Understanding Equipartition and Thermalization from Decoupling

Adrian Hutter

Master's Thesis
Institute for Theoretical Physics
ETH Zurich

Conducted at
Centre for Quantum Technologies
National University of Singapore

Supervisors:

Prof. Renato Renner
ETH Zurich

Prof. Stephanie Wehner
National University of Singapore

July 15, 2011

Abstract

The decoupling technique was originally developed for information-theoretical purposes. It describes the conditions under which the correlations in a bipartite state disappear if one part undergoes an evolution separated from the other. In the past years there has been enormous progress in understanding the foundations of statistical mechanics from first principles of quantum mechanics. By use of the decoupling technique we are able to reproduce and generalize major results of this development and to approach open problems.

As a first application of the decoupling technique we generalize the result of [Popescu *et al.*, Nat. Phys. **2**, 754-758 (2006)] about the apparent validity of the postulate of equal *a priori* probabilities to states which may be correlated to a reference. We express it in a form which allows to apply recent results about random two-qubit interactions. We give a criterion for the apparent validity of the postulate which is tight up to differences between different entropy measures. Similarly, we generalize the result of [Linden *et al.*, Phys. Rev. E **79**, 061103 (2009)] about the independence of the temporal average of a quantum mechanical system of its initial state to initial states which may be correlated to a reference.

We develop a criterion for whether a quantum-mechanical system has "forgotten" about its initial state which is tight up to differences between different entropy measures. We find that comparing two local entropies of just one particular state tells us whether generic initial states of the system have already evolved to the same state or not. After developing new bounds on the times which are necessary for entropy changes, we are able to provide lower bounds on the times which are necessary for a system to become independent of its initial state. We discover an intimate connection between a system becoming independent of its initial state and a loss of an observer's knowledge about the state of the system. As a further application of the decoupling technique we find sufficient conditions under which a system stays close to its initial state for all times, thereby extending the result of [Gogolin *et al.*, Phys. Rev. Lett. **106**, 040401 (2011)].

Acknowledgements

I thank Prof. Renato Renner for introducing me to the idea of deriving the foundations of statistical mechanics by use of the decoupling technique and for getting me in touch with Prof. Stephanie Wehner. I thank Lídia del Rio for the supervision of a semester project in which the result in Section 5.2.2 was derived. I thank Prof. Stephanie Wehner for inviting me to the National University of Singapore to write this thesis and for her instructive and friendly supervision during my four months in Singapore. I thank Mario Berta for insightful discussions about the smooth entropy formalism and correcture reading. Any remaining errors or omissions are mine.

You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage.

John von Neumann
Suggesting to Claude Shannon a name for his new uncertainty function,
as quoted in *Scientific American* Vol. 225 No. 3, (1971), p. 180

Notation

General	
log	Binary logarithm
ln	Natural logarithm
$\mathbb{C}$	Complex numbers
$\bar{c}$	Complex conjugate of $c \in \mathbb{C}$

Linear Algebra and Quantum Systems	
S, E, Ω, R	Labels for quantum systems (called system, environment, subspace and reference, respectively)
$\mathcal{H}_S, \mathcal{H}_E, \mathcal{H}_\Omega, \mathcal{H}_R$	Hilbert spaces associated with the systems S, E, Ω, R
$\mathcal{H}_S \otimes \mathcal{H}_E$	Tensor product space; Hilbert space of the joint system SE
d_A	Dimension of the Hilbert space $\mathcal{H}_A$
$\mathrm{Hom}(\mathcal{H}_A, \mathcal{H}_B)$	Homomorphisms (linear maps) $\mathcal{H}_A \to \mathcal{H}_B$
$\mathrm{End}(\mathcal{H}_A)$	$\mathrm{Hom}(\mathcal{H}_A, \mathcal{H}_A)$; endomorphisms on $\mathcal{H}_A$
$M^\dagger$	Adjoint of M
$\mathrm{Herm}(\mathcal{H}_A)$	Self-adjoint endomorphisms on $\mathcal{H}_A$; Hermitian operators on $\mathcal{H}_A$
$\lvert\psi\rangle_A, \lvert\phi\rangle_A, \lvert\nu\rangle_A$	Elements of $\mathcal{H}_A \cong \mathrm{Hom}(\mathbb{C}, \mathcal{H}_A)$
$\langle\psi\rvert_A, \langle\phi\rvert_A, \langle\nu\rvert_A$	Elements of $\mathrm{Hom}(\mathcal{H}_A, \mathbb{C})$
$\langle\psi\vert\phi\rangle$	Hilbert space scalar product between $\lvert\psi\rangle$ and $\lvert\phi\rangle$
$\mathcal{S}_=(\mathcal{H}_A)$	Set of normalized density operators on $\mathcal{H}_A$; c.f. eqn. (2.1)
$\mathcal{S}_\leq(\mathcal{H}_A)$	Set of subnormalized density operators on $\mathcal{H}_A$; c.f. eqn. (2.2)
$\rho, \sigma, \omega, \tau$	Density operators
$\mathcal{T}_{A\to B}$	Completely positive and trace-preserving linear map $\mathcal{H}_A \to \mathcal{H}_B$
$\mathbb{U}(A)$	Group of unitaries on the Hilbert space $\mathcal{H}_A$

Linear Algebra and Quantum Systems; continued

$\lambda_{\max}(\rho)$	Largest eigenvalue of the Hermitian operator ρ
$\mathbb{1}_A$	Identity operator on A
π_A	$\frac{\mathbb{1}_A}{d_A}$; fully mixed state on A
$\mathbb{1}_\Omega$	Projector onto the subspace $\mathcal{H}_\Omega \subseteq \mathcal{H}_S \otimes \mathcal{H}_E$
π_Ω	$\frac{\mathbb{1}_\Omega}{d_\Omega}$; equiprobable state on Ω
π_S^Ω	$\mathrm{Tr}_E\, \pi_\Omega$; canonical state of S with respect to Ω
$\lvert\Psi\rangle_{AA'}$	Fully entangled state between A and A'; c.f. eqn. (2.15)
$\Psi_{AA'}$, ϕ_A	Shorthand notation for the pure-state density matrices $\lvert\Psi\rangle\langle\Psi\rvert_{AA'}$, $\lvert\phi\rangle\langle\phi\rvert_A$, respectively (exclusively used with these two letters)
$J(\mathcal{T}_{A\to B})$	$\mathcal{T}_{A\to B}(\lvert\Psi\rangle\langle\Psi\rvert_{A'A})$; image of $\mathcal{T}$ under the Choi-Jamiołkowski isomorphism; c.f. eqn. (2.20)
H_{SE}	$\sum_k E_k \lvert E_k\rangle\langle E_k\rvert$; Hamilton operator of the system SE

Norms and Distance measures

$\lVert M\rVert_1$	$\mathrm{Tr}\sqrt{M^\dagger M}$; Schatten 1-norm
$\lVert\rho-\sigma\rVert_1$	Trace distance between ρ and σ
$\lVert M\rVert_2$	$\sqrt{\mathrm{Tr}(M^\dagger M)}$; Hilbert-Schmidt norm
$\lVert M\rVert_\infty$	$\sqrt{\lambda_{\max}(M^\dagger M)}$; Largest singular value of M; identical with $\lambda_{\max}(M)$ for $M \geq 0$
$F(\rho,\sigma)$	$\left\lVert\sqrt{\rho}\sqrt{\sigma}\right\rVert_1$; fidelity between ρ and σ
$P(\rho,\sigma)$	Purified distance between ρ and σ; c.f. eqn. (3.33)
$\mathcal{B}^\varepsilon(\rho)$	ε-ball; set of density operators which are ε-close (in P-distance) to ρ; c.f. eqn. (3.38)

Entropies

$H(A)_\rho$	$-\operatorname{Tr}(\rho_A \log \rho_A)$; von Neumann entropy of ρ_A
$H(A\|B)_\rho$	$H(AB)_\rho - H(B)_\rho$; conditional entropy of A given B on ρ_{AB}
$I(A:B)_\rho$	$H(A)_\rho + H(B)_\rho - H(AB)_\rho$; mutual information between A and B
$H_{\min}(A\|B)_\rho$	Conditional min-entropy; c.f. eqn. (3.10)
$H_{\max}(A\|B)_\rho$	Conditional max-entropy; c.f. eqn. (3.11)
$H^\varepsilon_{\min}(A\|B)_\rho$	ε-smooth conditional min-entropy; c.f. eqn. (3.40)
$H^\varepsilon_{\max}(A\|B)_\rho$	ε-smooth conditional max-entropy; c.f. eqn. (3.41)
$H_\alpha(A)_\rho$	$\frac{1}{1-\alpha} \log \operatorname{Tr} \rho_A^\alpha$; Rényi entropy of order α
$p(\rho_A)$	$\operatorname{Tr} \rho_A^2 \equiv 2^{-H_2(A)_\rho}$; purity of ρ_A

Contents

Chapter 1

Introduction

Consider filling a glass box with water from a freezing lake, closing it and putting in on a table. Initially you have a mixture of ice and liquid water of different temperatures in your box and the water molecules and ice blocks are moving around. After some time, the movements come to a rest and - depending on the temperature and pressure of the room in which the table stands - the ice blocks begin to melt or the water begins to freeze. Finally, the contents of the box are in a state where no further dynamics are observable.

While thermalization processes like this one are familiar to us from our everyday life, understanding them from a microscopic point of view is far from trivial. We could for instance try to comprehend how the final state, in which no further changes are observable, depends on the initial state of the contents of the box and the room. Our experience tells us that this final state only depends on some macroscopic properties of the room (like its pressure and temperature) and that basically all initial states of the room which share these macroscopic properties lead to the same final state. It does not matter, for example, if there is a chair standing on the left hand side or on the right hand side of the table. If the room is much larger than the box, the initial temperature and phase of the contents of the box should not matter either. It will matter, however, if there is an evacuated shell around the system preventing it from thermally interacting with its environment. The initial state of the room also does matter if there is a ticking bomb in it or a dog running against the table on which the box stands. So there may be a fraction of (untypical) states of the environment which have a large impact on the evolution of the system, besides imposing the macroscopic parameters. If we want to claim something like *almost all initial states of the environment of a system which share the same macroscopic parameters will eventually lead to the same state of the system* we will have to be precise about what

we wean by "almost all".

More importantly, a satisfying explanation of such thermalization processes should not find and adjust statements like the above example about independence of the environment *ad hoc* but should be able to derive them from a fundamental theory. In this thesis we are concerned with processes which happen far below the speed of light. The fundamental theory is therefore non-relativistic quantum mechanics as developed in the first half of the twentieth century.

Attempts to understand how systems of many interacting particles behave are in fact older than quantum mechanics. The central results of classical statistical physics were derived in the nineteenth century and are based on classical Hamiltonian dynamics. Classical statistical physics has produced many celebrated results and is generally very well confirmed by experiment. Trying to understand such phenomena directly from quantum mechanics may thus seem an unnecessarily ambitious goal. Even worse, the unitary time evolution of quantum mechanics seems at first sight to contradict irreversible processes like thermalization. This apparent contradiction actually led to suggestions of modifying quantum mechanics (references are given in [Gog10b]).

However, the foundations of many cornerstones of statistical physics like the postulate of equal *a priori* probabilities, ensemble averages or the Second Law of thermodynamics remained unclear in the classical framework troughout the twentieth century. In the recent years there has been an enormous progress in understanding statistical physics from first principles of quantum mechanics. Important instances are [Tas98, GM02, GLTZ06, PSW06b, LPSW09, GMM09, LPSW10, Gog10a, Sho11, GME11, RGE11]. We refer to [Gog10b] for an overview over major results of this approach.

Key is the insight that the interaction of the system in question with its environment is crucial for understanding its thermalization. If the system is isolated, the unitary time-evolution is unable to change the entropy of the system. An initial state of the system with zero entropy (a "pure" state) will then stay so forever and not be able to exhibit statistical properties. On the other hand, if the system is interacting with its environment a feature of quantum mechanics called *entanglement*, which may arise between the system and its environment, allows their joint pure state to do so as long as one focuses on the system part of this state only. The apparent contradiction between a quantum mechanical system approaching an equilibrium state and the ongoing unitary evolution can be resolved if the system is in contact with a much larger environment. The ongoing unitary evolution may then only change the state of the environment and leave the system close to its equilibrium state for most times.

In this thesis we bring the development which the above references stand for together with techniques developed in *quantum information theory* in the second half of the twentieth century and since. The most general way in which a quantum mechanical state can influence another state (of the same system at some later time or of another system) is mathematically described by a so-called *quantum channel*. If we for example ask ourselves how the state of the system under interest depends at a given time on the initial state of the environment it interacts with, this dependence is described by such a channel. Our main tool in this thesis will be the *decoupling technique* [Dup09, DBWR10] which describes the effects the application of a channel has.

Consider a bipartite quantum-mechanical state, that is, a state consisting of two parts which are classically and/or quantum mechanically correlated. We consider that we first apply a unitary to one part of the bipartite state and then input this part of the state to a quantum channel. The decoupling technique provides conditions for the channel output

a) being no longer correlated to (that is, being *decoupled* from) the other part of the bipartite state and

b) being in a definite state which does not depend on the local input state or the applied unitary but only on the channel.

This two predictions necessarily go hand in hand in the decoupling framework. By use of the decoupling technique we are able to reprodue and extend major results of the described approach [PSW06b, LPSW09, GME11]. When discussing how a system becomes independent of its initial state or approaches a canoncial state this references are solely concerned with the b) part of this predictions. We add to the quantum mechanical system S of interest and the environment E with which it interacts a third system R which we call the *reference*. The big picture is depicted in Figure 1.1. We imagine that the correlations between the reference and the initial state describe the knowledge an observer has about the initial microstate. The a) part then predicts decoupling between system and reference. We will see in several instances how the interactions between S and E destroy the correlations which S might have to an outside reference. So approaching a thermal state goes in our description necessarily hand in hand with the loss of an observer's initial information about the microstate of the system. Precisely speaking, while at $t = 0$ access to R might allow an observer to gain information about S, this is no longer the case when the system has thermalized.

As an example, let R be a thermometer which is used to measure the temperature of S before S is put in contact with E. From the point of view of an outside observer

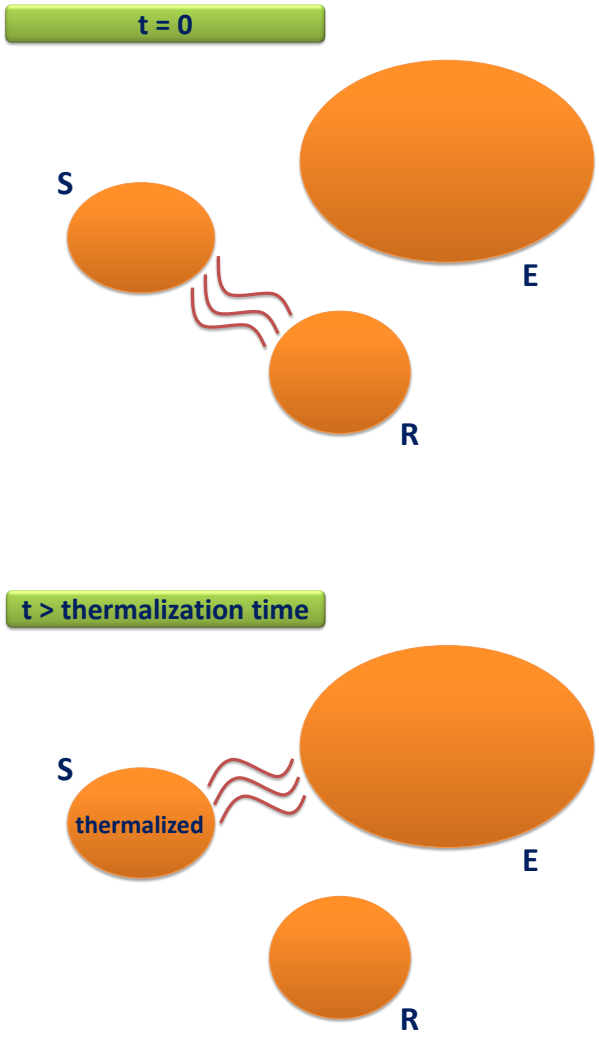

Figure 1.1: At $t = 0$ the system S is correlated to a reference R and put in contact with an environment E. The interactions between S and E then destroy the correlations between S and R, build correlations between S and E and leave S in a thermal state. What is not depicted in this figure is that due to the unitary evolution of SE the correlations between R and the joint system SE are for all times of the same strength as the initial correlations between R and S.

without access to R, this measurement leads to correlations between R and S. Having access to the thermometer after the measurement process, that is, learning the initial temperature of S, allows him or her to gain information about the microstate of S. Once the system has thermalized, access to the thermometer which has measured the initial temperature of S does no longer allow to gain any information about S.

Not only the consequences of decoupling involve the correlations between the reference and the system, these correlations themselves enter the condition for whether decoupling occurs in the first place. In cases where the reference is a quantum memory, correlations between the reference and the initial state are possible which are stronger than in cases where the reference is classical. Different initial correlation strengths can provably lead to different predictions about the decoupling behavior of the system under interest. There are cases in which a classical observer has lost all his or her initial knowledge about the microstate of S, in contrast to an observer with access to a reference which was quantum-mechanically entangled with the initial state. This aspect is somewhat similar to the results of [dRARDV11, BCCRR10]. In this references it is shown that the physics an observer with access to a quantum memory witnesses is different from what a classical observer witnesses.

1.1 Overview

Our approach requires a considerable amount of technical ingredients. The subsequent three Chapters 2, 3 and 4 are dedicated to the development of the tools and notation we apply in the thesis. In Chapter 3 we provide an introduction to the smooth entropy calculus and in Chapter 4 to the decoupling technique. With this tools at hand, strong theorems can be obtained with relative ease. While well-known facts constitute most of the contenct of these three chapters, some results like the chain rule Lemma 3.8 or the converse to the decoupling theorem in the form of Theorem 4.4 were developed specifically for this thesis.

In [PSW06b, LPSW09] it was shown how two corner stones of statistical mechanics, the postulate of equal *a priori* probabilities and thermalization, can be understood in a quantum world. In Chapters 5 and 6 we reproduce and extend these results by use of the introduced techniques. Notably we apply the results of [HL09] about random concatenations of two-qubit interactions

References like [LPSW09, LPSW10, Sho11, GME11] are exclusively concerned with long-term temporal averages and do not make statements about the state of the system at a given time. Providing bounds on the time-scales on which thermalization happens remains an open problem [LPSW10, HGJ11]. If we incorporate the temporal evolution into a quantum channel, the decoupling technique gives sufficient

and necessary conditions for the time a system needs to become independent of its initial state. These conditions depend on how fast different entropy measures can be changed. We will therefore first address this question in Chapter 7. In Chapter 8 we see that independence of the initial state of the environment is not a matter of time-scales at all. Combining the conditions obtained from the decoupling theorem and our results from Chapter 7 allows us to derive in Chapter 9 a rigorous lower bound on the time which is necessary for independence of the initial state of the system. Since initial state independence is a necessary condition for thermalization, this also lower-bounds the system's thermalization time. We discuss possible improvements of our criterion. Notably, it becomes literally tight in the case of a quantum memory suffering the influence of noise. Finally, we apply in Chapter 9 the results of Chapter 8 and prove a result somewhat converse to thermalization. We give sufficient conditions for a system staying close to its initial state of all times. Our result is very similar to the one derived in [GME11] but shows distinct advantages.

