978), pp. 273–290. DOI: 10.1007/BF01614224.

[237] A Lenard. "Thermodynamical proof of the Gibbs formula for elementary quantum systems". *J. Stat. Phys.* 19.6 (1978), pp. 575–586. DOI: 10.1007/BF01011769.

[238] C. Sparaciari, D. Jennings, and J. Oppenheim. "Energetic instability of passive states in thermodynamics". *Nature Comm.* 8.1 (2017). DOI: 10.1038/s41467-017-01505-4. arXiv: 1701.01703v1.

[239] E. Ruch, R. Schranner, and T. H. Seligman. "The mixing distance". *The Journal of Chemical Physics* 69.1 (1978), pp. 386–392. DOI: 10.1063/1.436364.

[240] S. Lloyd. "Almost any quantum logic gate is universal". *Phys. Rev. Lett.* 75 (1995), pp. 346–349. DOI: 10.1103/PhysRevLett.75.346.

[241] D. Burgarth, S. Bose, C. Bruder, and V. Giovannetti. "Local controllability of quantum networks". *Phys. Rev. A* 79 (2009). DOI: 10.1103/PhysRevA.79.060305.

[242] C. Gogolin, M. P. Müller, and J. Eisert. "Absence of thermalization in non-integrable systems". *Phys. Rev. Lett.* 106 (2011), p. 040401. DOI: 10.1103/PhysRevLett.106.040401.

[243] D. C. Brody, D. W. Hook, and L. P. Hughston. "Unitarity, ergodicity, and quantum thermodynamics". *J. Phys. A* 40 (2007), F503. DOI: 10.1088/1751-8113/40/26/F01.

[244] I. Peschel and V. Eisler. "Reduced density matrices and entanglement entropy in free lattice models". *J. Phys. A: Math. Theor.* 42.50 (2009), p. 504003. DOI: 10.1088/1751-8113/42/50/504003.

[245] A. C. Cassidy, C. W. Clark, and M. Rigol. "Generalized thermalization in an integrable lattice system". *Phys. Rev. Lett.* 106 (2011), p. 140405.

[246] P. Calabrese, F. H. Essler, and M. Fagotti. "Quantum quenches in the transverse field Ising chain: II. Stationary state properties". *J. Stat. Mech: Theory Exp.* 2012.07 (2012), P07022. DOI: 10.1088/1742-5468/2012/07/P07022.

[247] M. Fagotti and F. H. Essler. "Reduced density matrix after a quantum quench". *Phys. Rev. B* 87.24 (2013), p. 245107. DOI: 10.1103/PhysRevB.87.245107.

[248] Y. Guryanova, S. Popescu, A. J. Short, R. Silva, and P. Skrzypczyk. "Thermodynamics of quantum systems with multiple conserved quantities". *Nature comm.* 7 (2016).

[249] F. H. Essler and M. Fagotti. "Quench dynamics and relaxation in isolated integrable quantum spin chains". *J. Stat. Mech: Theory Exp.* 2016.6 (2016), p. 064002. DOI: 10.1088/1742-5468/2016/06/064002.

[250] A. E. Allahverdyan and T. M. Nieuwenhuizen. "Minimal work principle: Proof and counterexamples". *Phys. Rev. E* 71 (2005), p. 046107. DOI: 10.1103/PhysRevE.71.046107.

[251] I. Prigogine and D. Kondepudi. *Modern thermodynamics: From heat engines to dissipative structures, Second Edition*. John Wiley & Sons, Ltd, 2015. DOI: 10.1002/9781118698723.

[252] A. E. Allahverdyan, R. Balian, and T. M. Nieuwenhuizen. "Maximal work extraction from finite quantum systems". *EPL (Europhysics Letters)* 67.4 (2004), p. 565. DOI: 10.1209/epl/i2004-10101-2.

[253] M. Lewenstein, A. Sanpera, V. Ahufinger, B. Damski, A. Sen(De), and U. Sen. "Ultracold atomic gases in optical lattices: mimicking condensed matter physics and beyond". *Adv. Phys.* 56.2 (2007), pp. 243–379. DOI: 10.1080/00018730701223200.