Chapter 2

Basic concepts of Quantum Mechanics

We briefly introduce the quantum mechanical concepts and techniques used in this thesis by use of the notation applied henceforth. For a more thorough introduction the interested reader is referred to [NC00].

2.1 Hilbert spaces and density operators

A quantum mechanical system A is mathematically described by a Hilbert space $\mathcal{H}_A \cong \mathbb{C}^{d_A}$. Throughout this thesis, the dimensions of all Hilbert spaces will be assumed to be finite. Elements of $\mathcal{H}_A \cong \mathrm{Hom}(\mathbb{C}, \mathcal{H}_A)$ (we identify these two spaces) are written in the form $|\psi\rangle_A$ ("ket", with small greek letters) and elements of $\mathrm{Hom}(\mathcal{H}_A, \mathbb{C})$ in the form $\langle\psi|_A$ ("bra"). Unless explicitly stated otherwise, sets like $\{|i\rangle_A\}_{i=1,\ldots,d_A}$ (with small latin letters) will always denote an *orthonormal basis* of $\mathcal{H}_A$. We require that the physical kets be normalised, i.e. $\langle\psi|\psi\rangle = 1$.

The state of a quantum mechanical system A is described by a *density matrix* (or *density operator*) ρ_A. The set of all density matrices describing states of A corresponds to the set of positive semi-definite, normalized operators on $\mathcal{H}_A$:

$$\mathcal{S}_=(\mathcal{H}_A) := \{\rho_A \in \mathrm{Herm}(\mathcal{H}_A) : \rho_A \geq 0, \mathrm{Tr}\,\rho_A = 1\} \ . \tag{2.1}$$

For technical reasons, we also define the set of *sub-normalized density matrices*:

$$\mathcal{S}_\leq(\mathcal{H}_A) := \{\rho_A \in \mathrm{Herm}(\mathcal{H}_A) : \rho_A \geq 0, \mathrm{Tr}\,\rho_A \leq 1\} \ . \tag{2.2}$$

Density matrices of rank 1 or, equivalently, projectors onto elements of $\mathcal{H}_A$ are called *pure states*. These states can be written in the form

$$\rho_A = |\psi\rangle\langle\psi|_A \tag{2.3}$$

where $|\psi\rangle_A$ is determined by ρ_A up to an irrelevant complex phase. A state ρ_A is pure if and only if $\rho_A^2 = \rho_A \Leftrightarrow \operatorname{Tr}\rho_A^2 = 1$. Since density operators are positive semi-definite, every density operator can be written as a linear combination of orthogonal projectors (*spectral decomposition*)

$$\rho_A = \sum_{i=1}^{d_A} p_i |i\rangle\langle i|_A \tag{2.4}$$

where $p_i \geq 0$ and $\sum_{i=1}^{d_A} p_i = 1$. The states $|i\rangle_A$ are the eigenstates of the density operator ρ_A. By use of normalization there is only one element in $\mathcal{S}_=(\mathcal{H}_A)$ which is a multiple of the identity. We call the corresponding state *fully mixed* and denote it by

$$\pi_A := \frac{\mathbb{1}_A}{d_A} \ . \tag{2.5}$$

We will use the shorthand notation $\phi_A \equiv |\phi\rangle\langle\phi|_A$. While other small greek letters like ρ, σ, ω or τ denote general density operators, the letter ϕ is exclusively used for pure states.

2.2 Tensor product spaces, purification and entanglement

If we are interested in a bi- (tri-, ...) partite quantum mechanical system AB (ABC, ...), its Hilbert space is described by a *tensor product space*:

$$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B \tag{2.6}$$

where for $\mathcal{H}_A = \operatorname{span}_{\mathbb{C}} \{|i\rangle_A\}_{i=1,\ldots,d_A}$ and $\mathcal{H}_B = \operatorname{span}_{\mathbb{C}} \{|j\rangle_B\}_{j=1,\ldots,d_B}$ we have

$$\mathcal{H}_A \otimes \mathcal{H}_B = \operatorname{span}_{\mathbb{C}} \{|i\rangle_A \otimes |j\rangle_B\}_{i=1,\ldots,d_A; j=1,\ldots,d_B} \ . \tag{2.7}$$

We will often omit implicit identities which appear as tensor factors on one subsystem, that is

$$U_A \rho_{AB} U_A^\dagger \equiv (U_A \otimes \mathbb{1}_B)\rho_{AB}(U_A \otimes \mathbb{1}_B)^\dagger \tag{2.8}$$

or

$$\mathcal{T}_{A\to B}(\rho_{AR}) \equiv (\mathcal{T}_{A\to B}\otimes\mathcal{I}_R)(\rho_{AR}) \tag{2.9}$$

where $\mathcal{I}_R$ denotes the identity on $\mathrm{End}\,(\mathcal{H}_R)$. If we are only interested in the state on subsystem A of a bipartite state on AB, this state can be obtained by *tracing out* subsystem B

$$\rho_A = \mathrm{Tr}_B\,\rho_{AB} = \sum_{i,j=1}^{d_A} |i\rangle\langle j|_A\,\mathrm{Tr}\,[|j\rangle\langle i|_A\rho_{AB}]\ . \tag{2.10}$$

For $\rho_{AB} = \sum_{i,j=1}^{d_A}\sum_{k,l=1}^{d_B}\rho_{AB,ijkl}|i\rangle\langle j|_A\otimes|k\rangle\langle l|_B$ we find

$$\rho_A = \sum_{i,j=1}^{d_A}\sum_{k=1}^{d_B}\rho_{AB,ijkk}|i\rangle\langle j|_A\ . \tag{2.11}$$

Every pure state $|\psi\rangle_{AB}$ can be written in the form

$$|\psi\rangle_{AB} = \sum_i \sqrt{p_i}|i\rangle_A\otimes|i\rangle_B\ . \tag{2.12}$$

This is the so-called *Schmidt decomposition* of the state $|\psi\rangle_{AB}$. The *Schmidt coefficients* $\sqrt{p_i}$ are the square roots of the eigenvalues of the marginals $\rho_A = \mathrm{Tr}_B\,|\psi\rangle\langle\psi|_{AB}$ and $\rho_B = \mathrm{Tr}_A\,|\psi\rangle\langle\psi|_{AB}$, respectively, and $\{|i\rangle_A\}$ and $\{|i\rangle_A\}$ their eigenstates. We will several times make use of the fact that the eigenvalues of the reduced states ρ_A and ρ_B "on both sides" of a pure state are identical. We will usually omit the tensor product symbol $\otimes$ and simply write $|i\rangle_A|i\rangle_B \equiv |i\rangle_A\otimes|i\rangle_B$.

If we want to emphasize that a state is not necessarily pure, we call it a *mixed state*. Every mixed state can be described as a marginal of a pure state (its *purification*), that is for every $\rho_A\in\mathcal{S}_=(\mathcal{H}_A)$ we can find a space $\mathcal{H}_P$ and a pure state $|\psi\rangle_{AP}\in\mathcal{H}_A\otimes\mathcal{H}_P$ such that

$$\rho_A = \mathrm{Tr}_P\,|\psi\rangle\langle\psi|_{AP}\ . \tag{2.13}$$

For $\rho_A = \sum_i p_i|i\rangle\langle i|_A$ the state $|\psi\rangle_{AP}$ is of the form

$$|\psi\rangle_{AP} = \sum_i \sqrt{p_i}|i\rangle_A|i\rangle_P \tag{2.14}$$

and up to the choice of the basis $\{|i\rangle_P\}_i$, that is, up to a unitary acting on $\mathcal{H}_P$, uniquely determined by ρ_A. The fact that we can know "everything" about the

state of the joint system AP (which is in a pure state) but have an objective lack of knowlege about its marginal, the state on A (which is a mixture of different states), is genuinely quantum mechanical an cannot occur in classical systems.

The key to many genuinely quantum effects is a feature of quantum states called *entanglement*. A state in $\mathcal{S}_=(\mathcal{H}_A \otimes \mathcal{H}_B)$ is called *entangled* if it cannot be written in the form $\sum_i p_i \rho_A^{(i)} \otimes \rho_B^{(i)}$. As an example, the purification of a non-pure state is always entangled between the original and the purifying system. For $\mathcal{H}_{A'} \cong \mathcal{H}_A$ we call

$$|\Psi\rangle_{AA'} := \frac{1}{\sqrt{d_A}} \sum_{i=1}^{d_A} |i\rangle_A |i\rangle_{A'} \tag{2.15}$$

the *fully entangled state* between A and A'. Similarly as for ϕ_A we use the shorthand notation $\Psi_{AA'} \equiv |\Psi\rangle\langle\Psi|_{AA'}$. Note that the partial trace of the fully entangled state is the fully mixed state,

$$\mathrm{Tr}_{A'} \Psi_{AA'} = \pi_A \ . \tag{2.16}$$

In contrast to entangled states, we call a state ρ_{AZ} on a bipartite system *classical on* Z (with respect to the basis $\{|i\rangle_Z\}$) if

$$\rho_{AZ} \in \mathrm{Herm}(\mathcal{H}_A) \otimes \mathrm{span}_{\mathbb{C}}(\{|i\rangle\langle i|_Z\}_i) \ . \tag{2.17}$$

Such states are of the form

$$\rho_{AZ} = \sum_i p_i \rho_A^{(i)} \otimes |i\rangle\langle i|_Z \ . \tag{2.18}$$

If we discard the quantum system A, we say that a state on the sysem Z is classical with respect to $\{|i\rangle_Z\}$ if it is diagonal in this basis.

2.3 CPTPM's, Choi-Jamiołkowski isomorphism and Stinespring dilation

Let $\mathcal{T}_{A\to B}$ denote the most general way in which states on system A can influence states on system B. Since the set of density operators is convex in the sense that $\sum_i p_i \rho_A^{(i)}$ is a valid density operator whenever the states $\rho_A^{(i)}$ are valid density operators, we require the map $\mathcal{T}_{A\to B}$ to preserve the convex structure of this set, that

is

$$\mathcal{T}_{A\to B}\left(\sum_i p_i\rho_A^{(i)}\right) = \sum_i p_i \mathcal{T}_{A\to B}\left(\rho_A^{(i)}\right) \ . \tag{2.19}$$

We thus require that the map $\mathcal{T}_{A\to B}$ be linear. The requirement that the images of density operators are also valid density operators corresponds to two properties of the mapping called *trace preservation* and *complete positivity*. Whilst the first property is self-explanatory, the second one requires the map to not only map positive semi-definite operators to positive semi-definite operators, but to do so even if tensored with an identity operator of any finite dimension. A linear map fulfilling complete positivity and trace preservation is simply called a *CPTPM*. CPTPM's are sometimes referred to as *quantum channels* which carry quantum states over space and time.

Since the handling of density operators is much more pleasant than the handling of CPTPM's, it is useful to have a mapping between them known as the *Choi-Jamiołkowski isomorphism* [Jam72, Cho75]. Let A' be a "copy" of the system A, i.e. $\mathcal{H}_A \cong \mathcal{H}_{A'}$. Then, the Choi-Jamiołkowski isomorphism J is defined by

$$\begin{aligned} J : \mathrm{Hom}\left(\mathrm{End}(\mathcal{H}_A), \mathrm{End}(\mathcal{H}_B)\right) &\longrightarrow \mathrm{End}\left(\mathcal{H}_{A'}\otimes\mathcal{H}_B\right) \\ \mathcal{T}_{A\to B} &\longmapsto J(\mathcal{T}) = \mathcal{T}_{A\to B}\left(\Psi_{A'A}\right) \ . \end{aligned} \tag{2.20}$$

By use of purification and the Choi-Jamiołkowski isomorphism one can prove the following powerful theorem [Sti55]:

Theorem 2.1 (Stinespring extension)**.** *Let $\mathcal{T}_{A\to B}$ be a CPTPM. Then there exists a Hilbert space $\mathcal{H}_{B'}$ and an isometry $U_{A\to BB'} \in \mathrm{Hom}(\mathcal{H}_A, \mathcal{H}_B\otimes\mathcal{H}_{B'})$ such that $\forall\rho_A \in \mathcal{S}_=(\mathcal{H}_A)$ we have*

$$\mathcal{T}_{A\to B}(\rho_A) = \mathrm{Tr}_{B'}\left(U\rho_A U^\dagger\right) \ . \tag{2.21}$$

By extending $\mathcal{H}_A$ in a way so as to make the isometry $U_{A\to BB'}$ unitary, we see that every CPTPM can be seen as a "fragment" of a unitary on a large enough space.

2.4 Hamiltonians and unitary time evolution

Elements of $\mathrm{Herm}(\mathcal{H}_A)$ are called *observables* and correspond to principally measurable physical quantities. The dynamics of a quantum mechanical system A are governed by the observable corresponding to the energy, which is known as the *Hamilton operator* of the system. We write it in the form

$$H_A = \sum_k E_k |E_k\rangle\langle E_k|_A \in \mathrm{Herm}(\mathcal{H}_A) \tag{2.22}$$

where E_k is the eigenvalue (i.e. the energy) of the energy eigenstate $|E_k\rangle_A$. We assume this operator to be time-independent. The following *von Neumann equation* is a generalization of the Schrödinger equation for mixed states. Setting $\hbar = 1$ it is given by

$$\frac{\partial \rho_A}{\partial t} = -\,\mathbb{i}\,[H_A, \rho_A] \ . \tag{2.23}$$

It is solved by

$$\rho_A(t) = U_A \rho_A(0) U_A^\dagger \tag{2.24}$$

with a unitary

$$U_A = \sum_k e^{-\,\mathbb{i}\,E_k t} |E_k\rangle\langle E_k|_A \in \mathbb{U}(\mathcal{H}_A) \ . \tag{2.25}$$

In component-wise form, we have

$$\rho_A(t) = \sum_{i,j} \rho_{A,ij}(t) |E_i\rangle\langle E_j|_A = \sum_{i,j} e^{-\,\mathbb{i}(E_i - E_j)t} \rho_{A,ij}(0) |E_i\rangle\langle E_j|_A \tag{2.26}$$

where

$$\rho_{A,ij}(t) = \langle E_i|\rho_A(t)|E_j\rangle \ . \tag{2.27}$$

Since the temporal average of the oscillating term is given by

$$\langle e^{-\,\mathbb{i}(E_i - E_j)t} \rangle_t = \lim_{t\to\infty} \frac{1}{t} \int_0^t e^{-\,\mathbb{i}(E_i - E_j)t} = \delta_{E_i, E_j} \tag{2.28}$$

we find for non-degenerate energies (i.e. $\delta_{E_i,E_j} = \delta_{i,j}$) the state

$$\langle \rho_A(t) \rangle_t := \lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau \rho_A(t) \mathrm{d}t = \sum_k \rho_{A,kk} |E_k\rangle\langle E_k| = \sum_k |E_k\rangle\langle E_k|\rho_A(0)|E_k\rangle\langle E_k| \ . \tag{2.29}$$

In the case of degenerate energy levels, the $|E_k\rangle\langle E_k|$ terms in the last expressions are to be replaced by projectors onto the subspaces spanned by all eigenstates with the same energy.

2.5 Interaction strengths

In this thesis we will consider a quantum mechanical system S which is interacting with an environment E. Let the Hamiltonian of the joint system SE be given by H_{SE}. It can be decomposed into a term acting non-trivially only on S, a term acting non-trivially only on E and an "interaction term". So,

$$H_{SE} = H_S \otimes \mathbb{1}_E + \mathbb{1}_S \otimes H_E + H_{int} \ . \tag{2.30}$$

Given H_{SE} the above decomposition is not unique. We may add a term H_0 to the decomposition which is proportional to $\mathbb{1}_{SE}$. Then, requiring that the operators H_S and H_E be traceless and that both partial traces of H_{int} vanish [LPSW10, Gog10a, Gog10b, RKIA11] makes the decomposition unique. Explicitly, we find in this case

$$H_0 := \frac{1}{d_S d_E} \operatorname{Tr}_{SE}(H_{SE}) \cdot \mathbb{1}_{SE} \ , \tag{2.31}$$

$$H_S := \frac{1}{d_E} \operatorname{Tr}_E(H_{SE}) - \frac{1}{d_S d_E} \operatorname{Tr}_{SE}(H_{SE}) \cdot \mathbb{1}_S \ , \tag{2.32}$$

$$H_E := \frac{1}{d_S} \operatorname{Tr}_S(H_{SE}) - \frac{1}{d_S d_E} \operatorname{Tr}_{SE}(H_{SE}) \cdot \mathbb{1}_E \tag{2.33}$$

and

$$\begin{aligned} H_{int} &:= H_{SE} - H_0 - H_S \otimes \mathbb{1}_E - \mathbb{1}_S \otimes H_E \\ &= H_{SE} + \operatorname{Tr}_{SE}(H_{SE}) \cdot \pi_{SE} - \operatorname{Tr}_E(H_{SE}) \otimes \pi_E - \pi_S \otimes \operatorname{Tr}_S(H_{SE}) \ . \end{aligned} \tag{2.34}$$

The operators H_0, $H_S \otimes \mathbb{1}_E$ and $\mathbb{1}_S \otimes H_E$ will be eliminated when computing certain commutators and traces. Bounds on physical quantities in S (specifically, the rate with which local entropy measures can be changed) will thus contain the term $\|H_{int}\|_\infty$ as their sole explicit dependence on H_{SE} which is then called the "interaction strength" [GHH07, Gog10a, Gog10b, RKIA11]. Such bounds will be derived in Chapter 7.

Shifting all energy levels by a certain amount is physically irrelevant. It can be seen from (2.23) or (2.26), for example, that only the differences between the energy levels and not their absolute value is relevant for the quantum mechanical evolution. H_{int} as defined in (2.34) is indeed invariant under a shift of all energy levels.

Still, an upper bound which involves the term $\|H_{int}\|_\infty$ can easily be improved without any further work. If we add a multiple of the identity to H_{int} alone it will be

eliminated in a commutator in all of the references cited. We may therefore shift all eigenvalues of H_{int} by a certain constant which is such that $\|H_{int}\|_\infty$ is minimized. We will therefore use the quantity

$$\Delta(H_{int}) := 2 \min_{\lambda \in \mathbb{R}} \|H_{int} - \lambda \mathbb{1}_{SE}\|_\infty = \lambda_{\max}(H_{int}) + \lambda_{\max}(-H_{int}) \tag{2.35}$$

to measure interaction strengths. Unlike $\|H_{int}\|_\infty$, $\Delta(H_{int})$ is now also invariant under an irrelevant shift of the eigenvalues of H_{int}. $\Delta(H_{int})$ is twice the minimal value $\|H_{int}\|_\infty$ can take when shifting all energy levels by the same amount. This allows to replace $\|H_{int}\|_\infty$ by $\frac{1}{2}\Delta(H_{int})$. Furthermore, when using $\Delta(H_{int})$ as the measure of the interaction strength the second summand in definition (2.34) becomes irrelevant. For this reason, when calculating $\Delta\left(H_{int}\right)$, we will use

$$H_{int} = H_{SE} - \operatorname{Tr}_E H_{SE} \otimes \pi_E - \pi_S \otimes \operatorname{Tr}_S H_{SE} \ . \tag{2.36}$$

An upper bound which involves $\Delta(H_{int})$ can in general be further optimized. A commutator or trace which becomes zero for $H_S \otimes \mathbb{1}_E$ or $\mathbb{1}_S \otimes H_E$ will also become so for $\mu \cdot H_S \otimes \mathbb{1}_E$ or $\nu \cdot \mathbb{1}_S \otimes H_E$, respectively, with $\mu, \nu \in \mathbb{R}$. Instead of (2.36) we may therefore define

$$\tilde{H}_{int} := H_{SE} - \mu \cdot \operatorname{Tr}_E H_{SE} \otimes \pi_E - \nu \cdot \pi_S \otimes \operatorname{Tr}_S H_{SE} \tag{2.37}$$

where $\mu, \nu \in \mathbb{R}$ are chosen such that $\Delta\left(\tilde{H}_{int}\right)$ is minimized. Simple examples show that this leads in general to better results. For instance, let $H_{SE} = |1\rangle\langle 1|_S \otimes |1\rangle\langle 1|_E$. We find

$$\|H_{int}\|_\infty = \max\left\{1 - \frac{1}{d_S} - \frac{1}{d_E} + \frac{1}{d_S d_E}, \frac{1}{d_S} - \frac{1}{d_S d_E}, \frac{1}{d_E} - \frac{1}{d_S d_E}, \frac{1}{d_S d_E}\right\} \ , \tag{2.38}$$

and

$$\Delta\left(H_{int}\right) = 1 - \frac{1}{\max\left\{d_S, d_E\right\}} \ . \tag{2.39}$$

A straightforward case-by-case analysis shows that the optimal values to define $\tilde{H}_{int}$ are given by $\mu = \frac{d_E}{2}$ and $\nu = \frac{d_S}{2}$ which yields $\Delta\left(\tilde{H}_{int}\right) = \frac{1}{2}$. At least in this example, the bounds do not only become stronger when replacing $\|H_{int}\|_\infty$ by $\frac{1}{2}\Delta\left(H_{int}\right)$ and $\Delta\left(H_{int}\right)$ by $\Delta\left(\tilde{H}_{int}\right)$, but also significantly simpler.

The relevant commutators or traces which are vanishing for $H_S \otimes \mathbb{1}_E$ or $\mathbb{1}_S \otimes H_E$ vanish for any operator of the form $P_S \otimes \mathbb{1}_E$ or $\mathbb{1}_S \otimes Q_E$, respectively, with $P_S \in$

$\mathrm{Herm}(\mathcal{H}_S)$ and $Q_E \in \mathrm{Herm}(\mathcal{H}_E)$. An upper bound involving $\Delta\left(\tilde{H}_{int}\right)$ may thus be optimized even further by use of

$$\hat{H}_{int} := H_{SE} - P_S \otimes \pi_E - \pi_S \otimes Q_E \tag{2.40}$$

where P_S and Q_E are chosen such that $\Delta\left(\hat{H}_{int}\right)$ is minimized. Finding the optimal operators $P_S \in \mathrm{Herm}(\mathcal{H}_S)$ and $Q_E \in \mathrm{Herm}(\mathcal{H}_E)$ such that

$$\min\left\{\lambda \geq 0 : -\lambda \mathbb{1}_{SE} \leq H_{SE} - P_S \otimes \pi_E - \pi_S \otimes Q_E \leq \lambda \mathbb{1}_{SE}\right\} \tag{2.41}$$

is minimized actually defines a semidefinite program.

We have no further results on this issue and leave it as an open problem how to determine $\Delta\left(\hat{H}_{int}\right)$. We do not know whether there are cases where $\hat{H}_{int}$ differs from $\tilde{H}_{int}$.

In [LPSW10] an upper bound proportional to $\|H_S \otimes \mathbb{1}_E + H_{int}\|_\infty$ is derived. Our discussion applies analogously.

Chapter 3

Entropy measures

3.1 Classical entropy measures

The Shannon entropy as introduced in [Sha48] is defined as

$$H(\{p_1, \ldots, p_n\}) := -\sum_{i=1}^{n} p_i \log p_i \tag{3.1}$$

where log denotes the binary logarithm. It describes the averaged uncertainty about the outcome of a random variable distributed according to the probability distribution $\{p_1, \ldots, p_n\}$. For the case of $n = 2$ we find the binary Shannon entropy

$$H(p) := H(\{p, 1-p\}) = -p \log(p) - (1-p) \log(1-p) \ . \tag{3.2}$$

By use of Stirling's approximation one can show that for large n

$$\log \binom{n}{pn} \approx nH(p) \ . \tag{3.3}$$

3.2 Von Neumann entropy

The *von Neumann entropy* $H(A)_\rho$ is a generalization of the Shannon entropy to the case of quantum states

$$H(A)_\rho := -\operatorname{Tr}(\rho_A \log \rho_A) \ . \tag{3.4}$$

It corresponds to the Shannon entropy of the eigenvalues of ρ_A. Its conditional version is defined by

$$H(A|B)_\rho := H(AB)_\rho - H(B)_\rho \ . \tag{3.5}$$

This quantity, roughly speaking, measures how uncertain we are on average about A if we have access to B. For a state $\rho_{AB} = \sum_i p_i \rho_A^{(i)} \otimes |i\rangle\langle i|_B$ which is classical on B we have, for example,

$$H(A|B)_\rho = \sum_i p_i H(A)_{\rho_A^{(i)}} \ . \tag{3.6}$$

The *mutual information* between systems A and B measures how strongly the two systems are correlated in a state ρ_{AB}:

$$I(A:B)_\rho := H(A)_\rho - H(A|B)_\rho = H(A)_\rho + H(B)_\rho - H(AB)_\rho \ . \tag{3.7}$$

We have

$$-\log d_A \leq H(A|B)_\rho \leq \log d_A \tag{3.8}$$

and

$$0 \leq I(A:B)_\rho \leq 2\log d_A \tag{3.9}$$

where $H(A|B)_\rho < 0$ and $I(A:B)_\rho > \log d_A$ can only be achieved if A and B are quantum-mechanically entangled. Table 3.1 captures some archetypical examples of bipartite quantum states which are used repeatedly throughout the thesis.

ρ_{AB}	$H(A\|B)$	$I(A:B)$
$\pi_A \otimes \|1\rangle\langle 1\|_B$	$\log d_A$	0
$\rho_A \otimes \rho_B$	$H(A)_\rho$	0
$\sum_i p_i \|i\rangle\langle i\|_A \otimes \|i\rangle\langle i\|_B$	0	$H(\{p_i\}_i)$
Ψ_{AB}	$-\log d_A$	$2\log d_A$

Table 3.1: Values of $H(A|B)$ and $I(A:B)$ for different states ρ_{AB}

3.3 Rényi entropies

We review multiple entropy measures which provide alternatives to the usual von Neumann entropy. $H_{\min}$ and $H_{\max}$ are introduced due to their direct physical relevance, H_R, H_0 and H_2 as auxiliary quantities.

For $\rho_{AB} \in \mathcal{S}_\leq(\mathcal{H}_{AB})$ we define the min-entropy of A conditioned on B as

$$H_{\min}(A|B)_\rho := \sup_{\sigma_B \in \mathcal{S}_=(\mathcal{H}_B)} \sup\left\{\lambda \in \mathbb{R} : 2^{-\lambda}\mathbb{1}_A \otimes \sigma_B \geq \rho_{AB}\right\} \tag{3.10}$$

and the max-entropy of A conditioned on B as

$$H_{\max}(A|B)_\rho := \sup_{\sigma_B \in \mathcal{S}_=(\mathcal{H}_B)} \log \left[F(\rho_{AB}, \mathbb{1}_A \otimes \sigma_B)\right]^2 \ . \tag{3.11}$$

For a trivial system B they simplify to $H_{\min}(A)_\rho = -\log \lambda_{\max}(\rho)$ and $H_{\max}(A)_\rho = 2 \log \operatorname{Tr} \sqrt{\rho_A}$. From [TCR09, Lemma 2] and [TCR10, Lemma 20] we have for $\rho_{AB} \in \mathcal{S}_=(\mathcal{H}_{AB})$ that

$$-\log d_{\min} \leq H_{\min}(A|B)_\rho \leq H(A|B)_\rho \leq H_{\max}(A|B)_\rho \leq \log d_A \tag{3.12}$$

where $d_{\min} := \min\{d_A, d_B\}$.

An example of the operational significance of $H_{\min}(A|B)_\rho$ is that its negative quantifies the maximal number of fully entangled bits achievable from ρ_{AB} with local operations restricted to B. $H_{\max}(A|B)_\rho$ quantifies, for instance, how random A appears (when used to generate a key, for example) from the point of view of an adversary with access to B [KRS09].

For technical reasons we introduce a quantity known as *Rényi entropy of order 2* or *quantum collision entropy*.

$$H_2(A)_\rho = -\log(\operatorname{Tr} \rho_A^2) \ . \tag{3.13}$$

Some authors introduce a quantity

$$d^{\text{eff}}(\rho_A) := \frac{1}{\operatorname{Tr} \rho_A^2} \equiv 2^{H_2(A)_\rho} \tag{3.14}$$

called *effective dimension* [PSW06b, LPSW09] and its inverse

$$p(\rho_A) := \operatorname{Tr} \rho_A^2 \equiv 2^{-H_2(A)_\rho} \tag{3.15}$$

called *purity* [GHH07, Gog10b, RKIA11]. While the purity is often easier to handle analytically than the logarithmic entropy measures, we will, however, stick to the latter in this thesis whenever possible. This is because the entropic quantities are extensive, i.e. they scale with the size of the system under consideration which allows for a better interpretation in terms of physical quantities like the number of particles in the system. Furthermore, we will generalize results of [PSW06b] and [LPSW09] to a form that makes entropy measures necessary which, unlike the effective dimension, have a conditional form.

Let $p_1 \geq p_2 \geq \ldots \geq p_{d_A} \geq 0$ denote the eigenvalues of ρ_A (thus $\sum_{i=1}^{d_A} p_i = 1$). Then the inequalities

$$\sqrt{\sum_{i=1}^{d_A} p_i^2} \geq p_1 = p_1 \cdot (\sum_{i=1}^{d_A} p_i) \geq \sum_{i=1}^{d_A} p_i^2 \tag{3.16}$$

translate into

$$\sqrt{p(\rho)} \geq \|\rho\|_\infty \geq p(\rho) \tag{3.17}$$

or

$$\frac{1}{2} H_2(A)_\rho \leq H_{\min}(A)_\rho \leq H_2(A)_\rho \ . \tag{3.18}$$

As a further auxiliary quantity we introduce the *Rényi entropy of order 0* which is simply defined as

$$H_0(A)_\rho = \log \operatorname{rank}(\rho_A) \ . \tag{3.19}$$

The last auxiliary entropy measure we introduce is

$$H_R(A)_\rho := -\sup \left\{ \lambda \in \mathbb{R} : \rho_A \geq 2^\lambda \cdot \rho_A^0 \right\} \tag{3.20}$$

where ρ_A^0 denotes the projector onto $\operatorname{supp}(\rho_A)$.

All entropy measures introduced above can be seen as special cases or limits of the *Rényi entropy of order α* which is defined as

$$H_\alpha(A)_\rho := \frac{1}{1-\alpha} \log \operatorname{Tr} \rho_A^\alpha \ . \tag{3.21}$$

The factor $\frac{1}{1-\alpha}$ is such hat the entropies coincide for every α if the distribution of the eigenvalues of ρ_A is flat, i.e. if ρ_A is proportional to a projector. While $H_{\frac{1}{2}} = H_{\max}$ we recover the von Neumann entropy H in the limit $\alpha \to 1$ and $H_{\min}$ and H_R in the limits $\alpha \to \infty$ and $\alpha \to -\infty$, respectively.

In conclusion, we have

$$\underbrace{H_{-\infty}}_{H_R} \geq \underbrace{H_0}_{\log \operatorname{rank} \rho} \geq \underbrace{H_{\frac{1}{2}}}_{H_{\max}} \geq \underbrace{H_1}_{H} \geq \underbrace{H_2}_{-\log p(\rho)} \geq \underbrace{H_\infty}_{H_{\min}} \tag{3.22}$$

where the inequalities are either trivial or due to [TCR10, Lemma 3].

3.4 Fidelity and distance measures

Severel notions of the distance or similarity of two quantum states are useful for different problems. The 1-norm of an operator $M \in \operatorname{End}(\mathcal{H})$ is

$$\|M\|_1 := \operatorname{Tr} \sqrt{M^\dagger M} \ . \tag{3.23}$$

For $M = \sum_i \lambda_i |i\rangle\langle i| \in \mathrm{Herm}(\mathcal{H})$ we have

$$\|M\|_1 = \mathrm{Tr}\,|M| = \sum_i |\lambda_i| \ . \tag{3.24}$$

The metric $\|\rho - \sigma\|_1$ induced by this norm is called *trace distance*.[1] Its relevance is basically due to the fact that the maximal probability to distinguish correctly between ρ and σ by a measurement solely depends on their trace distance and increases linearly with it [Hel69]. In fact it is given by

$$p_{\checkmark,\max} = \frac{1}{2}\left(1 + \frac{1}{2}\|\rho - \sigma\|_1\right) \ . \tag{3.25}$$

If two states have a small trace distance, *Fannes' inequality* [NC00] tells us that their von Neumann entropy is similar.

Theorem 3.1 (Fannes' inequality)**.** *Let $\rho_A, \sigma_A \in \mathcal{S}_=(\mathcal{H}_A)$ and define $\eta(x) := -x\log(x)$. Then*

$$|H(A)_\rho - H(A)_\sigma| \leq \log d_A \cdot \|\rho - \sigma\|_1 + \eta\left(\|\rho - \sigma\|_1\right) \ . \tag{3.26}$$

While the trace distance is invariant under unitaries, it is in general not so under purifications. For this reason, another distance measure is more appropriate for some purposes.

A notion of the similarity of two states is given by the *fidelity* which generalizes the Hilbert space scalar product to mixed states. For $\rho, \sigma \in \mathcal{S}_\leq(\mathcal{H})$ it is defined by

$$F(\rho, \sigma) := \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1 \ . \tag{3.27}$$

If one of the states is pure, say $\rho = |\psi\rangle\langle\psi|$, we have

$$F(|\psi\rangle\langle\psi|, \sigma) := \sqrt{\langle\phi|\sigma|\phi\rangle} \ . \tag{3.28}$$

The fidelity can only increase under CPTPM's (e.g. partial traces) [NC00], i.e.

$$F\left(\mathcal{T}(\rho), \mathcal{T}(\sigma)\right) \geq F(\rho, \sigma) \ . \tag{3.29}$$

Many important properties involving the fidelity can be derived from the following theorem [Uhl76].

[1] Note that he trace distance is often introduced with an additional factor $\frac{1}{2}$ in the literature in order to bound it by 0 and 1. We omit this factor since it would merely lead to an additional factor $\frac{1}{2}$ in many results.

Theorem 3.2 (Uhlmann's theorem)**.** *For $\rho, \sigma \in \mathcal{S}_=(\mathcal{H})$ we have*

$$F(\rho, \sigma) = \max_{|\psi\rangle, |\phi\rangle} |\langle\psi|\phi\rangle| \tag{3.30}$$

where the maximum is over all purifications $|\psi\rangle$ of ρ and $|\phi\rangle$ of σ. For a fixed purification $|\psi\rangle$ it suffices to maximize over all $|\phi\rangle$.

The fidelity and the trace distance are essentially equivalent measures of the distance/similarity of two states $\rho, \sigma \in \mathcal{S}_=(\mathcal{H})$, as shown by the *Fuchs-van de Graaf inequalities* [FvdG99]

$$1 - F(\rho, \sigma) \leq \frac{1}{2} \|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho, \sigma)^2} \ . \tag{3.31}$$

By use of the fidelity we can define a distance measure satisfying many natural conditions. We first indroduce the *generalized fidelity* for subnormalized states $\rho, \sigma \in \mathcal{S}_\leq(\mathcal{H})$

$$\bar{F}(\rho, \sigma) := F(\rho, \sigma) + \sqrt{(1 - \operatorname{Tr}\rho)\,(1 - \operatorname{Tr}\sigma)} \tag{3.32}$$

which coincides with the usual fidelity if at least one of the states is normalized. This allows us to define the *purified distance*

$$P(\rho, \sigma) = \sqrt{1 - \bar{F}(\rho, \sigma)^2} \ . \tag{3.33}$$

As its name suggests, the purified distance is the minimal trace distance between purifications of the normalized states ρ and σ for (if the trace distance is defined with a prefactor $\frac{1}{2}$). For subnormalized states $\rho, \sigma \in \mathcal{S}_\leq(\mathcal{H})$ the purified distance satisfies the following properties [TCR10]:

- It is a metric.
- It cannot increase under CPTPM's.
- It is invariant under extensions and purifications in the sense that for every extension (purification) $\bar{\rho}$ of ρ we can find an extension (purification) $\bar{\sigma}$ of σ such that $P(\rho, \sigma) = P(\bar{\rho}, \bar{\sigma})$.

We can find a statement similar to the Fuchs-van de Graaf inequalities for the purified distance and for subnormalized states.