[254] N. Strohmaier, Y. Takasu, K. Günther, R. Jördens, M Köhl, H. Moritz, and T. Esslinger. "Interaction-controlled transport of an ultra-cold Fermi gas". *Phys. Rev. Lett.* 99 (2007), p. 220601. DOI: 10.1103/PhysRevLett.99.220601.

[255] R. Modak and M. Rigol. "Work extraction in an isolated quantum lattice system: Grand canonical and generalized Gibbs ensemble predictions". *Phys. Rev. E* 95 (6 2017), p. 062145. DOI: 10.1103/PhysRevE.95.062145. eprint: arXiv:1704.05474.

[256] P. Hänggi, G.-L. Ingold, and P. Talkner. "Finite quantum dissipation: the challenge of obtaining specific heat". *New J. Phys.* 10.11 (2008), p. 115008. DOI: 10.1088/1367-2630/10/11/115008.

[257] M. F. Gelin and M. Thoss. "Thermodynamics of a subensemble of a canonical ensemble". *Phys. Rev. E* 79.5 (2009), p. 051121. DOI: 10.1103/PhysRevE.79.051121.

[258] I. Kim and G. Mahler. "Clausius inequality beyond the weak-coupling limit: The quantum Brownian oscillator". *Phys. Rev. E* 81.1 (2010), p. 011101. DOI: 10.1103/PhysRevE.81.011101.

[259]   M. Campisi, D. Zueco, and P. Talkner. "Thermodynamic anomalies in open quantum systems: Strong coupling effects in the isotropic XY model". *Chem. Phys.* 375 (2010), p. 187. DOI: `10.1016/j.chemphys.2010.04.026`.

[260]   S. Hilt, B. Thomas, and E. Lutz. "Hamiltonian of mean force for damped quantum systems". *Phys. Rev. E* 84.3 (2011), p. 031110. DOI: `10.1103/PhysRevE.84.031110`.

[261]   M Carrega, P Solinas, A Braggio, M Sassetti, and U Weiss. "Functional integral approach to time-dependent heat exchange in open quantum systems: general method and applications". *New J. Phys.* 17.4 (2015), p. 045030. DOI: `10.1088/1367-2630/17/4/045030`.

[262]   M. Esposito, M. A. Ochoa, and M. Galperin. "Quantum thermodynamics: A nonequilibrium Green's function approach". *Phys. Rev. Lett.* 114.8 (2015), p. 080602. DOI: `10.1103/PhysRevLett.114.080602`.

[263]   M. Esposito, M. A. Ochoa, and M. Galperin. "Nature of heat in strongly coupled open quantum systems". *Phys. Rev. B* 92.23 (2015), p. 235440. DOI: `10.1103/PhysRevB.92.235440`.

[264]   D. Gelbwaser-Klimovsky and A. Aspuru-Guzik. "Strongly coupled quantum heat machines". *The journal of physical chemistry letters* 6.17 (2015), pp. 3477–3482. DOI: `10.1021/acs.jpclett.5b01404`.

[265]   A. Bruch, M. Thomas, S. V. Kusminskiy, F. von Oppen, and A. Nitzan. "Quantum thermodynamics of the driven resonant level model". *Phys. Rev. B* 93.11 (2016), p. 115318. DOI: `10.1103/PhysRevB.93.115318`.

[266]   R. Uzdin, A. Levy, and R. Kosloff. "Quantum heat machines equivalence, work extraction beyond markovianity, and strong coupling via heat exchangers". *Entropy* 18.4 (2016), p. 124. DOI: `10.3390/e18040124`.

[267]   P. Strasberg, G. Schaller, N. Lambert, and T. Brandes. "Nonequilibrium thermodynamics in the strong coupling and non-Markovian regime based on a reaction coordinate mapping". *New J. Phys.* 18.7 (2016), p. 073007. DOI: `10.1088/1367-2630/18/7/073007`.

[268]   A. Kato and Y. Tanimura. "Quantum heat current under non-perturbative and non-Markovian conditions: Applications to heat machines". *The Journal of chemical physics* 145.22 (2016), p. 224105. DOI: `10.1063/1.4971370`.