Lemma 3.3. *For $\rho, \sigma \in \mathcal{S}_{\leq}(\mathcal{H})$ we have*

$$\frac{1}{2}\left\|\rho-\sigma\right\|_1 \leq P(\rho,\sigma) \leq \sqrt{2\left\|\rho-\sigma\right\|_1} \ . \tag{3.34}$$

If $\rho, \sigma \in \mathcal{S}_{=}(\mathcal{H})$ we have

$$\frac{1}{2}\left\|\rho-\sigma\right\|_1 \leq P(\rho,\sigma) \leq \sqrt{\left\|\rho-\sigma\right\|_1} \ . \tag{3.35}$$

Proof. Combining [TCR10, Definition 1 and Lemma 6] we have

$$\frac{1}{2}\left\|\rho-\sigma\right\|_1 + \frac{1}{2}\left|\operatorname{Tr}\rho - \operatorname{Tr}\sigma\right| \leq P\left(\rho,\sigma\right) \leq \sqrt{\left\|\rho-\sigma\right\|_1 + \left|\operatorname{Tr}\rho - \operatorname{Tr}\sigma\right|} \ . \tag{3.36}$$

The second statement then follows trivially, the first statement follows with the observation that

$$\left|\operatorname{Tr}\rho - \operatorname{Tr}\sigma\right| \leq \left\|\rho-\sigma\right\|_1 \ . \tag{3.37}$$

□

By use of the purified distance we are able to define neighbourhoods of mixed states. For $\rho \in \mathcal{S}_{\leq}(\mathcal{H})$ and $\varepsilon \geq 0$ with $\operatorname{Tr}\rho \geq \varepsilon^2$ we define an *ε-ball* in $\mathcal{S}_{\leq}(\mathcal{H})$ around ρ as

$$\mathcal{B}^{\varepsilon}(\rho) := \left\{\sigma \in \mathcal{S}_{\leq}(\mathcal{H}) : P(\rho,\sigma) \leq \varepsilon\right\} \ . \tag{3.38}$$

From the triangle inequality for P we find the following triangle inequality for the ε-balls:

$$\tau \in \mathcal{B}^{\varepsilon}(\rho) \wedge \sigma \in \mathcal{B}^{\varepsilon'}(\tau) \Rightarrow \sigma \in \mathcal{B}^{\varepsilon+\varepsilon'}(\rho) \ . \tag{3.39}$$

For more details about the purified distance and ε-balls we refer to [TCR10].

3.5 Smooth entropy measures

A problem with the conditional min- and max-entropies introduced in Section 3.3 is that they are sensitive to small variations of the state on which they are defined whereas the physical quantities we are bounding with them generally are not. Following an idea first introduced to quantum mechanics in [RW04] we will therefore

use "smooth" versions of these entropy measures.[2] Roughly speaking, the smoothing means that states which are highly untypical do not have to be taken into account.

For $\varepsilon \geq 0$ and $\rho_{AB} \in \mathcal{S}_{\leq}(\mathcal{H}_{AB})$ we define the ε-smooth min-entropy of A conditioned on B as

$$H_{\min}^{\varepsilon}(A|B)_{\rho} := \sup_{\hat{\rho}_{AB} \in \mathcal{B}^{\varepsilon}(\rho_{AB})} H_{\min}(A|B)_{\hat{\rho}} \tag{3.40}$$

and the ε-smooth max-entropy of A conditioned on B as

$$H_{\max}^{\varepsilon}(A|B)_{\rho} := \inf_{\hat{\rho}_{AB} \in \mathcal{B}^{\varepsilon}(\rho_{AB})} H_{\max}(A|B)_{\hat{\rho}} \ . \tag{3.41}$$

Since all Hilbert spaces in this thesis are finite dimensional, we can and will replace the suprema and infima by maxima and minima, respectively. In particular, we will make use of the fact that there is a state in the ε-ball which achieves the extremal value. Note that $H_{\min}^{\varepsilon}(A|B)_{\rho}$ is monotously increasing and $H_{\max}^{\varepsilon}(A|B)_{\rho}$ monotously decreasing in ε.

The relevance of smooth entropies is due to the fact that they are relevant in one-shot scenarios, where ε usually plays the role of an error probability. On the other hand, the von Neumann entropy is mainly relevant in an i.i.d. scenario. $H_{\max}^{\varepsilon}(A|B)$, for example, quantifies the entanglement cost of quantum state merging with a certain error probability [Ber09]. It also quantifies the work cost to erase system A conditioned on a memory B, except with a certain probability [dRARDV11].

Throughout the thesis we will find upper bounds for distances which contain a term like $2^{-\frac{1}{2}H_{\min}^{\varepsilon}(\Omega|R)}$ (which monotously decreases with growing ε) and a term like 12ε. Such a bound is strictly stronger than a non-smooth one, since the value of ε which optimizes (i.e. minimizes) the bound is in general different from 0.

The smooth min- and max-entropy are dual to each other in the sense that if $\rho_{ABC} \in \mathcal{S}_{\leq}(\mathcal{H}_{ABC})$ is pure we have [TCR10]

$$H_{\min}^{\varepsilon}(A|B)_{\rho} = -H_{\max}^{\varepsilon}(A|C)_{\rho} \ . \tag{3.42}$$

Furthermore, $H_{\min}^{\varepsilon}(A|B)_{\rho}$ is invariant under isometries acting on A or B, i.e. it does not depend on the Hilbert space used to represent the density operator locally. These two properties of the smooth entropy measures crucially depend on the choice of P as the relevant distance measure. The smooth entropy measures share natural

[2] Note that some of the references like [Ren05, TCR09, Dup09] use older definitions for the smoothing procedure and relations found therein may therefore not be valid with the definition used in this thesis, which is based on [TCR10].

properties with the usual von Neumann entropy like *strong subadditivity* [TCR10, footnote 7]

$$\begin{aligned} H^{\varepsilon}_{\min}(A|BC)_{\rho} &\leq H^{\varepsilon}_{\min}(A|B)_{\rho} \quad \text{and} \\ H^{\varepsilon}_{\max}(A|BC)_{\rho} &\leq H^{\varepsilon}_{\max}(A|B)_{\rho} \ . \end{aligned} \tag{3.43}$$

It can be seen from the Schmidt decomposition (2.12) that for a pure state ϕ_{AB} the entropies of the marginals on the A- and B-subsystem are identical. This observation generalizes to the case of the smooth entropy measures.

Lemma 3.4. *Let $\phi_{AB} \in \mathcal{S}_{=}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a pure state. Then,*

$$\begin{aligned} H^{\varepsilon}_{\min}(A)_{\phi} &= H^{\varepsilon}_{\min}(B)_{\phi} \quad \textit{and} \\ H^{\varepsilon}_{\max}(A)_{\phi} &= H^{\varepsilon}_{\max}(B)_{\phi} \ . \end{aligned} \tag{3.44}$$

Proof. Since $\mathrm{Tr}_A \, \phi_{AB}$ and $\mathrm{Tr}_B \, \phi_{AB}$ have the same eigenvalues, there is an isometry mapping one to the other. The statement then follows directly from the invariance of the smooth entropy measures under isometries. □

The smooth min- and the 2-entropy are essentially equivalent measures of quantum entropy in the sense that besides (3.18) we have

$$H^{\varepsilon}_{\min}(A)_{\rho} + \frac{1}{2} \log \frac{2}{\varepsilon^2} \geq H_2(A)_{\rho} \geq H_{\min}(A)_{\rho} \tag{3.45}$$

where the first inequality stems from [TCR09, Theorem 7]. While the smooth min-entropy has an operational relevance and is easy to handle by use of chain rules (c.f. Section 3.6), the collision entropy is sometimes easier to handle analytically.

While the smooth entropy measures coincide (for $\varepsilon \to 0$) with the von Neumann entropy for probability distributions which are essentially flat[3] they are strictly more general. We discover the von Neumann entropy from the smooth entropy measures in an i.i.d. (*independent and identically distributed*) scenario.

Theorem 3.5 (Fully Quantum Asymptotic Equipartition Property)**.** *[TCR09] Let $\varepsilon > 0$ and let $\rho_{AB} \in \mathcal{S}_{=}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Then,*

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} H^{\varepsilon}_{\min}(A|B)_{\rho^{\otimes n}} = H(A|B)_{\rho} \tag{3.47}$$

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} H^{\varepsilon}_{\max}(A|B)_{\rho^{\otimes n}} = H(A|B)_{\rho} \ . \tag{3.48}$$

[3] As it is the case in the first, third and fourth example in Table 3.1, while we obtain

$$H^{\varepsilon}_{\min}(A|B)_{\rho_A \otimes \rho_B} = H^{\varepsilon}_{\min}(A)_{\rho_A} \tag{3.46}$$

for a product state.

3.6 Chain rules

In order to deal with the introduced smooth entropy measures, chain rules are indispensable.

Lemma 3.6. *[DBWR10, Lemma A.6.] Let $\varepsilon > 0$, $\varepsilon', \varepsilon'' \geq 0$ and $\rho_{ABC} \in \mathcal{S}_=(\mathcal{H}_{ABC})$. Then,*

$$H_{\min}^{\varepsilon'}(A|BC)_\rho \leq H_{\min}^{\varepsilon+2\varepsilon'+\varepsilon''}(AB|C)_\rho - H_{\min}^{\varepsilon''}(B|C)_\rho + \log \frac{2}{\varepsilon^2} \ . \tag{3.49}$$

In the other direction (i.e. in order to lower-bound $H_{\min}^{\varepsilon}(A|B)_\rho$), we will use two chain rules neither of which is stronger than the other. Since we will not need it, we omit the conditioning system C. While the chain rule given in Lemma 3.7 may look more useful than the chain rule in Lemma 3.8, the huge advantage of the latter is that it allows us to express the condition for decoupling (c.f. Chapter 4) in a way which is tight up to differences between smooth min- and max-entropies and small correction terms.

Lemma 3.7. *For any $\varepsilon \geq 0$, $\rho_{AB} \in \mathcal{S}_\leq(\mathcal{H}_A \otimes \mathcal{H}_B)$ we have*

$$H_{\min}^{\varepsilon}(A|B)_\rho \geq H_{\min}^{\varepsilon}(AB)_\rho - \log d_B \ . \tag{3.50}$$

Proof. Choose $\tilde{\rho}_{AB} \in \mathcal{B}^\varepsilon(\rho)_{AB}$ such that $H_{\min}(AB)_{\tilde{\rho}} = H_{\min}^{\varepsilon}(AB)_\rho$. From [Ren05, Lemma 3.1.10.][4] we have

$$H_{\min}(A|B)_{\tilde{\rho}} \geq H_{\min}(AB)_{\tilde{\rho}} - H_0(B)_{\tilde{\rho}} \ . \tag{3.51}$$

By definition $H_{\min}^{\varepsilon}(A|B)_\rho \geq H_{\min}(A|B)_{\tilde{\rho}}$ and $H_0(B)_{\tilde{\rho}} \leq \log d_B$ and hence the assertion. □

Lemma 3.8. *Let $\varepsilon > 0$ and $\rho_{AB} \in \mathcal{S}_=(\mathcal{H}_A \otimes \mathcal{H}_B)$. Then,*

$$H_{\min}^{\varepsilon}(A|B)_\rho \geq H_{\min}^{\frac{\varepsilon}{2}}(AB)_\rho - H_{\max}^{\frac{\varepsilon}{2}}(B)_\rho - 2 \cdot \log \frac{24}{\varepsilon^2} \ . \tag{3.52}$$

Proof. Most of our proof follows a similar line of argument like the proof of [BCR09, Lemma B.12.]. Since $H_R(A)_\rho$ is the negative logarithm of the smallest non-zero eigenvalue of ρ it is obvious that $H_R(A)_\rho \geq H_0(A)_\rho$. Using (3.51) we find

$$\begin{aligned} H_{\min}(A|B)_\rho &\geq H_{\min}(AB)_\rho - H_0 \\ &\geq H_{\min}(AB)_\rho - H_R(B)_\rho \ . \end{aligned} \tag{3.53}$$

[4] Note that the quantity called $H_{\max}$ in [Ren05] differs from ours. $H_{\min}(A|B)_\rho \geq H_{\min}(AB)_\rho - H_{\max}(B)_\rho$ does not hold.

By the definition of the smooth min-entropy and (3.53) we have

$$\begin{aligned} H_{\min}^{\varepsilon}(A|B)_{\rho} &\geq \max_{\hat{\rho}_{AB}\in\mathcal{B}^{\varepsilon}(\rho_{AB})} \{H_{\min}(AB)_{\hat{\rho}} - H_R(B)_{\hat{\rho}}\} \\ &\geq \max_{\omega_{AB}\in\mathcal{B}^{\frac{\varepsilon}{2}}(\rho_{AB})} \left\{ \max_{\Pi_B} \left[H_{\min}(AB)_{\Pi_B\omega\Pi_B} - H_R(B)_{\Pi_B\omega\Pi_B} \right] \right\} . \end{aligned} \tag{3.54}$$

The maximum $\max_{\Pi_B}$ ranges over all $0 \leq \Pi_B \leq \mathbb{1}_B$ such that $\Pi_B\omega_{AB}\Pi_B \in \mathcal{B}^{\frac{\varepsilon}{2}}(\omega_{AB})$ and hence by use of the triangle inequality eqn. (3.39) $\Pi_B\omega_{AB}\Pi_B \in \mathcal{B}^{\varepsilon}(\rho_{AB})$. Using the auxiliary Lemma A.1 we find

$$H_{\min}^{\varepsilon}(A|B)_{\rho} \geq \max_{\omega_{AB}\in\mathcal{B}^{\frac{\varepsilon}{2}}(\rho_{AB})} \left\{ H_{\min}(AB)_{\omega} - \min_{\Pi_B} \left[H_R(B)_{\Pi_B\omega\Pi_B} \right] \right\} . \tag{3.55}$$

As a next step we choose $\omega_{AB} = \tilde{\omega}_{AB} \in \mathcal{B}^{\frac{\varepsilon}{2}}(\rho_{AB})$ such that $H_{\min}^{\frac{\varepsilon}{2}}(AB)_{\rho} = H_{\min}(AB)_{\tilde{\omega}}$. Hence we get

$$H_{\min}^{\varepsilon}(A|B)_{\rho} \geq H_{\min}^{\frac{\varepsilon}{2}}(AB)_{\rho} - \min_{\Pi_B} \left[H_R(B)_{\Pi_B\tilde{\omega}\Pi_B} \right] , \tag{3.56}$$

where now the maximum $\max_{\Pi_B}$ ranges over all $0 \leq \Pi_B \leq \mathbb{1}_B$ such that $\Pi_B\tilde{\omega}_{AB}\Pi_B \in \mathcal{B}^{\frac{\varepsilon}{2}}(\tilde{\omega}_{AB})$. Using Lemma A.2 we can choose $0 \leq \Pi_B \leq \mathbb{1}_B$ with $\Pi_B\tilde{\omega}_{AB}\Pi_B \in \mathcal{B}^{\frac{\varepsilon}{2}}(\tilde{\omega}_{AB})$ such that

$$H_R(B)_{\Pi_A\tilde{\omega}\Pi_A} \leq H_{\max}^{\frac{\varepsilon^2}{24}}(B)_{\tilde{\omega}} - 2 \cdot \log \frac{\varepsilon^2}{24} . \tag{3.57}$$

From this we finally obtain

$$\begin{aligned} H_{\min}^{\varepsilon}(A|B)_{\rho} &\geq H_{\min}^{\frac{\varepsilon}{2}}(AB)_{\rho} - H_{\max}^{\frac{\varepsilon^2}{24}}(B)_{\tilde{\omega}} + 2 \cdot \log \frac{\varepsilon^2}{24} \\ &\geq H_{\min}^{\frac{\varepsilon}{2}}(AB)_{\rho} - H_{\max}^{\frac{\varepsilon}{2}}(B)_{\rho} + 2 \cdot \log \frac{\varepsilon^2}{24} . \end{aligned} \tag{3.58}$$

□

3.7 Uncertainty about the initial state

Consider a system S in contact with an environment E. We do not make any assumptions about the "system" and the "environment" apart from the requirement that the Hilbert space of the joint system be a product space $\mathcal{H}_{SE} \cong \mathcal{H}_S \otimes \mathcal{H}_E$.

For example, looking at a large number of equivalent particles, S may consist of any subset of them. In particlar, the environment does not have to be a heat bath or in thermal equilibrium.

We introduce a subspace $\mathcal{H}_\Omega \subseteq \mathcal{H}_S \otimes \mathcal{H}_E$ which describes some knowledge we have about the initial state. When talking about macroscopic physical systems, we usually do not have exact knowledge about their microstate, which would mean that we simultanously have measured the state of an enormous number of constituent particles. We will usually only have knowledge about macroscopic observables like the volume, pressure, temperature or magnetization of the system. We do therefore in general not assume that we know the exact initial state of SE but only that its support is restricted to some substace of $\mathcal{H}_S \otimes \mathcal{H}_E$.

Imagine that we measure an operator $A_{SE} = \sum_i a_i |a_i\rangle\langle a_i|_{SE} \in \mathrm{Herm}(\mathcal{H}_S \otimes \mathcal{H}_E)$. Realistic measurement devices always show a finite imprecision. We can therefore only know that the measurement outcome lies in an interval $[\Sigma, \Sigma + \Delta]$. In this case, the support of the initial density matrix is restricted to

$$\mathcal{H}_\Omega := \mathrm{span}_{\mathbb{C}} \left\{ |\psi\rangle \in \mathcal{H}_S \otimes \mathcal{H}_E : A|\psi\rangle = \xi|\psi\rangle, \xi \in [\Sigma, \Sigma + \Delta] \right\} \subseteq \mathcal{H}_S \otimes \mathcal{H}_E \ . \quad (3.59)$$

If the operator A_{SE} commutes with the Hamiltonian H_{SE}, the temporal evolution will preserve the space $\mathcal{H}_\Omega$ and the support of the state under interest will be restricted to $\mathcal{H}_\Omega$ for all times. In general, however, this will not be the case.

The measurement leads to correlations between SE and the measurement device. We take in this thesis a more general point of view and assume that the state of SE is initially correlated to any system R, which we simply calle the *reference*. R does not have to be a classical measurement device but may also be a quantum memory. Let the density matrix of the joint initial state on SER be given by

$$\rho_{SER} \in \mathcal{S}_=(\mathcal{H}_\Omega \otimes \mathcal{H}_R) \ . \quad (3.60)$$

This in particular implies according to (3.12) that

$$-\log d_\Omega \leq H_{\min}^\varepsilon(SE|R)_\rho \leq \log d_\Omega \quad (3.61)$$

for $\varepsilon \to 0$.