[269]   T. Philbin and J Anders. "Thermal energies of classical and quantum damped oscillators coupled to reservoirs". *J. Phys. A: Math. Theor.* 49.21 (2016), p. 215303. DOI: `10.1088/1751-8113/49/21/215303`.

[270]   D. Newman, F. Mintert, and A. Nazir. "Performance of a quantum heat engine at strong reservoir coupling". *Phys. Rev. E* 95.3 (2017), p. 032139. DOI: `10.1103/PhysRevE.95.032139`.

[271]   R. Kubo. "Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems". *J. Phys. Soc. Jpn.* 12.6 (1957), pp. 570–586. DOI: `10.1143/JPSJ.12.570`.

[272]   H. Mori. "A quantum-statistical theory of transport processes". *J. Phys. Soc. Jpn.* 11.10 (1956), pp. 1029–1044. DOI: `10.1143/JPSJ.11.1029`.

[273]   H. Mori. "Transport, Collective Motion, and Brownian Motion". *Progress of theoretical physics* 33.3 (1965), pp. 423–455. DOI: `10.1143/PTP.33.423`.

[274]   D. Petz and G. Toth. "The Bogoliubov inner product in quantum statistics". *Lett. Math. Phys.* 27.3 (1993), pp. 205–216. DOI: `10.1007/BF00739578`.

[275]   O. Bratteli and D. W. Robinson. *Operator algebras and quantum statistical mechanics. Vol. 2: Equilibrium states. Models in quantum statistical mechanics*. Springer-Verlag Berlin Heidelberg, 1981.

[276]   M. M. Wolf, F. Verstraete, M. B. Hastings, and J. I. Cirac. "Area laws in quantum systems: mutual information and correlations". *Phys. Rev. Lett.* 100.7 (2008), p. 070502. DOI: `10.1103/PhysRevLett.100.070502`.

[277]   U. Weiss. *Quantum dissipative systems*. Vol. 13. World scientific, 2012.

[278]   A. Rivas, A. D. K. Plato, S. F. Huelga, and M. B. Plenio. "Markovian master equations: a critical study". *New J. Phys.* 12.11 (2010), p. 113032. DOI: `10.1088/1367-2630/12/11/113032`.

[279]   J. Eisert, M. Cramer, and M. B. Plenio. "Colloquium: Area laws for the entanglement entropy". *Rev. Mod. Phys.* 82.1 (2010), p. 277. DOI: `10.1103/RevModPhys.82.277`.

[280]   A. E. Allahverdyan and T. M. Nieuwenhuizen. "Extraction of Work from a Single Thermal Bath in the Quantum Regime". *Phys. Rev. Lett.* 85 (2001), p. 1799. DOI: `10.1103/PhysRevLett.85.1799`.

[281] T. M. Nieuwenhuizen and A. Allahverdyan. "Statistical thermodynamics of quantum Brownian motion: Construction of perpetuum mobile of the second kind". *Phys. Rev. E* 66.3 (2002), p. 036102. DOI: 10.1103/PhysRevE.66.036102.

[282] J. Eisert and M. B. Plenio. "Quantum and classical correlations in quantum Brownian motion". *Phys. Rev. Lett.* 89.13 (2002), p. 137902. DOI: 10.1103/PhysRevLett.89.137902.

[283] V. Ambegaokar. "Equilibration of a dissipative quantum oscillator". *Ann. Phys.* 16.5-6 (2007), pp. 319–324. DOI: 10.1002/andp.200610236.

[284] Y Subaşı, C. Fleming, J. Taylor, and B. L. Hu. "Equilibrium states of open quantum systems in the strong coupling regime". *Phys. Rev. E* 86.6 (2012), p. 061132. DOI: 10.1103/PhysRevE.86.061132.

[285] R. B. Griffiths. "Spontaneous Magnetization in Idealized Ferromagnets". *Phys. Rev.* 152.1 (1966), 240–246. DOI: 10.1103/physrev.152.240.