As an example, if the observer has some classical information about the initial state, this might be described as $\rho_{SER} = \sum_i p_i \rho_{SE}^{(i)} \otimes |i\rangle\langle i|_R$, which yields a non-negative conditional min-entropy [DBWR10, Lemma A.5]

$$H_{\min}(SE|R)_\rho = -\log \left(\sum_i p_i 2^{-H_{\min}(SE)_{\rho^{(i)}}} \right) \ . \quad (3.62)$$

Now assume that we know the initial state only up to a unitary acting on $\mathcal{H}_\Omega$, so it may be every state of the form $U_\Omega \rho_{SER} U_\Omega^\dagger$ with $U_\Omega \in \mathbb{U}(\Omega)$. If we discard R for a moment, this is equivalent with saying that we know the eigenvalues of ρ_{SE} but not its eigenstates. If ρ_{SE} is pure, this means that it may be any pure state on $\mathcal{H}_\Omega$. As is apparent from eqn. (3.62), $H_{\min}^\varepsilon(SE|R)$ describes the averaged uncertainty about the initial state an observer with access to R has, which is due to the *mixture* on SE. This term is invariant under any unitary $U_{SE} \in \mathbb{U}(\mathcal{H}_S \otimes \mathcal{H}_E)$. If we now know the initial state only up to a unitary on $\mathcal{H}_\Omega$, this additional uncertainty will manifest itself in $\log d_\Omega$ terms. With the interpretation of $\mathcal{H}_\Omega$ given in (3.59), $\log d_\Omega$ is the larger, the larger the measurement uncertainty Δ is. The sum of this term and the entropic term describes therefore our total ignorance about the initial state.

$$\text{uncertainty about initial state } = \underbrace{H_{\min}^\varepsilon(SE|R)_\rho}_{\text{due to mixture}} + \underbrace{\log d_\Omega}_{\text{about eigenstates}} . \tag{3.63}$$

While $H_{\min}^\varepsilon(SE|R)_\rho$ might be negative if there are quantum mechanical correlations between the initial state and the reference, this total uncertainty about the initial state is according to eqn. (3.61) always non-negative.

The relevance of the previous paragraph is the following. We will derive statements which show that if the uncertainty about the initial state as defined above is high enough, then almost all states (in a Haar measure sense) of the form $U_\Omega \rho_{SER} U_\Omega^\dagger$ will yield the same output after the application of a certain channel. In these cases, the higher the uncertainty about the channel input (as introduced above) the lower is the variance in the channel outputs. The "uncertainty about the initial state" therefore quantifies the lack of distinguishability of different states of the form $U_\Omega \rho_{SER} U_\Omega^\dagger$ are after applying the channel.

In Chapter 5 we will adapt a different interpretation of $U_\Omega \rho_{SER} U_\Omega^\dagger$. There, we understand $\mathcal{H}_\Omega$ as describing some macroscopic constraint or conservation law. Correspondingly, we understand U_Ω as a physical evolution ρ_{SER} may undergo, which is restricted to $\mathcal{H}_\Omega$. Accordingly, we interpret $\log d_\Omega$ as the uncertainty which is due to the *evolution* U_Ω and make predictions which hold after most evolutions restricted to $\mathcal{H}_\Omega$.

Chapter 4

Decoupling

The idea of decoupling was first introduced in [HOW05] and generalized in [Dup09]. A concise introduction to the decoupling technique which makes use of the smoothing procedure introduced in the previous chapter can be found in [DBWR10]. The decoupling technique was originally developed for information-theoretical purposes like the problem of transmission of quantum data over a noisy quantum channel. In this problem we want to ensure that there are strong correlations in the final state between the input and output of a channel. The basic idea is that this can be achieved by *destroying* the correlations between this final state and a purifying environment. Further information-theoretical applications include information locking [Dup09] and state merging [ADHW06, DBWR10]. Physical applications include the study of information retrieval from an evaporating black hole (assuming that the internal dynamics of the black hole can be modelled by a Haar measure random unitary) [HP07] and evaluation of the work cost which is necessary for the erasure of a quantum mechanical system [dRARDV11].

4.1 The theorem

Consider an initial state ρ of system A which is correlated to a reference system R. We first apply a unitary U_A on system A and then a mapping $\mathcal{T}_{A\to B}$. The decoupling theorem predicts that if the averaged entropy of the initial state given access to the reference (measured by $H_{\min}^{\varepsilon}(A|R)_{\rho}$) is high enough and/or the mapping is good enough at destroying correlations (measured by $H_{\min}^{\varepsilon}(A'|B)_{\tau}$ where $\tau_{A'B} = J(\mathcal{T})$) the final state on B will be decoupled from the reference R and will be in a definite state which does not depend on the channel input (and in particular on U_A). The

higher the sum of the two terms is, the closer is the channel output on average to the decoupled state.

To get a feeling for the two entropic terms informally introduce above, consider the simple examples in Table 4.1 which are also given in [DBWR10].

Description of initial state	ρ_{AR}	$H_{\min}^{\varepsilon}(A\|R)_{\rho}$
Fully mixed on A, indep. of R	$\pi_A \otimes \rho_R$	$\log d_A$
Classically correlated, pure on A	$\sum_{i=1}^{d_A} p_i \|i\rangle\langle i\|_A \otimes \|i\rangle\langle i\|_R$	0
Fully entangled	Ψ_{AR}	$-\log d_A$
Description of mapping	$\mathcal{T}$	$H_{\min}^{\varepsilon}(A'\|B)_{\tau}$
Erasure of A	$\sigma_A \mapsto \|0\rangle\langle 0\|_A$	$\log d_A$
Orthogonal measurement on A	$\sigma_A \mapsto \sum_{k=1}^{d_A} \|k\rangle\langle k\|_A \sigma \|k\rangle\langle k\|_A$	0
Identity on A	$\sigma_A \mapsto \sigma_A$	$-\log d_A$

Table 4.1: Entropic quantities specifying the initial state ρ_{AR} and the mapping $\mathcal{T}_{A\to B}$ in the case $\varepsilon \to 0$. $\{|i\rangle_R\}_{i=1,\ldots,d_A}$ and $\{|k\rangle_A\}_{k=1,\ldots,d_A}$ are orthonormal bases whereas $\{|i\rangle_A\}_{i=1,\ldots,d_A}$ is not necessarily orthonormal.

Formally, we have the following theorem which generalizes most previous decoupling results [DBWR10].

Theorem 4.1 (Decoupling theorem)**.** *Let $\varepsilon > 0$, $\rho_{AR} \in \mathcal{S}_{=}(\mathcal{H}_{AR})$ and $\mathcal{T}_{A\to B}$ a CPTPM with Choi-Jamiołkowski representation $\tau_{A'B} = J(\mathcal{T})$. Then,*

$$\int_{\mathbb{U}(A)} ||\mathcal{T}(U\rho_{AR}U^{\dagger}) - \tau_B \otimes \rho_R||_1 \mathrm{d}U \leq 2^{-\frac{1}{2}H_{\min}^{\varepsilon}(A|R)_{\rho} - \frac{1}{2}H_{\min}^{\varepsilon}(A'|B)_{\tau}} + 12\varepsilon \tag{4.1}$$

where $\int_{\mathbb{U}(A)} \ldots \mathrm{d}U$ denotes the integral over the Haar measure on all unitaries U on A.

Note that the state on system B appearing in the theorem is the one obained from applying the mapping to a uniform input,

$$\begin{aligned} \tau_B &= \mathrm{Tr}_{A'}\, \tau_{A'B} \\ &= \mathrm{Tr}_{A'}\, \mathcal{T}_{A\to B}\left(\Psi_{A'A}\right) \\ &= \mathcal{T}_{A\to B}\left(\mathrm{Tr}_{A'}\, \Psi_{A'A}\right) \\ &= \mathcal{T}_{A\to B}\left(\pi_A\right) \ . \end{aligned} \tag{4.2}$$

As a further example for $H_{\min}^{\varepsilon}(A'|B)_{\tau}$, consider a system A consisting of $m+n$ qubits and the mapping $\mathcal{T}_{A\to B}$ which is just the partial trace over n qubits, leaving

the remaining m qubits which form system B untouched. Then $H_{\min}^{\varepsilon}(A'|B)_{\tau} = n - m$ for small ε. The more we trace out and the less we leave untouched the better we are at destroying correlations between A and its copy A' and thus at decoupling. We recover the identity and erasure in Table 4.1 as special cases. We recall that by Theorem 2.1 every channel $\mathcal{T}_{A\to B}$ can be written as a concatenation of an isometry $U_{A\to BB'}$ (which leaves entropies unchanged) and a partial trace over B'.[1] The entropy $H_{\min}^{\varepsilon}(A'|B)_{\tau}$ therefore basically measures how much is traced out when applying $\mathrm{Tr}_{B'}$. From this insight we can anticipate an important principle we will encounter several times during this thesis. When discussing dependencies of the system S under interest on its environment E we are dealing with mappings of the form $\mathcal{T}_{E\to S} = \mathrm{Tr}_{S'} \circ U_{E\to SS'}$. From the definition of an isometry we have $d_E \leq d_S d_{S'}$. So when E is considerably larger than S, the partial trace over S' and thus $\mathcal{T}_{E\to S}$ is always "good at decoupling". If we obtain decoupling the channel output does not depend on the input. Hence independence of the environment is the generic case for systems which are considerably smaller than it. *A system cannot sensitively depend on another system which is much larger than itself.* The same argument does not apply when discussing dependencies $S \to S$ (like the dependency of the temporal average of S or its state at some later time on its initial state). Hence for S becoming independent of its own initial state is a more subtle issue than becoming independent of the initial state of E. The ability of partial traces to "destroy correlations" may be seen as the fundamental reason behind this observation, which has already been made in [LPSW09].

4.2 Haar measure averages

For a non-negative random variable X with expectation value $\mathcal{E}(X)$ *Markov's inequality* tells us that

$$\Pr[X > K] < \frac{\mathcal{E}(X)}{K} \ . \tag{4.3}$$

If $\mathcal{E}(X)$ is small choosing e.g. $K = \sqrt{\mathcal{E}(X)}$ yields a high probability for a small outcome. If we average over the natural Haar measure on a high dimensional ball, a much stronger statement known as *Levy's Lemma* can be obtained.

Consider a function

$$f : \{|\psi\rangle \in \mathcal{H} : \langle\psi|\psi\rangle = 1\} \longrightarrow \mathbb{C}$$

[1] B' is here in general *not* a copy of B.

defined on a high dimensional unit sphere which has a finite Lipschitz constant. Since $|\psi\rangle$ and $e^{\mathrm{i}\,\alpha}|\psi\rangle$ yield the same pure state $|\psi\rangle\langle\psi|$, we restrict ourselves to functions $f(|\psi\rangle)$ which do not depend on the phase of their argument. We write $\langle f(|\psi\rangle)\rangle_{\psi}$ to denote the Haar measure average of f over all elements of the unit sphere. Levy's Lemma [MS01] tells us that if we pick $|\phi\rangle$ from the Haar measure on the unit sphere we have an exponentially small probability that $f(|\phi\rangle)$ is far from $\langle f(|\psi\rangle)\rangle_{\psi}$. The set of all pure states of a quantum mechanical system can be identified with the unit sphere in a Hilbert space. Levy's lemma is therefore useful to turn statements about the average of functions defined on these pure states into exponentially strong statements about the probability of obtaining an outcome which considerably differs from this average [PSW06b, LPSW09].

In this thesis we do not only consider kets but general quantum mechanical states, that is we work with the set $\mathcal{S}_{=}(\mathcal{H})$. For a function

$$f : \mathcal{S}_{=}(\mathcal{H}) \longrightarrow \mathbb{C}$$

we write

$$\int_{\mathbb{U}(\mathcal{H})} f(U\rho U^{\dagger})\mathrm{d}U$$

to denote the average over all unitaries which is taken with respect to the Haar measure on the group of unitaries $\mathbb{U}(\mathcal{H})$. Equivalently, we average f over all density operators which have the same eigenvalues as ρ and Haar distributed eigenstates. Averaging over all pure states is a special case of averaging over unitaries in the sense that

$$\langle f(|\psi\rangle\langle\psi|)\rangle_{\psi} = \int_{\mathbb{U}(\mathcal{H})} f(U|\psi\rangle\langle\psi|U^{\dagger})\mathrm{d}U \ . \tag{4.4}$$

When an application of the decoupling theorem tells us that the distance of the channel output from the decoupled state is on average small, we can deduce by Markov's inequality that the probability of obtaining an outcome above the average decreases inversely with the distance of that outcome to the average. In fact, a much stronger statement can be proven by use of the measure concentration properties of the Haar measure. Lemma A.3 gives an extension of Levy's Lemma for such Haar measure averages over unitaries.

Theorem 4.2. *In the scenario of Theorem 4.1, the probability of a violation is exponentially small in the sense that*

$$\begin{aligned}\Pr_{U}\left\{\left\|\mathcal{T}_{A\to B}(U\rho_{AR}U^{\dagger}) - \tau_B \otimes \rho_R\right\|_1 \geq 2^{-\frac{1}{2}H^{\varepsilon}_{\min}(A|R)_{\rho}-\frac{1}{2}H^{\varepsilon}_{\min}(A'|B)_{\tau}} + 12\varepsilon + \delta\right\} \\ \leq 2e^{-d_A\delta^2/16}\end{aligned} \tag{4.5}$$

where the probability is computed over the choice of U from the Haar measure on $\mathbb{U}(A)$.

For large d_A, we can obtain a small violation with very high probability by choosing $\delta = d_A^{-1/3}$. This statement holds for any channel $\mathcal{T}$ we apply the decoupling theorem to. We will not state it explicitly every time, but keep in mind that the relevant Lipschitz constant has been calculated once and for all.

Proof. It is shown in the proof of [Dup09, Theorem 3.9.] that the Lipschitz constant of the function

$$f(U) = \left\| \mathcal{T}_{A\to B}(U\rho_{AR}U^\dagger) - \tau_B \otimes \rho_R \right\|_1 \tag{4.6}$$

is upper-bounded by

$$2\max\left\{ \|\mathcal{T}(X)\|_1 : X \in \mathrm{Herm}(\mathcal{H}_A), \|X\|_1 \leq 1 \right\} \cdot \sqrt{\|\rho_A\|_\infty} \ .$$

Since $\rho_A \in \mathcal{S}_=(\mathcal{H}_A)$ we have $\sqrt{\|\rho_A\|_\infty} \leq 1$. Any $X \in \mathrm{Herm}(\mathcal{H}_A)$ can be written as $X = P_1 - P_2$ with $P_1, P_2 \in \mathrm{Herm}(\mathcal{H}_A)$, $P_1, P_2 \geq 0$. Since $\mathcal{T}$ is trace-preserving and positive (i.e. maps positive operators to positive operators)

$$\begin{aligned} \|\mathcal{T}(X)\|_1 &\leq \|\mathcal{T}(P_1)\|_1 + \|\mathcal{T}(P_2)\|_1 \\ &= \mathrm{Tr}\left[\mathcal{T}(P_1)\right] + \mathrm{Tr}\left[\mathcal{T}(P_2)\right] \\ &= \mathrm{Tr}\, P_1 + \mathrm{Tr}\, P_2 \\ &= \|X\|_1 \ , \end{aligned} \tag{4.7}$$

so

$$\max\left\{ \|\mathcal{T}(X)\|_1 : X \in \mathrm{Herm}(\mathcal{H}_A), \|X\|_1 \leq 1 \right\} \leq 1 \tag{4.8}$$

and the Lipschitz constant of f is upper-bounded by 2. Lemma A.3 tells us that

$$\Pr_U \left\{ |f(U) - \langle f \rangle_U| \geq \delta \right\} \leq 2e^{-d_A\delta^2/16} \tag{4.9}$$

so

$$\begin{aligned} \Pr_U \left\{ f(U) \geq 2^{-\frac{1}{2}H^\varepsilon_{\min}(A|R)_\rho - \frac{1}{2}H^\varepsilon_{\min}(A'|B)_\tau} + 12\varepsilon + \delta \right\} &\leq \Pr_U \left\{ f(U) \geq \langle f \rangle_U + \delta \right\} \\ &\leq \Pr_U \left\{ |f(U) - \langle f \rangle_U| \geq \delta \right\} \\ &\leq 2e^{-d_A\delta^2/16} \end{aligned} \tag{4.10}$$

where the first inequality is due to Theorem 4.1. □

4.3 Converse

For many situations it is not only useful to know under which conditions decoupling is achieved, but also under which conditions a channel will not decouple a system from another. A converse to the above theorems has been derived in [DBWR10]. We provide here a slight generalization of this theorem (the state η_A is arbitrary and not necessarily given by ρ_A). Furthermore, we provide slightly improved smoothing parameters ($\sqrt{\varepsilon}$ instead of $\sqrt{2\varepsilon}$). The proof is almost identical to the one given in [DBWR10] and reproduced here in a more explicit way for completemess and in order to make it better understandable to readers less familiar with the subject.