[286] F. J. Dyson, E. H. Lieb, and B. Simon. "Phase transitions in quantum spin systems with isotropic and nonisotropic interactions". *J. Stat. Phys.* 18.4 (1978), 335–383. DOI: 10.1007/bf01106729.

[287] T. Koma and H. Tasaki. "Symmetry Breaking and Finite Size Effects in Quantum Many-Body Systems". *J. Stat. Phys.* 76 (1994), pp. 745–803. DOI: 10.1007/BF02188685.

[288] D. Poulin. "Lieb-Robinson Bound and Locality for General Markovian Quantum Dynamics". *Phys. Rev. Lett.* 104 (2010), p. 190401. DOI: 10.1103/PhysRevLett.104.190401.

[289] T. Barthel and M. Kliesch. "Quasi-locality and efficient simulation of Markovian quantum dynamics". *Phys. Rev. Lett.* 108 (2011), p. 230504. DOI: 10.1103/PhysRevLett.108.230504.

[290] B. Nachtergaele, A. Vershynina, and V. A. Zagrebnov. "Lieb-Robinson Bounds and Existence of the Thermodynamic Limit for a Class of Irreversible Quantum Dynamics". *AMS Cont. Math.* 552 (2011), pp. 161–175. eprint: 1103.1122.

[291] J. Eisert and T. Prosen. "Noise-driven quantum criticality" (2010). arXiv: 1012.5013.

[292] M. Höning, M. Moos, and M. Fleischhauer. "Critical exponents of steady-state phase transitions in fermionic lattice models". *Phys. Rev. A* 86 (2012), p. 013606. DOI: 10.1103/PhysRevA.86.013606.

[293] K. Temme. "Lower bounds to the spectral gap of Davies generators". *J. Math. Phys.* 54 (2013), p. 122110. DOI: 10.1063/1.4850896.

[294] M. J. Kastoryano and J. Eisert. "Rapid mixing implies exponential decay of correlations". *J. Math. Phys.* 54 (2013), p. 102201. DOI: 10.1063/1.4822481.

[295] Z. Cai and T. Barthel. "Algebraic versus Exponential Decoherence in Dissipative Many-Particle Systems". *Phys. Rev. Lett.* 111 (2013). DOI: 10.1103/PhysRevLett.111.150403.

[296] M. Znidaric. "Relaxation times of dissipative many-body quantum systems". *Phys. Rev. E* 92 (2015), p. 042143. DOI: 10.1103/PhysRevE.92.042143.

[297] O. Bratteli and D. W. Robinson. *Operator Algebras and Quantum Statistical Mechanics: Volume 1: C*-and W*-Algebras. Symmetry Groups. Decomposition of States.* Springer-Verlag Berlin Heidelberg, 1987.

[298] R. Alicki. "On the detailed balance condition for non-Hamiltonian systems". *Rep. Math. Phys.* 10.2 (1976), pp. 249–258. DOI: 10.1016/0034-4877(76)90046-X.

[299] A. Kossakowski, A. Frigerio, V. Gorini, and M. Verri. "Quantum detailed balance and KMS condition". *Commun. Math. Phys.* 57.2 (1977), pp. 97–110. DOI: 10.1007/bf01625769.

[300] V Jakšić and C.-A. Pillet. "On entropy production in quantum statistical mechanics". *Comm. Math. Phys.* 217.2 (2001), pp. 285–293. DOI: 10.1007/s002200000339.

[301] F. Fagnola and V. Umanità. "Generators of Detailed Balance Quantum Markov Semigroups". *Infinite Dimensional Analysis, Quantum Probability and Related Topics* 10.03 (2007), pp. 335–363. DOI: 10.1142/s0219025707002762.

[302] F. Fagnola and V. Umanità. "Generators of KMS Symmetric Markov Semigroups on B(h): Symmetry and Quantum Detailed Balance". *Comm. Math. Phys.* 298.2 (2010), pp. 523–547. DOI: 10.1007/s00220-010-1011-1.

[303] M. J. Kastoryano and F. G.S. L. Brandao. "Quantum Gibbs samplers: the commuting case". *Comm. Math. Phys.* 344 (2016), pp. 915–957. DOI: 10.1007/s00220-016-2641-8.

[304] L. Landau, J. F. Perez, and W. F. Wreszinski. "Energy gap, clustering, and the Goldstone theorem in statistical mechanics". *J. Stat. Phys.* 26.4 (1981), pp. 755–766. DOI: 10.1007/bf01010937.

[305]   F. Strocchi. *Symmetry breaking*. Vol. 643. Springer Science & Business Media, 2005.

[306]   F. Verstraete, M. M. Wolf, and J. I. Cirac. "Quantum computation and quantum-state engineering driven by dissipation". *Nature Phys.* 5 (2009). arXiv:0803.1447, pp. 633–636. DOI: `10.1038/nphys1342`.

[307]   S. Diehl, E. Rico, M. A. Baranov, and P. Zoller. "Topology by Dissipation in Atomic Quantum Wires". *Nature Phys.* 7 (2011), p. 971. DOI: `10.1038/nphys2106`.

[308]   M. J. Kastoryano, M. M. Wolf, and J. Eisert. "Precisely Timing Dissipative Quantum Information Processing". *Phys. Rev. Lett.* 110 (11 2013), p. 110501. DOI: `10.1103/PhysRevLett.110.110501`.

[309]   H. Weimer, M. Müller, I. Lesanovsky, P. Zoller, and H. P. Büchler. "A Rydberg quantum simulator". *Nature Phys.* 6 (2010), pp. 382–388. DOI: `10.1038/nphys1614`.

[310]   K. V. Hovhannisyan, M. Perarnau-Llobet, M. Huber, and A. Acín. "Entanglement generation is not necessary for optimal work extraction". *Phys. Rev. Lett.* 111.24 (2013), p. 240401. DOI: `10.1103/PhysRevLett.111.240401`.

[311]   M. Perarnau-Llobet, K. V. Hovhannisyan, M. Huber, P. Skrzypczyk, N. Brunner, and A. Acín. "Extractable work from correlations". *Phys. Rev. X* 5.4 (2015), p. 041011. DOI: `10.1103/PhysRevX.5.041011`.

[312]   M. Huber, M. Perarnau-Llobet, K. V. Hovhannisyan, P. Skrzypczyk, C. Klöckl, N. Brunner, and A. Acín. "Thermodynamic cost of creating correlations". *New J. Phys.* 17.6 (2015), p. 065008. DOI: `10.1088/1367-2630/17/6/065008`.

[313]   D. E. Bruschi, M. Perarnau-Llobet, N. Friis, K. V. Hovhannisyan, and M. Huber. "Thermodynamics of creating correlations: Limitations and optimal protocols". *Phys. Rev. E* 91.3 (2015), p. 032118. DOI: `10.1088/1367-2630/17/6/065008`.

[314]   N. Friis, M. Huber, and M. Perarnau-Llobet. "Energetics of correlations in interacting systems". *Phys. Rev. E* 93.4 (2016), p. 042135. DOI: `10.1103/PhysRevE.93.042135`.

[315]   A. R. Brown and L. Susskind. "Second law of quantum complexity". *Phys. Rev. D* 97.8 (2018). DOI: `10.1103/physrevd.97.086015`. arXiv: `1701.01107`.

A   Back matter

A.1   Summary of the thesis

To understand in detail the relation between unitary quantum theory that describes our world at the microscopic scale and thermodynamics, which was long believed only to apply to macroscopic objects, is one of the most interesting and long-standing problems in physics. Recently, this problem has received renewed attention, in particular from the community of quantum information theory, but also from the field of statistical mechanics, inspired from stochastic thermodynamics. These results suggest that thermodynamics is also relevant for individual quantum systems, provided that they can be brought into contact with thermal baths.

In this thesis, I use recently developed tools to provide new results both on fundamental questions, but also on questions which are of practical relevance for potential miniaturized thermal machines. In terms of fundamental questions, the results in this thesis contribute to understanding how the basic laws of thermodynamics and statistical mechanics can be understood directly from unitary quantum mechanics. In particular, I discuss and answer the following questions: i) How can we quantify the third law of thermodynamics using information theoretic methods? ii) How can we quantify the "thermodynamic value" of a state-transition in quantum systems? iii) How can we axiomatically characterize the non-equilibrium free energy and relative entropy? iv) How can we justify statistical ensembles from an operational perspective, without having to introduce either some probability measures or an information theoretic entropy measure? v) How can we understand the equilibration of closed quantum systems, how long does it take and how difficult is it to avoid?