Theorem 4.3. *Let $\rho_{AR} \in \mathcal{S}_=(\mathcal{H}_{AR})$ and $\mathcal{T}_{A\to B}$ a CPTPM and $\tau_{A'B} = d_A\sqrt{\rho_{A'}}J(\mathcal{T})\sqrt{\rho_{A'}}$. Let $\eta_A \in \mathcal{S}_=(\mathcal{H}_A)$ be an arbitrary state. For any $\varepsilon' > 0$ and ε'', $\varepsilon''' \geq 0$, suppose that*

$$H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(A|R)_\rho + H_{\max}^{\varepsilon''}(A'B)_\tau - H_{\min}^{\varepsilon'''}(B)_{\mathcal{T}(\eta_A)} < -\log\frac{2}{\varepsilon'^2} \ . \tag{4.11}$$

Then,

$$\|\mathcal{T}(\rho_{AR}) - \mathcal{T}(\eta_A)\otimes\rho_R\|_1 > \varepsilon \ . \tag{4.12}$$

Proof. We will actually prove the contrapositive of the statement given in the theorem. We assume that $\|\mathcal{T}(\rho_{AR}) - \mathcal{T}(\eta_A)\otimes\rho_R\|_1 \leq \varepsilon$ and will show that

$$H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(A|R)_\rho \geq -H_{\max}^{\varepsilon''}(A'B)_\tau + H_{\min}^{\varepsilon'''}(B)_{\mathcal{T}(\eta_A)} - \log\frac{2}{\varepsilon'^2} \ . \tag{4.13}$$

For every CPTPM $\mathcal{T}$ we can find a Stinespring extension (Theorem 2.1), that is, we can write $\mathcal{T}_{A\to B}$ as a concatenation of an isometry $U_{A\to BB'}$ and a partial trace over B'. Let ρ_{ARP} be a purification of ρ_{AR} and define $\sigma_{BB'RP} := U_{A\to BB'}\rho_{ARP}U_{A\to BB'}^\dagger$. By use of the invariance of the smooth min-entropy under local isometries, we have

$$H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(A|R)_\rho = H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(BB'|R)_\sigma \ . \tag{4.14}$$

Applying the chain rule Lemma 3.6 gives

$$H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(A|R)_\rho \geq H_{\min}^{\varepsilon''}(B'|BR)_\sigma + H_{\min}^{\varepsilon'''+\sqrt{\varepsilon}}(B|R)_\sigma - \log\frac{2}{\varepsilon'^2} \ . \tag{4.15}$$

By use of the strong subadditivity of min-entropy (3.43), the fact that σ is pure, the duality between min- and max-entropy and Lemma 3.4 we obtain

$$\begin{aligned} H_{\min}^{\varepsilon''}(B'|BR)_\sigma &\geq H_{\min}^{\varepsilon''}(B'|BRP)_\sigma \\ &= -H_{\max}^{\varepsilon''}(B')_\sigma \\ &= -H_{\max}^{\varepsilon''}(BRP)_\sigma \ . \end{aligned} \tag{4.16}$$

Inserting this into (4.15) yields

$$H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(A|R)_\rho \geq -H_{\max}^{\varepsilon''}(BRP)_\sigma + H_{\min}^{\varepsilon'''+\sqrt{\varepsilon}}(B|R)_\sigma - \log\frac{2}{\varepsilon'^2} \ . \tag{4.17}$$

By assumption, $\mathcal{T}(\rho_{AR})$ is at most ε apart (in $\|\ldots\|_1$-distance) from a product state $\mathcal{T}(\eta_A)\otimes\rho_R$. Using Lemma 3.3 we therefore find

$$P\left(\mathcal{T}(\rho_{AR}), \mathcal{T}(\eta_A)\otimes\rho_R\right) \leq \sqrt{\varepsilon} \ . \tag{4.18}$$

We conclude that $\mathcal{T}(\eta_A)\otimes\rho_R \in \mathcal{B}^{\sqrt{\varepsilon}}\left(\mathcal{T}(\rho_{AR})\right)$. Since $\sigma_{BR} = \mathcal{T}(\rho_{AR})$ and by use of the triangle inequality (3.39) we find

$$H_{\min}^{\varepsilon'''+\sqrt{\varepsilon}}(B|R)_\sigma \geq H_{\min}^{\varepsilon'''}(B|R)_{\mathcal{T}(\eta_A)\otimes\rho_R} = H_{\min}^{\varepsilon'''}(B)_{\mathcal{T}(\eta_A)} \ . \tag{4.19}$$

Inserting this into (4.17) yields

$$H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(A|R)_\rho \geq -H_{\max}^{\varepsilon''}(BRP)_\sigma + H_{\min}^{\varepsilon'''}(B)_{\mathcal{T}(\eta_A)} - \log\frac{2}{\varepsilon'^2} \ . \tag{4.20}$$

In order to deal with the $-H_{\max}^{\varepsilon''}(BRP)_\sigma$ term, we evaluate $\sigma_{BRP} = \mathcal{T}_{A\to B}(\rho_{ARP})$. By construction, ρ_{ARP} is pure and therefore allows for a Schmidt decomposition (2.12) $\rho_{ARP} = |\nu\rangle\langle\nu|_{ARP}$ with $|\nu\rangle_{ARP} = \sum_i \sqrt{p_i}|i\rangle_A|i\rangle_{RP}$. Hence,

$$\sigma_{BRP} = \mathcal{T}(\rho_{ARP}) = \sum_{i,j}\sqrt{p_i}\sqrt{p_j}\mathcal{T}(|i\rangle\langle j|_A)\otimes|i\rangle\langle j|_{RP} \ . \tag{4.21}$$

We compare this to

$$\begin{aligned}\tau_{A'B} &= d_A\sqrt{\rho_{A'}}J(\mathcal{T})\sqrt{\rho_{A'}} \\ &= d_A\left(\sum_k\sqrt{p_k}|k\rangle\langle k|_{A'}\right)\mathcal{T}\left(\frac{1}{d_A}\sum_{i,j}|i\rangle_A|i\rangle_{A'}\langle j|_A\langle j|_{A'}\right)\left(\sum_l\sqrt{p_l}|l\rangle\langle l|_{A'}\right) \\ &= \sum_{i,j}\sqrt{p_i}\sqrt{p_j}\mathcal{T}(|i\rangle\langle j|_A)\otimes|i\rangle\langle j|_{A'} \ .\end{aligned} \tag{4.22}$$

Since σ_{BRP} and $\tau_{A'B}$ have the same eigenvalues, there is an isometry $V_{RP\to A'}$ mapping one to the other. Using the duality statement (3.42) it follows that not only the smooth min-entropy but also the smooth max-entropy is invariant under isometries, so

$$H_{\max}^{\varepsilon''}(BRP)_\sigma = H_{\max}^{\varepsilon''}(A'B)_\tau \tag{4.23}$$

which concludes the proof. □

The following theorem makes a statement about how far apart we are on average from *any* decoupled state if we first apply a randomly chosen unitary.

Theorem 4.4. *Let $\rho_{AR} \in \mathcal{S}_=(\mathcal{H}_{AR})$ and $\mathcal{T}_{A\to B}$ be a CPTPM with Choi-Jamiołkowski representation $\tau_{A'B} = J(\mathcal{T})$. For any $\varepsilon' > 0$ and ε'', $\varepsilon''' \geq 0$, suppose that*

$$H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(A|R)_\rho + H_{\max}^{\varepsilon''}(A'B)_\tau - H_{\min}^{\varepsilon'''}(B)_\tau < -\log \frac{2}{\varepsilon'^2} \ . \tag{4.24}$$

Then there is no state $\omega_B \in \mathcal{S}_=(\mathcal{H}_B)$ such that

$$\int_{\mathbb{U}(A)} \left\| \mathcal{T}(U\rho_{AR}U^\dagger) - \omega_B \otimes \rho_R \right\|_1 \mathrm{d}U \leq \frac{\varepsilon}{2} \ . \tag{4.25}$$

In particular we have

$$\int_{\mathbb{U}(A)} \left\| \mathcal{T}(U\rho_{AR}U^\dagger) - \mathcal{T}(\pi_A) \otimes \rho_R \right\|_1 \mathrm{d}U > \varepsilon \ . \tag{4.26}$$

Proof. The proof consists of two parts. First we show (4.26) the proof of which is a formalization of [DBWR10, footnote 7]. Then we show that if this is true the integral cannot be small for any state ω_B.

Apply Theorem 4.3 to a state $\tilde{\rho}_{A\tilde{R}}$ where $\mathcal{H}_{\tilde{R}} = \mathcal{H}_R \otimes \mathcal{H}_V$. Chose $\eta_A = \tilde{\rho}_A$. We think of V as being a classical register which holds the randomly chosen unitary U (the dimension of V is $|\mathbb{U}(A)|$, the cardinality of $|\mathbb{U}(A)|$, which is infinite). The input state is given by

$$\tilde{\rho}_{A\tilde{R}} = \tilde{\rho}_{ARV} := \int_{\mathbb{U}(A)} U\rho_{AR}U^\dagger \otimes |U\rangle\langle U|_V \mathrm{d}U \ . \tag{4.27}$$

Since $\tilde{\rho}_A = \int_{\mathbb{U}(A)} U\rho_A U^\dagger \mathrm{d}U = \pi_A$ we have $\tilde{\tau}_{A'B} = d_A\sqrt{\tilde{\rho}_A}J(\mathcal{T})\sqrt{\tilde{\rho}_A} = J(\mathcal{T})$. From Theorem 4.3 we have that

$$\begin{aligned}
\varepsilon &< \|\mathcal{T}(\tilde{\rho}_{A\tilde{R}}) - \mathcal{T}(\tilde{\rho}_A) \otimes \tilde{\rho}_{\tilde{R}}\|_1 \\
&= \left\| \int_{\mathbb{U}(A)} \mathcal{T}(U\rho_{AR}U^\dagger) \otimes |U\rangle\langle U|_V \mathrm{d}U - \mathcal{T}(\pi_A) \otimes \rho_R \otimes \int_{\mathbb{U}(A)} |U\rangle\langle U|_V \mathrm{d}U \right\|_1 \\
&= \left\| \int_{\mathbb{U}(A)} \left\{ \mathcal{T}(U\rho_{AR}U^\dagger) - \mathcal{T}(\pi_A) \otimes \rho_R \right\} \otimes |U\rangle\langle U|_V \mathrm{d}U \right\|_1 \\
&= \int_{\mathbb{U}(A)} \left\| \left\{ \mathcal{T}(U\rho_{AR}U^\dagger) - \mathcal{T}(\pi_A) \otimes \rho_R \right\} \otimes |U\rangle\langle U|_V \right\|_1 \mathrm{d}U \\
&= \int_{\mathbb{U}(A)} \left\| \mathcal{T}(U\rho_{AR}U^\dagger) - \mathcal{T}(\pi_A) \otimes \rho_R \right\|_1 \mathrm{d}U \ .
\end{aligned} \tag{4.28}$$

The third equality is due to the fact that all operators in the integral act on mutually orthogonal states due to the V-factor.

Now, assume by contradiction that there is a state $\omega_B \in \mathcal{S}_=(\mathcal{H}_B)$ such that

$$\int_{\mathbb{U}(A)} \left\| \mathcal{T}(U\rho_{AR}U^\dagger) - \omega_B \otimes \rho_R \right\|_1 \mathrm{d}U \leq \frac{\varepsilon}{2} \ . \tag{4.29}$$

Then, by use of the triangle inequality,

$$\begin{aligned} \frac{\varepsilon}{2} &\geq \int_{\mathbb{U}(A)} \left\| \mathcal{T}(U\rho_{AR}U^\dagger) - \mathcal{T}(\pi_A) \otimes \rho_R \right\|_1 \mathrm{d}U - \left\| \mathcal{T}(\pi_A) \otimes \rho_R - \omega_B \otimes \rho_R \right\|_1 \\ &> \varepsilon - \left\| \mathcal{T}(\pi_A) - \omega_B \right\|_1 \ . \end{aligned} \tag{4.30}$$

Furthermore, by use of the convexity of the trace distance,

$$\begin{aligned} \int_{\mathbb{U}(A)} \left\| \mathcal{T}(U\rho_{AR}U^\dagger) - \omega_B \otimes \rho_R \right\|_1 \mathrm{d}U &\geq \left\| \mathcal{T}\Big(\int_{\mathbb{U}(A)} U\rho_{AR}U^\dagger \mathrm{d}U \Big) - \omega_B \otimes \rho_R \right\|_1 \\ &= \left\| \mathcal{T}(\pi_A \otimes \rho_R) - \omega_B \otimes \rho_R \right\|_1 \\ &= \left\| \mathcal{T}(\pi_A) - \omega_B \right\|_1 \ . \end{aligned} \tag{4.31}$$

The first equality is due to Lemma A.4. Combining inequalities (4.30) and (4.31) yields

$$\int_{\mathbb{U}(A)} \left\| \mathcal{T}(U\rho_{AR}U^\dagger) - \omega_B \otimes \rho_R \right\|_1 \mathrm{d}U > \frac{\varepsilon}{2} \tag{4.32}$$

in contradiction to (4.29). □

Since we are usually interested in the limit of small, positive epsilons, it may at first sight seem disturbing that the term on the right hand side of (4.11) and (4.24) diverges in this limit. Keep in mind, however, that the divergence is only logarithmic and that the epsilons do not depend on the size of the systems (but usually refer to error probabilities). The entropic terms, on the other hand, grow proportionally with the size of the system. In the thermodynamic limit the logarithmic divergence is therefore negligible.

Both converse theorems may in principle be applied to any mapping $\mathcal{T}$ investigated henceforth. However, we will only do this explicitly if the application leads to non-trivial new insights and if the emerging terms are analytically manageable.

While Theorem 4.3 makes a statement about one particular input state, Theorem 4.4 averages over possible input states. A major advantage of Theorem 4.4 is that

$\tau_{A'B}$ is now given by the Choi-Jamiołkowski isomorphism $J(\mathcal{T})$ of the channel $\mathcal{T}_{A\to B}$ and not any more by the difficult-to-handle state $d_A\sqrt{\rho_{A'}}J(\mathcal{T})\sqrt{\rho_{A'}}$ as in Theorem 4.3.

Applying the chain rule Lemma (3.8) to the decoupling theorem expressed in (4.1) we obtain

$$\begin{aligned}\int_{\mathbb{U}(A)} ||\mathcal{T}(U\rho_{AR}U^\dagger) - \tau_B \otimes \rho_R||_1 \mathrm{d}U \\ \leq 2^{-\frac{1}{2}H^\varepsilon_{\min}(A|R)_\rho - \frac{1}{2}H^{\frac{\varepsilon}{2}}_{\min}(A'B)_\tau + \frac{1}{2}H^{\frac{\varepsilon}{2}}_{\max}(B)_\tau + \log\frac{24}{\varepsilon^2}} + 12\varepsilon \ .\end{aligned} \tag{4.33}$$

For small ε we can therefore express the condition for decoupling slightly informally as

$$H^\varepsilon_{\min}(A|R)_\rho + H^\varepsilon_{\min}(A'B)_\tau - H^\varepsilon_{\max}(B)_\tau \gtrapprox 0 \ . \tag{4.34}$$

If this is fulfilled, we have by Theorem 4.2 an exponentially small probability that an input state $U_A\rho_{AR}U_A^\dagger$ with U_A drawn from the Haar measure yields a channel output which is further away from the decoupled state than a certain distance. This distance is exponentially small in the l.h.s. of (4.34).

From (4.24) we obtain for small ε the condition

$$H^\varepsilon_{\min}(A|R)_\rho + H^\varepsilon_{\max}(A'B)_\tau - H^\varepsilon_{\min}(B)_\tau \lessapprox 0 \tag{4.35}$$

for an input state $U_A\rho_{AB}U_A^\dagger$ on average yielding an output state which is further away from decoupling than some finite quantity which can be found from the precise form of the condition.

Comparing these two conditions, we see that the condition for decoupling is tight up to differences between smooth min- and max-entropies and correction terms of order $\mathcal{O}(\log\frac{1}{\varepsilon})$. We will call this "essentially tight" but keep in mind that the differences between smooth min- and max-entropies may be arbitrarily large. The decoupling criterion *is* tight whenever the different entropy measures coincide, that is, if the relevant density operators are essentially flat, i.e. proportional to projectors. In Chapter 9 the difference between the two entropy measures will manifest itself in a *time interval* for which we do not know whether a quantum mechanical system is already independent of its initial state or not.

Chapter 5

The postulate of equal a priori probability

In this chapter we will try to understand a cornerstone of classical statistical physics, the postulate of equal *a priori* probabilities, from the underlying quantum mechanics. We will make a statement about how typical states satisfying a certain global constraint look like locally and will not be concerned with states evolving under a particular Hamiltonian. In the following chapter we will investigate what we can tell about temporal averages under the evolution governed by a given Hamiltonian. After an intermediate chapter about the rates with which entropies can be changed, we will make statements about the state of the system after evolving for a given time under a given Hamiltonian. That is, our predictions will become more and more specific throughout the subsequent chapters and, correspondingly, the problems more and more difficult. In the present and the following chapter we mainly extend existing results, in the chapter about the state at a given time we take a shot at unsolved problems.

5.1 Motivation

Standard approaches to statistical physics are based on the postulate of equal *a priori* probabilities, which is used to justify the application of the microcanonical and the canonical ensemble. One assumes that in equlibrium all states of the system (described as points in phase space) satisfying a certain constraint (usually energy conservation) have the same *a priori* probability.[1] In the density operator formalism

[1] As an example, c.f. [Kit58, p. 12f]:

used in quantum mechanics, this is equivalent to postulating that the density operator is fully mixed on the subspace of the total Hilbert space of the system which describes the constraint. This means that one describes the joint system by the state which achieves the maximal entropy compatible with the constraint, which is known as *Jaynes' Principle* [Jay57].

Crucially, the postulate describes *equilibrium* states. Specific initial states might obviously violate it. We therefore need the assumption that all initial states evolve into states satisfying the postulate. In classical statistical physics, this evolution takes place in phase space and is described by Hamilton's equations. In quantum mechanics, it is described by a unitary acting on the states on a Hilbert space.

Consider a system in contact with a (usually much larger) environment. If we are only interested in the details of the system, then its state can be obtained by *tracing out* the environment. According to the postulate, we will therefore in equilibrium trace out the environment of a state which is fully mixed on a subspace of the joint Hilbert space. The state obtained this way will be called the *canonical state* of the system (corresponding to the constraint described by the subspace). Evaluating this canonical state is a standard task in statistical physics. If the environment is described as a heat bath (a reservoir with a fixed temperature), it will often turn out to be of *Boltzmannian* form. In [Tas98, GM02, GLTZ06] the Boltzmann form of the canonical state is derived under particular assumptions on the Hamiltonian.

The goal of this chapter is to justify the application of the postulate of equal *a priori* probability from the principles of quantum mechanics. This will be done by showing that all relevant (for statistical physics) consequences of the postulate can be derived by use of the decoupling theorem. More precisely, consider an initial state of the system and its environment, about which an observer has some knowledge. We show that almost every evolution compatible with a global constraint will turn the system and its environment into the canonical state of the system, about which the observer will have lost all his or her initial information.

Note that we will *not* prove that most evolutions lead system and environment to the equiprobable state (as suggested by the postulate of equal *a priori* probability) but only that after most evolutions the system will be in the same state *as if* system and environment were in the equiprobable state. This explains why the postulate

---

If the energy of the system is prescribed to be in the range δE at E_0, we may, according to the preceding section, form a satisfactory ensemble by taking the density as equal to zero except in the selected narrow range δE at E_0. We specify the ensemble by

$$P(E) = \begin{cases} \text{Constant (for energy in } \delta E \text{ at } E_0) \\ 0 \text{ (outside this range).} \end{cases} \tag{5.1}$$

leads to correct results despite possibly not being fulfilled in physical systems. Quoting [PSW06b][2] we therefore suggest that the main postulate of statistical mechanics, the equal *a priori* probability postulate, should be abandoned as missleading and unnecessary.

The temporal evolution will in this chapter not be governed by a Hamiltonian. Rather we consider all evolutions compatible with a macroscopic constraint or conservation law. Accordingly there will neither be a notion of time in this chapter. Given an initial state we will after a brief enough time interval still be close to it. Our result and the postulate therefore are statements about the state of the system after having let the evolution enough time. For a discussion of what "enough time" is, we refer to Chapter 9.

We will proof two versions of the main theorem of this chapter which are based on the chain rules Lemma 3.7 and Lemma 3.8, respectively. The first version allows for an easier understanding in terms of entropic quantities and to directly reproduce the main results of [PSW06a, PSW06b]. The second version comes up with more scary-looking epsilon-terms, but allows us to give a condition for the fulfillment of the postulate which is tight up to differences between smooth min- and max-entropies and small correction terms.