In the second part of the thesis I discuss in detail how experimental restrictions, which become important at the quantum scale, influence the ultimate thermodynamic bounds for thermal machines. In particular, the results in this thesis provide thermodynamic bounds on work-extraction and efficiencies of thermal machines in situations where 1.) an experimenter only has limited field strengths available, 2.) an experimenter cannot control the interactions between particles, but external fields arbitrarily well, and 3.) situations in which a small quantum system can only be strongly coupled to heat baths. These bounds are tight and I provide explicit examples illustrating the different behaviours.

Finally, I come back to a classic problem in statistical physics: The emergence of spontaneous symmetry breaking. Here, I provide general and rigorous new results that show how symmetry-breaking stationary states emerge from fluctuations in order parameters in dissipative lattice models.

A.2   *Zusammenfassung*

Eines der spannendsten Probleme in der Physik ist zu verstehen wie genau die Thermodynamik aus der mikroskopischen Quantentheorie hervorgeht. Dieses klassische Problem hat in den letzten Jahren erneute Aufmerksamkeit erfahren, einerseits aus Sicht der Quanteninformationstheorie, andererseits aus Sicht der statistischen Mechanik, insbesondere motiviert durch Ergebnisse der stochastischen Thermodynamik. Die Ergebnisse dieser Arbeiten deuten darauf hin, dass thermodynamische Konzepte nicht nur für makroskopische Systeme, sondern auch für ein einzelne Quantensysteme relevant sind, wenn diese in Kontakt mit Wärmebädern gebracht werden können.

In dieser Dissertation verwende ich kürzlich entwickelte Methoden, um sowohl neue Resultate in Bezug auf fundamentale Fragestellungen, als auch Resultate welche für potentielle mikroskopische thermische Maschinen relevant sind, herzuleiten. Die Resultate in Bezug auf fundamentale Fragestellungen helfen dabei zu verstehen wie Thermodynamik und statistische Mechanik aus der unitären Quantenmechanik heraus verstanden werden können. Insbesondere diskutiere (und beantworte ich) dabei die folgenden Fragen: i) Wie können wir den dritten Hauptsatz der Thermodynamik mithilfe von informationstheoretischen Methoden quantifizieren? ii) Wie lässt sich der "thermodynamische Wert" von Zustandsänderungen in Quantensystemen aus operationaler Sichtweise quantifizieren? iii) Wie können wir die freie Energie sowie die relative Entropie für Quantensysteme axiomatisch charakterisieren? iv) Wie können kanonische statistische Gesamtheiten aus operationaler Sichtweise gerechtfertigt werden, ohne Wahrscheinlichkeitsmaße oder informationstheoretische Entropien einzuführen? v) Wie können wir das Äquilibrierungsverhalten geschlossener Quantensysteme verstehen, wie lange dauert es bis ein solches System äquilibriert und wie schwierig ist es ein solches Verhalten zu verhindern?

Im zweiten Teil der Arbeit diskutiere ich im Detail welche Auswirkungen zusätzliche experimentelle Einschränkungen auf die theoretischen thermodynamischen Schranken für die Effizienz von thermischen Maschinen im Quantenregime haben. Insbesondere diskutiere ich theoretische Schranken für die Extraktion von Arbeit und den Wirkungsgrad von thermischen Maschinen in Situationen in denen 1.) nur beschränkte Feldstärken in einem Experiment zur Verfügung stehen, 2.) in denen ein_e Experimentator_in in der Lage ist externe Felder zu kontrollieren, aber nicht die Wechselwirkung zwischen einzelnen Spins und 3.) Situationen in denen ein Quantensystem nur durch eine starke Wechselwirkung in Kontakt mit einem Wärmebad gebracht werden kann. Diese neuen Schranken sind strikt und ich illustriere sie mit mehreren Beispielen.

Schließlich komme ich zurück zu einem klassischen Problem der statistischen Physik: Das Auftreten von spontaner Symmetriebrechung. Hier präsentiere ich allgemeine und rigorose Resultate, welche zeigen wie spontane Symmetriebrechung aus Fluktuationen in lokalen Ordnungsparametern in dissipativen Gittermodellen hervorgeht.

A.3 Anteil des Autors bei Konzeption, Durchführung und Verfassung der zugrundeliegenden Arbeiten

Diese Dissertation basiert auf den Publikationen [1–4, 6–12] des Verfassers, welche in Zusammenarbeit mit anderen Wissenschaftler_innen entstanden sind. Im Folgenden wird für jede dieser Publikationen der Anteil des Verfasser bei Konzeption, Durchführung und Verfassung der Publikation aufgeführt.

[1] Der Verfasser war federführend in diesem Projekt. Er hat wesentliche Teile zur Konzeption, Herleitung der Ergebnisse sowie Verfassung der Publikation beigetragen.

[2] Der Verfasser war maßgeblich an Konzeption des Projekts, der Herleitung der Hauptergebnisse, sowie der Formulierung der Publikation beteiligt. Insbesondere hat er wichtige Beiträge zum Beweis der Hauptergebnisse geleistet, sowie die Abbildungen der Publikation erstellt.

[3] Der Verfasser hat substanzielle Beiträge zur Herleitung der Ergebnisse sowie zur Formulierung der Publikation beigetragen.

[4] Der Verfasser war federführend in diesem Projekt. Er hat wesentliche Beiträge zur Konzeption, Herleitung der Hauptergebnisse und der Formulierung weiter Teile der Publikation beigetragen. Insbesondere hat er die Beweise der Hauptergebnisse ausgearbeitet.

[6] Der Verfasser hat wesentlich zur Konzeption der Arbeit beigetragen. Weiterhin hat er substanzielle Beiträge zur analytischen Herleitung der Ergebnisse für das Ising-Modell beigetragen und wesentlich an der Formulierung des technischen Anhangs mitgewirkt.

[7] Der Verfasser war federführend in diesem Projekt. Er hat wichtige Beiträge zur Konzeption des Projektes geliefert, sowie das Hauptergebnis hergeleitet und an der Formulierung der Publikation mitgewirkt.

[8] Der Verfasser war federführend in diesem Projekt und hat maßgeblich zur Konzeption, zur Herleitung der Hauptergebnisse sowie der Formulierung der Publikation beigetragen.

[9] Der Verfasser hat wichtige Beiträge zur Herleitung der Ergebnisse und der Formulierung der Publikation beigetragen. Insbesondere hat er wichtige Beiträge für die Herleitung der Störungstheorie der Korrekturterme geleistet.

[10] Der Verfasser war federführend in diesem Projekt. Seine Beiträge waren wesentlich für die Konzeption, die Herleitung der analytischen Argumente sowie der Ausformulierung der Publikation.

[11] Der Verfasser hat zur Konzeption des Projektes sowie der Formulierung der Publikation beigetragen. Weiterhin hat er Beiträge zur Herleitung und Formulierung der Hauptergebnisse geleistet, insbesondere der technischen Ausarbeitung des Beweises des Hauptergebnisses.

[12] Der Verfasser hat wichtige Beiträge zur Konzeption, Herleitung der Ergebnisse sowie der Formulierung der Publikation beigetragen. Insbesondere hat er wesentliche Teile zur Formulierung und zum Beweis von Theorem 3 (Theorem 8.4 in dieser Arbeit) beigetragen.

A.4   Eigenständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbstständig und nur mit Hilfe der angegebenen Hilfsmittel angefertigt habe. Alle Stellen der Arbeit, die wörtlich oder sinngemäß aus Veröffentlichungen oder aus anderweitigen fremden Quellen entnommen wurden, sind als solche kenntlich gemacht. Ich habe die Arbeit noch nicht in einem früheren Promotionsverfahren eingereicht.

. . . .
Henrik Wilming, 28. November 2017