5.2 Generalization of previous results

5.2.1 Informal version

If we formally describe the global constraint which is imposed on the joint system SE by $\mathcal{H}_\Omega \subseteq \mathcal{H}_S \otimes \mathcal{H}_E$ the postulate of equal *a priori* probabilities predicts the joint system to be in the state π_Ω. The system S is then in the state

$$\pi_S^\Omega := \mathrm{Tr}_E \, \pi_\Omega \tag{5.2}$$

which we call the *canonical state* of S with respect to Ω.

Let an initial state ρ_{SER} evolve under some constraint Ω. Our goal is to characterize the evolved state $U_\Omega \rho_{SER} U_\Omega^\dagger$ and to investigate how the correlations between the reference and and the system (describing an observer's knowledge about the system) develop.

[2] "In effect, we propose to replace the postulate of equal *a priori* probabilities by the principle of apparently equal *a priori* probabilities, which states that as far as the system is concerned almost every state of the universe seems similar to the average."

The first version of the theorem can be summarized as follows. If

$$H_{\min}^{\varepsilon}(SE|R)_{\rho} + \log d_{\Omega} \gtrapprox 2 \log d_S \tag{5.3}$$

for $\varepsilon \to 0$, the evolved state $U_{\Omega}\rho_{SER}U_{\Omega}^{\dagger}$ will very likely be characterized by the following two properties:

1. The system will be decoupled from the reference: $I(S:R)_{U\rho U^{\dagger}} \approx 0$.
2. The system will be in the canonical state: $\mathrm{Tr}_{ER}\left(U\rho U^{\dagger}\right) \approx \pi_S^{\Omega}$.

The probability for this not to happen is exponentially small. As discussed in Section 3.7, condition (5.3) ensures that there is "enough uncertainty" about the evolved state and quantifies the two sources of this uncertainty. Those are the averaged uncertainty about the initial state an observer with access to the reference has, $H_{\min}^{\varepsilon}(SE|R)_{\rho}$, and the uncertainty which is due to the evolution, $\log d_{\Omega}$.

5.2.2 Formal version

Formally we prove the following theorem.

Theorem 5.1. *Let $\mathcal{H}_{\Omega} \subseteq \mathcal{H}_S \otimes \mathcal{H}_E$ and let $\rho_{SER} \in \mathcal{S}_{=}(\mathcal{H}_{\Omega} \otimes \mathcal{H}_R)$. Let $\varepsilon \geq 0$. Then,*

$$\begin{aligned}\int_{\mathbb{U}(\Omega)} &|| \mathrm{Tr}_E(U_{\Omega}\rho_{SER}U_{\Omega}^{\dagger}) - \pi_S^{\Omega} \otimes \rho_R||_1 \mathrm{d}U \\ &\leq \sqrt{d_S} \cdot 2^{-\frac{1}{2}H_{\min}^{\varepsilon}(E)_{\pi_{\Omega}} - \frac{1}{2}H_{\min}^{\varepsilon}(SE|R)_{\rho}} + 12\varepsilon \\ &\leq \frac{d_S}{\sqrt{d_{\Omega}}} \cdot 2^{-\frac{1}{2}H_{\min}^{\varepsilon}(SE|R)_{\rho}} + 12\varepsilon \ . \end{aligned} \tag{5.4}$$

Proof. Theorem 4.1 with $A = \Omega$, $\mathcal{H}_{\Omega} \subseteq \mathcal{H}_S \otimes \mathcal{H}_E$, $\mathcal{H}_{\Omega} \cong \mathcal{H}_{\Omega'}$, $\rho_{AR} = \rho_{\Omega R}$, $B = S$ and $\mathcal{T}_{A\to B} = \mathrm{Tr}_E$[3] gives

$$\int_{\mathbb{U}(\Omega)} || \mathrm{Tr}_E(U\rho_{\Omega R}U^{\dagger}) - \tau_S \otimes \rho_R||_1 \mathrm{d}U \leq 2^{-\frac{1}{2}H_{\min}^{\varepsilon}(\Omega'|S)_{\tau} - \frac{1}{2}H_{\min}^{\varepsilon}(\Omega|R)_{\rho_{\Omega R}}} + 12\varepsilon \ . \tag{5.5}$$

From eqn. (4.2) we have

$$\tau_S = \mathrm{Tr}_E \, \pi_{\Omega} = \pi_S^{\Omega} \ . \tag{5.6}$$

[3]Note that $\mathrm{Tr}_E : \mathcal{S}_{=}(\mathcal{H}_S \otimes \mathcal{H}_E) \supseteq \mathcal{S}_{=}(\mathcal{H}_{\Omega}) \to \mathcal{S}_{=}(\mathcal{H}_S)$ is a CPTPM.

Since $\mathcal{S}_=(\mathcal{H}_\Omega) \subseteq \mathcal{S}_=(\mathcal{H}_S \otimes \mathcal{H}_E)$ we can write ρ_{SER} instead of $\rho_{\Omega R}$. This yields the left hand side of eqn. (5.4). It remains to show that $2^{-\frac{1}{2}H^\varepsilon_{\min}(\Omega'|S)_\tau} \leq \sqrt{d_S} \cdot 2^{-\frac{1}{2}H^\varepsilon_{\min}(E)_{\pi_\Omega}}$. Using the chain rule Lemma 3.7 we have

$$H^\varepsilon_{\min}(\Omega'|S)_\tau \geq H^\varepsilon_{\min}(S\Omega')_\tau - \log d_S \ . \tag{5.7}$$

Since $\tau_{S\Omega'} = \operatorname{Tr}_E \Psi_{\Omega\Omega'}$ and $\Psi_{\Omega\Omega'} \in \mathcal{S}_=(\mathcal{H}_S \otimes \mathcal{H}_E \otimes \mathcal{H}_{S'} \otimes \mathcal{H}_{E'})$ is pure, we have according to Lemma 3.4

$$\begin{aligned} H^\varepsilon_{\min}(S\Omega')_\tau &= H^\varepsilon_{\min}(S\Omega')_{\Psi_{\Omega\Omega'}} \\ &= H^\varepsilon_{\min}(E)_{\Psi_{\Omega\Omega'}} \\ &= H^\varepsilon_{\min}(E)_{\pi_\Omega} \ . \end{aligned} \tag{5.8}$$

Combinig this with (5.7) yields $H^\varepsilon_{\min}(\Omega'|S)_\tau \geq H^\varepsilon_{\min}(E)_{\pi_\Omega} - \log d_S$ which proves the stronger bound.

In order to obtain the weaker bound, we apply the strong subadditivity of the min-entropy (3.43) and again the chain rule to get

$$\begin{aligned} H^\varepsilon_{\min}(E)_{\pi_\Omega} &\geq H_{\min}(E)_{\pi_\Omega} \\ &\geq H_{\min}(E|S)_{\pi_\Omega} \\ &\geq H_{\min}(SE)_{\pi_\Omega} - \log d_S \\ &= \log d_\Omega - \log d_S \ . \end{aligned} \tag{5.9}$$

□

5.2.3 Discussion

Theorem 5.1 generalizes the Fully Quantum Slepian-Wolf theorem [ADHW06], which corresponds to the special case of $\mathcal{H}_\Omega = \mathcal{H}_S \otimes \mathcal{H}_E$. With the interpretation of $\mathcal{H}_\Omega$ given in eqn. (3.59) and taking the operator to be the Hamiltonian, this special case of our result corresponds to the fully degenerate case.

Consider the special case where the system is initially in a pure state $\rho_{SER} = \phi_{SE} \otimes |r\rangle\langle r|_R$, so $H^\varepsilon_{\min}(SE|R)_\rho = 0$ and the reference might be discarded. In this case, equation (5.4) turned into an exponentially strong statement by Theorem 4.2 simplifies to the main result of [PSW06b], equations (2) and (3). The similarity with these equations is striking. Note that these almost identical results were achieved using totally different approaches. While the derivation in [PSW06a] is technically challenging, ours is based on the decoupling theorem and a few key properties like strong subadditivity and a chain rule for smooth entropy measures.

What makes our result more general than the result in [PSW06b] are the following three points: First, we do not have the restriction that the initial state of SE has to be pure. Second, we include the observer's knowledge about the initial state and allow for the possibilities of incomplete knowledge or entanglement. Third, averaging over unitaries instead of averaging over pure states allows for a more physical interpretation of the averaging process. We will comment in more detail about this point in Section 5.4. As discussed in Section 4.2 averaging over all unitaries acting on a certain space (as we do) is strictly more general than averaging over all pure states from a certain space (as in [PSW06b]) in that it can also be applied to mixed states.

As for the first point, there are two possible interpretations of mixed states which both show the relevance of extending the result to mixed initial states. Every mixed state may be seen as the marginal of a pure state. We may think of the "system" and the "environment" as first being in contact with a further environment which we call the "laboratory" L and imagine that the joint system SEL is initially in a pure state. The interaction between SE and L will lead to quantum mechanical correlations, i.e. entanglement between them. We then isolate SE from L. While SEL is still in a pure state, the state of SE and its evolution has to be described by a mixed state. We can also understand the mixture of the initial state as in classical information theory, namely as simply describing subjective ignorance. Since every real measurement process (even a classical one) has only finite precision there is always a positive amount of entropy of this kind.

Note that the result of [PSW06b] cannot straightforwardly be generalized to mixed states by introducing a purifying system P with $\mathcal{H}_P \cong \mathcal{H}_\Omega$ and formally including it into E, a method we will use later in this thesis where it is appropriate.

The condition $H_{\min}^{\varepsilon}(SE|R)_\rho + \log d_\Omega \gtrapprox 2 \log d_S$ tells us that we need a high d_Ω to compensate for a low $H_{\min}^{\varepsilon}(SE|R)_\rho$ and vice versa. In other words, we do not predict decoupling if our knowledge about the initial state is too precise and if the global constraint is too severe. If the reference R is classical, we always have $H_{\min}^{\varepsilon}(SE|R)_\rho \geq 0$ and therefore $\log d_\Omega \gtrapprox 2 \log d_S$ is always sufficient for decoupling. However, if R is a quantum memory which might initally be entangled with SE, we might have $H_{\min}^{\varepsilon}(SE|R)_\rho < 0$.

5.2.4 Example: Weakly interactive spins

We evaluate the stronger bound in (5.4) for one of the simplest Hamiltonians possible. This allows us to extend the example exhibited in [PSW06b]. Both sides of the

condition

$$H_{\min}^{\varepsilon}(E)_{\pi_\Omega} + H_{\min}^{\varepsilon}(SE|R)_\rho \gtrapprox \log d_S \tag{5.10}$$

grow linearly with the size of the joint system if the ratio of the number of particles in S and E and their correlations to R are fixed. We would like to obtain a condition in terms of intensive quantities. In this case condition (5.10) turns into a simple inequality for a large enough system.

Consider a set of n weakly-interactive spin-$\frac{1}{2}$-systems in a magnetic field in z-direction that causes an energy gap Δ between the up- and down-states. With weakly-interactive we mean that the Hamiltonian is given by the number of up-spins plus a small perturbation or interaction term. The interaction term should allow to flip spins around while keeping the total number of up spins conserved. Let Ω denote the constraint given by energy conservation. We assume that the interaction term is weak enough such that all states with the same number of up-states are in the same energy-range described by the space $\mathcal{H}_\Omega$ (c.f. (3.59)). We consider that the system S is composed by a fraction α of the spins and call the remaining $(1-\alpha)n$ spins "the environment" E. The fraction of spins (in system and environment) in the up-state will be denoted by p, which, in equilibrium, is a function of the temperature of the spins. The observer's knowledge about the initial state ρ of the spins will be described by use of a reference R.

Corollary 5.2. *Consider the setting described above with $n \gg 1$ and $\frac{\alpha}{2} \leq p \leq 1 - \frac{\alpha}{2}$. If the condition*

$$H(p) + \frac{H_{\min}^{\varepsilon}(SE|R)_\rho}{n} > 2\alpha \tag{5.11}$$

is fullfilled (for $\varepsilon \to 0$), almost every evolution the spins can undergo will decouple the system from the reference and leave it in the canonical state.

Proof. Since the space $\mathcal{H}_\Omega$ is known, we can make use of the stronger bound in (5.4) and thus condition (5.10). From the definitions, we have $d_S = 2^{\alpha n}$ and $d_\Omega = \binom{n}{pn}$. π_Ω is a mixture of states all of which have probability $\binom{n}{pn}^{-1}$. The eigenvalues (probabilities) of the states in $\mathrm{Tr}_S\, \pi_\Omega$ are given by this factor times the number of states in $\mathcal{H}_\Omega$ wich lead to a particular state when tracing out S. The number of up-spins in S is upper-bouned by pn and lower-bounded by $\alpha n - (n - pn)$, so

$$\lambda_{\max}(\mathrm{Tr}_S\, \pi_\Omega) = \max_{\ell \in [\alpha n - (n-pn), pn]} \binom{\alpha n}{\ell} \cdot \binom{n}{pn}^{-1} \tag{5.12}$$

$$= \binom{\alpha n}{\frac{\alpha n}{2}} \cdot \binom{n}{pn}^{-1}. \tag{5.13}$$

For the second equality to be valid, we made use of the assumption $\frac{\alpha n}{2} \leq pn \leq n - \frac{\alpha n}{2}$. From (3.3) we obtain

$$\log \lambda_{\max}(\mathrm{Tr}_S \, \pi_\Omega) \approx n \, (\alpha - H(p)) \ . \tag{5.14}$$

So we have $H_{\min}^{\varepsilon}(E)_{\pi_\Omega} \approx n \, (H(p) - \alpha)$ and (5.10) yields

$$n \, (H(p) - \alpha) + H_{\min}^{\varepsilon}(SE|R)_\rho \gtrapprox \alpha n \ . \tag{5.15}$$

In the limit of large n (5.15) is fullfilled if

$$H(p) + \frac{H_{\min}^{\varepsilon}(SE|R)_\rho}{n} > 2\alpha \ . \tag{5.16}$$

□

In this simple example, (5.11) can also be obtained from the weaker bound, condition (5.3). In the derivation of the weaker bound from the stronger one, we make use of the inequality $H_{\min}^{\varepsilon}(E)_{\pi_\Omega} \geq H_{\min}^{\varepsilon}(E|S)_{\pi_\Omega}$. We therefore expect condition (5.10) to provide better conditions than condition (5.3) in the case where the states in $\mathcal{H}_\Omega$ are far from product. If we interpret $\mathcal{H}_\Omega$ as being the span of the set of energy eigenstates whose energy lies in a certain interval (as in the above example) this is the case if these energy eigenstates are strongly entangled, i.e. if the interaction between S and E is strong.

If inequality (5.11) is fullfilled, almost every evolution the spins can undergo will destroy all correlations between the reference and the system and leave the set of qubits we consider the system in the canonical state. There might be different reasons for this not to happen, corresponding to the three terms in (5.11):

- We might have too precise knowledge about the initial state, leading to a small $\frac{H_{\min}^{\varepsilon}(SE|R)_\rho}{n}$. If we allow for quantum-mechanical entanglement between R and SE, it may even be negative.

- The constraint set by energy conservation might be too severe leaving the evolution "no space" to destroy the correlations between the system and the reference. Physically, the temperature might be so low that almost all spins point in the same direction, leading to a small p and thus to a small $H(p)$.

- The environment might be too small in comparison to the system (leaving, again, not enough space for the evolution).

In a physical situation our knowledge about the different particles in what we call the system and the environment might not be specific for each particle but of the same form for all particles in the system and the environment. If the initial state is of the form

$$\rho_{SER} = \eta_{SR}^{\otimes \alpha n} \otimes \sigma_{ER}^{\otimes (1-\alpha)n} \tag{5.17}$$

with a product system

$$S = S_1 \dots S_{\alpha n} \tag{5.18}$$

and similarly for E and R we have by use of Theorem 3.5

$$\begin{aligned}
&\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} H_{\min}^{\varepsilon}(SE|R)_{\rho} \\
&\quad = \alpha \cdot \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{\alpha n} H_{\min}^{\varepsilon}(S|R)_{\eta^{\otimes \alpha n}} + (1-\alpha) \cdot \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{(1-\alpha)n} H_{\min}^{\varepsilon}(E|R)_{\sigma^{\otimes (1-\alpha)n}} \\
&\quad = \alpha \cdot H(S|R)_{\eta} + (1-\alpha) \cdot H(E|R)_{\sigma} \ .
\end{aligned} \tag{5.19}$$

Criterion (5.11) then simply becomes

$$H(p) + \alpha \cdot H(S|R)_{\eta} + (1-\alpha) \cdot H(E|R)_{\sigma} > 2\alpha \ . \tag{5.20}$$

We will evaluate criterion (5.11) for different choices of ρ_{SER}, describing different accuracies of knowledge about the initial state. These will be:

1. We know the initial pure state of the system.
2. We are initially fully entangled with the system.
3. We know the initial pure state of the system and the environment.
4. We are initially fully entangled with the system and the environment.

1. If we know the exact initial state of the system (a state with ℓ up-spins), but are completely ignorant about the rest of the spins (the environment), $2^{H_{\min}^{\varepsilon}(SE|R)_{\rho}}$ corresponds to the number of possible states of the environment which complement the state of the system to a state with pn up-spins. This number is given by $\binom{n-\alpha n}{pn-\ell}$. Using (3.3) we obtain

$$\log \binom{n-\alpha n}{pn-\ell} \approx (n-\alpha n) H(\frac{p - \frac{\ell}{n}}{1-\alpha}) \approx (1-\alpha) n H(p) \tag{5.21}$$

where the approximation holds for large n and small α. Consequently, $H_{\min}^{\varepsilon}(SE|R)_\rho \approx (1-\alpha)nH(p)$.

Remarkably, this approximation as well as the state the system will approach if the condition is fulfilled is independent of ℓ: If we have a system in contact with a much larger environment, the initial energy of the system is irrelevant for the evolved state at some later time. We get here a first glance at what we will later call independence of the initial state of the system.

Inserting this in our condition yields $H(p) > \frac{2\alpha}{2-\alpha} \approx \alpha$.

2. If we are initially *fully entangled* with the system and know nothing about the microstate of the environment, we have $H_{\min}^{\varepsilon}(SE|R)_\rho \approx -\alpha n + (1-\alpha)nH(p)$ which yields the condition $H(p) > \frac{3\alpha}{2-\alpha} \approx \frac{3\alpha}{2}$.

3. If acces to R tells us the exact initial state of SE, $\rho_{SER} = \sum_i p_i |i\rangle\langle i|_\Omega \otimes |i\rangle\langle i|_R$, we have $H_{\min}^{\varepsilon}(SE|R)_\rho = 0$ and the condition is $H(p) > 2\alpha$, in accordance with [PSW06b, Eq. (86)].

4. Finally, if we are fully entangled with the initial state of the system and the environment, $H_{\min}^{\varepsilon}(SE|R)_\rho = -\log d_\Omega = -\log \binom{n}{np} \approx -nH(p)$. In this case, (5.11) is not fullfilled for any temperature so our results do not predict decoupling.

In conclusion, we obtain the conditions listed in Table 5.1.

Knowledge about initial state of SE	To achieve decoupling, we need ...
$H(S\|R) = 0$, $H(E\|R) = (1-\alpha)nH(p)$	$H(p) > \alpha$
$H(S\|R) = -\alpha n$, $H(E\|R) = (1-\alpha)nH(p)$	$H(p) > \frac{3\alpha}{2}$
$H(SE\|R) = 0$	$H(p) > 2\alpha$
$H(SE\|R) = -nH(p)$	?

Table 5.1: Different levels of an observer's knowledge about the system and the environment and sufficient conditions for an evolution of the system to a canonical state decoupled from the observer. The conditions are expressed in terms of the temperature-dependent fraction of up-spins p and the fraction of spins which forms the system α. Note that the first and second condition were derived under the assumption that α is small.

In the $\alpha \ll \frac{1}{\sqrt{n}}$ regime, the canonical state π_S^Ω can be shown to be of the form, $\pi_S^\Omega = (p|1\rangle\langle 1| + (1-p)|0\rangle\langle 0|)^{\otimes \alpha n} \propto \exp(-\beta H_S)$, which takes the usual Boltzmann form. The fraction of spins up, p, is linked to the inverse temperature of the spins via $\beta = \frac{1}{\Delta}\ln(\frac{1-p}{p})$, as shown in [PSW06a]. The function $H(p(\beta))$ maps the inverse temperature-interval $(0,\infty)$ bijectively to the entropy-interval $\left(0, \frac{1}{2}\right)$.

5.3 Essentially tight version

5.3.1 Informal version

In the following we will prove a criterion for the two predictions about the evolved state $U_\Omega \rho_{SER} U_\Omega^\dagger$ which is tight up to differences between smooth min- and max-entropies. Except with an exponentially small probability, we obtain a canonical state on S which is decoupled from R if

$$H_{\min}^{\varepsilon}(SE|R)_\rho + H_{\min}^{\varepsilon}(E)_{\pi_\Omega} - H_{\max}^{\varepsilon}(S)_{\pi_\Omega} \gtrapprox 0 \ . \tag{5.22}$$

This will not be the case generically if

$$H_{\min}^{\varepsilon}(SE|R)_\rho + H_{\max}^{\varepsilon}(E)_{\pi_\Omega} - H_{\min}^{\varepsilon}(S)_{\pi_\Omega} \lessapprox 0 \ . \tag{5.23}$$

The differences $H_{\min}^{\varepsilon}(E)_{\pi_\Omega} - H_{\max}^{\varepsilon}(S)_{\pi_\Omega}$ and $H_{\max}^{\varepsilon}(E)_{\pi_\Omega} - H_{\min}^{\varepsilon}(S)_{\pi_\Omega}$ quantify, so to speak, how much the space $\mathcal{H}_\Omega$ describing the constraint is weighted towards the tensor factors $\mathcal{H}_S$ or $\mathcal{H}_E$ of the joint system. If, in a simplest case, $\mathcal{H}_\Omega = \mathcal{H}_{\Omega_S} \otimes \mathcal{H}_{\Omega_E}$ we have for $\varepsilon \to 0$

$$H_{\min}^{\varepsilon}(E)_{\pi_\Omega} - H_{\max}^{\varepsilon}(S)_{\pi_\Omega} = H_{\max}^{\varepsilon}(E)_{\pi_\Omega} - H_{\min}^{\varepsilon}(S)_{\pi_\Omega} = \log d_{\Omega_E} - \log d_{\Omega_S} \ . \tag{5.24}$$

5.3.2 Formal version

Theorem 5.3. *Let $\mathcal{H}_\Omega \subseteq \mathcal{H}_S \otimes \mathcal{H}_E$ and let $\rho_{SER} \in \mathcal{S}_=(\mathcal{H}_\Omega \otimes \mathcal{H}_R)$.*
Achievability. *Let $\varepsilon \geq 0$. Then,*

$$\begin{aligned} &\int_{\mathbb{U}(\Omega)} || \operatorname{Tr}_E(U_\Omega \rho_{SER} U_\Omega^\dagger) - \pi_S^\Omega \otimes \rho_R ||_1 \mathrm{d}U \\ &\qquad \leq 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(SE|R)_\rho + \frac{1}{2} H_{\max}^{\frac{\varepsilon}{2}}(S)_{\pi_\Omega} - \frac{1}{2} H_{\min}^{\frac{\varepsilon}{2}}(E)_{\pi_\Omega} + \log \frac{24}{\varepsilon^2}} + 12\varepsilon \ . \end{aligned} \tag{5.25}$$

Converse. *For any $\varepsilon' > 0$ and ε'', $\varepsilon''' \geq 0$, suppose that*

$$H_{\min}^{\varepsilon' + 2\varepsilon'' + \varepsilon''' + \sqrt{\varepsilon}}(SE|R)_\rho + H_{\max}^{\varepsilon''}(E)_{\pi_\Omega} - H_{\min}^{\varepsilon'''}(S)_{\pi_\Omega} < -\log \frac{2}{\varepsilon'^2} \ . \tag{5.26}$$

Then there is no state $\omega_S \in \mathcal{S}_=(\mathcal{H}_S)$ such that

$$\int_{\mathbb{U}(\Omega)} \left\| \operatorname{Tr}_E(U_\Omega \rho_{SER} U_\Omega^\dagger) - \omega_S \otimes \rho_R \right\|_1 \mathrm{d}U \leq \frac{\varepsilon}{2} \tag{5.27}$$

and in particular

$$\int_{\mathbb{U}(\Omega)} \left\| \mathrm{Tr}_E(U_\Omega \rho_{SER} U_\Omega^\dagger) - \pi_S^\Omega \otimes \rho_R \right\|_1 \mathrm{d}U > \varepsilon \ . \tag{5.28}$$

Proof. The sufficiency statement follows by a similar proof as in the first version of the theorem. Using Lemma 3.8 instead of Lemma 3.7 to lower-bound $H_{\min}^{\varepsilon}(\Omega'|S)_\tau$ we find

$$H_{\min}^{\varepsilon}(\Omega'|S)_\tau \geq H_{\min}^{\frac{\varepsilon}{2}}(S\Omega')_\tau - H_{\max}^{\frac{\varepsilon}{2}}(S)_\tau - 2 \cdot \log \frac{24}{\varepsilon^2} \tag{5.29}$$

with $\tau_{S\Omega'} = \mathrm{Tr}_E \, \Psi_{\Omega\Omega'}$. Lemma 3.4 again yields

$$\begin{aligned} H_{\min}^{\varepsilon}(\Omega'|S)_\tau &\geq H_{\min}^{\frac{\varepsilon}{2}}(E)_{\Psi_{\Omega\Omega'}} - H_{\max}^{\frac{\varepsilon}{2}}(S)_\tau - 2 \cdot \log \frac{24}{\varepsilon^2} \\ &= H_{\min}^{\frac{\varepsilon}{2}}(E)_{\pi_\Omega} - H_{\max}^{\frac{\varepsilon}{2}}(S)_{\pi_\Omega} - 2 \cdot \log \frac{24}{\varepsilon^2} \ . \end{aligned} \tag{5.30}$$

The converse statement follows from Theorem 4.4, which gives the condition

$$\begin{aligned} -\log \frac{2}{\varepsilon'^2} &> H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(\Omega|R)_\rho + H_{\max}^{\varepsilon''}(\Omega' S)_\tau - H_{\min}^{\varepsilon'''}(S)_\tau \\ &= H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(\Omega|R)_\rho + H_{\max}^{\varepsilon''}(E)_\tau - H_{\min}^{\varepsilon'''}(S)_\tau \\ &= H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(\Omega|R)_\rho + H_{\max}^{\varepsilon''}(E)_{\pi_\Omega} - H_{\min}^{\varepsilon'''}(S)_{\pi_\Omega} \ . \end{aligned} \tag{5.31}$$

The first equality is due to an application of Lemma 3.4. □

5.4 Unitary 2-designs

In this chapter we have adapted a point of view in which the average over all unitaries U on $\mathcal{H}_\Omega$ is interpreted as the average over the evolutions the joint system can undergo while subject to the constraint Ω. The averaging is over the Haar measure which seems a natural choice. The defining property of the Haar measure, namely being the unique left-invariant measure on $\mathbb{U}(\mathcal{H}_\Omega)$, translates into the physical statement that if the evolution U_Ω we average about is preceded by another evolution respecting Ω, this does not lead to different averages.

However, a given physical system may not be able to exactly realize every unitary or not even to approximate every unitary efficiently. Furthermore, it has been shown that, under assumption of the Church-Turing thesis, some (even finite dimensional)

unitary operators may not be realized as dynamics by *any* Hamiltonian [Nie97]. A more physical way of averaging over possible evolutions is thus given by a *unitary 2-design* which is defined as a finite set of unitaries which reproduces the first and second moments of the Haar distribution.

It has been shown [HL09] that random concatenations of two-qubit interactions approximate the first and second moment of the Haar distribution (and thus constitute *approximate 2-designs*) in a time which is polynomial in the number of qubits. It has also been shown [Sze11] that approximate 2-designs approximate the predictions of the decoupling theorem.

Since the random two-qubit interactions do not preserve any non-trivial subspace of $\mathcal{H}_S \otimes \mathcal{H}_E$ we set $\mathcal{H}_\Omega = \mathcal{H}_S \otimes \mathcal{H}_E$. The fraction $\frac{d_S}{\sqrt{d_\Omega}}$ in Theorem 5.1 thus turns into $\sqrt{\frac{d_S}{d_E}}$. The dimensions grow exponentially with the number of particles. Therefore, if SE consists of a large number of particles this fraction is small whenever we consider less then half of the particles to constitute the "system". For a classical reference the conditional entropy term which then will also appear in the exponent in Theorem 5.1 is non-negative. Hence by our theorem *random two-qubit interactions efficiently decouple any subset of the qubits which consists of less than half of them from a classical reference and leave it in the canonical state.*

In order to formalize the above, we first provide a formal version of the two cited theorems. An ε-approximate unitary 2-design is a finite set of unitaries and probabilities attached to them, so that applying those unitaries (weighted with the corresponding probabilities) to a state approximates the first and second moments of the Haar distribution to accuracy ε. For lack of need we omit a formal definition of the term (which would require to first introduce the *diamond norm* for channels) but refer to [HL09].

Theorem 5.4. *[HL09, Theorems 2.9. and 2.10.] Consider a distribution μ on $\mathbb{U}(4)$ which is either a universal gate set on it or an approximate unitary 2-design on two qubits. Draw T random unitaries according to μ. For each such unitary, chose a random pair from n qubits and apply the unitary to that pair. Then there is $C(\mu)$ such that $\forall \varepsilon > 0$ and $\forall T > C(\mu)(n^2 + n \log \frac{1}{\varepsilon})$ the mapping obtained this way constitues an ε-approximate unitary 2-design.*

Examples of universal gate sets on $\mathbb{U}(4)$ include $\mathbb{U}(4)$ itself or any entangling gate (e.g. CNOT) together with all single qubit gates.

Theorem 5.5. *[Sze11, p. 34] Let $\rho_{AR} \in \mathcal{S}_\leq(\mathcal{H}_A \otimes \mathcal{H}_R)$ and let $\mathcal{T}_{A\to B}$ be a CPTPM with Choi-Jamiołkowski isomorphism $\tau_{A'B}$. Let $\langle \ldots \rangle_\triangle$ denote the weighted averaged*

over the application of unitaries U_A from an ε-approximate 2-design on $\mathcal{H}_A$. Then for small enough δ we have

$$\left\langle \left\| \mathcal{T}_{A\to B}(U_A \rho_{AR} U_A^\dagger) - \tau_B \otimes \rho_R \right\|_1 \right\rangle_\triangle \leq \sqrt{1 + 4\varepsilon d_A^4} \cdot 2^{-\frac{1}{2} H_{\min}^\delta(A'|B)_\tau - \frac{1}{2} H_{\min}^\delta(A|R)_\rho} + 8 d_A \varepsilon \delta + 12\delta \ . \tag{5.32}$$

Combining this two theorems with Theorem 5.1 we find the following theorem.

Theorem 5.6. *Let $\mathcal{H}_A := (\mathbb{C}^2)^{\otimes n}$. Let $\Lambda \subseteq \{1, \ldots, n\}$, $\bar{\Lambda} := \{1, \ldots, n\} \setminus \Lambda$ and $\alpha := \frac{|\Lambda|}{n}$. Let μ denote the probability distribution corresponding to any universal gate set on $\mathbb{U}(4)$ or an approximate 2-design on two qubits. Draw T random unitaries according to μ and apply each of them to a random pair chosen from the n qubits. Let $\langle \ldots \rangle_\triangle$ denote the average over possible evolutions $U_A \in \mathbb{U}(\mathcal{H}_A)$ achievable with the T random unitaries weighted with the corresponding probabilities according to μ. Then there is $C(\mu)$ such that $\forall \varepsilon \geq 0$, $\forall T > C(\mu) n^2$, $\forall \rho_{AR} \in \mathcal{S}_= (\mathcal{H}_A \otimes \mathcal{H}_R)$ we have*

$$\left\langle \left\| \mathrm{Tr}_{\bar{\Lambda}} \left(U_A \rho_{AR} U_A^\dagger \right) - \pi_\Lambda \otimes \rho_R \right\|_1 \right\rangle_\triangle \leq 2^{-\frac{1}{2}(1-2\alpha)n} \cdot 2^{-\frac{1}{2} H_{\min}^\varepsilon(A|R)_\rho} + 12\varepsilon \ . \tag{5.33}$$

As discussed informally above, for large n and non-negative $H_{\min}^\varepsilon(A|R)_\rho$ a time which grows quadratically in the number of qubits is sufficient to decouple less than half of them from the reference and maximize their entropy. The requirement that we have to consider less than half of the qubits for an initial state with non-negative conditional entropy coincides with the condition obtained in Corollary 5.2 in the limit where the magnetic field-to-temperature ratio is zero. As in Corollary 5.2 we cannot predict decoupling for any $\alpha > 0$ if the qubits are initially fully entangled with the reference.

Proof. From Theorem 5.4 we know that the T random unitaries drawn according to μ and applied to a random pair of qubits constitute an *ε-approximate unitary 2-design* if $T > C(\mu)(n^2 + n \log \frac{1}{\varepsilon})$. We apply Theorem 5.5 where A is the system of the n qubits, B the qubits described by the index set Λ and $\mathcal{T}_{A\to B} = \mathrm{Tr}_{\bar{\Lambda}}$ the partial trace over spaces corresponding to the qubits which are not in this set. This gives us

$$\left\langle \left\| \mathrm{Tr}_{\bar{\Lambda}} \left(U_A \rho_{AR} U_A^\dagger \right) - \tau_\Lambda \otimes \rho_R \right\|_1 \right\rangle_T \leq \sqrt{1 + 4\varepsilon (2^n)^4} \cdot 2^{-\frac{1}{2} H_{\min}^\varepsilon(A'|\Lambda)_{\tau_{\Lambda A'}}} \cdot 2^{-\frac{1}{2} H_{\min}^\varepsilon(A|R)_\rho} + 8 \cdot 2^n \varepsilon \delta + 12\varepsilon \tag{5.34}$$

where $\tau_{\Lambda A'} = \mathrm{Tr}_{\bar{\Lambda}} \Psi_{AA'}$ is the Choi-Jamiołkowski isomorphism of the partial trace and $\Psi_{AA'}$ denotes the fully entangled state between the n qubits and a copy of them. τ_Λ

is thus equal to π_Λ. The sufficient number of gates given by $C(\mu)(n^2 + n\log\frac{1}{\varepsilon})$ turns for $\varepsilon = (2^n)^{-m}$ into $C(\mu)(m+1)n^2$. Choosing an $m > 4$ we can make $\sqrt{1+4\varepsilon(2^n)^4}$ arbitrarily close to 1 and $8\cdot 2^n\varepsilon\delta$ arbitrarily small and then absorb the factor $m+1$ into $C(\mu)$. For large n it is sufficient to chose $m = 5$. Following the same steps as in the proof of Theorem 5.1 we can show that

$$H_{\min}^{\varepsilon}(A'|\Lambda)_{\omega_{\Lambda A'}} \geq \log(2^n) - 2\log(2^{|\Lambda|}) = n(1-2\alpha) \tag{5.35}$$

which after relabling $\delta \mapsto \varepsilon$ concludes the proof. □

Chapter 6

Thermalization: Temporal averages

6.1 The long-term perspective

As poined out in [LPSW09], the straighforward-looking process of thermalization actually consists of four aspects which may be addressed independently.

1. *Equilibration.* The initial state of the system evolves towards some particular equilibrium state and remains close to it for almost all times.

2. *Independence of the initial state of the environment.* The equilibrium state of the system does not depend on the precise initial state of the environment but only on macroscopic parameters describing it, like its temperature.

3. *Independence of the initial state of the system.* If the system is small compared to the environment, its equilibrium state should be independent of its initial state.

4. *Boltzmannity of the equilibrium state.* The equilibrium state state of the system takes the familiar Boltzmann form $\frac{1}{Z(\beta)}e^{-\beta H_S}$.

We will address the first three points in this thesis and refer to [RGE11] for a derivation of the fourth point in a weak-coupling limit between the system and the environment and a discussion of what the relevant weak-coupling limit is.

In order to show that almost all initial states equilibrate we want the quantity

$$\left\langle \left\| \rho_S(t) - \langle \rho_S(t) \rangle_t \right\|_1 \right\rangle_{t,\rho(0)} \tag{6.1}$$

to be small. With $\langle \ldots \rangle_{\rho(0)}$ we denote a not yet precisely defined average over initial states. For the second and third point we want a quantity of the form

$$\left\langle \left\| \langle \rho_S(t) \rangle_t - \langle \rho_S(t) \rangle_{t,\rho(0)} \right\|_1 \right\rangle_{\rho(0)} \tag{6.2}$$

to be small. If we can show that these two quantities are small, then by the triangle inequality the quantity

$$\left\langle \left\| \rho_S(t) - \langle \rho_S(t) \rangle_{t,\rho(0)} \right\|_1 \right\rangle_{t,\rho(0)} \tag{6.3}$$

will be small as well, i.e. almost all initial states will lead to the system being for almost all times close to a definite equilibrium state.

Expressions obtained from the decoupling theorem are always of the form

$$\int \|\ldots\|_1 \, \mathrm{d}U \leq \ldots$$

where the integration is with respect to the Haar measure on the group of unitaries acting on a certain space. Hence there is hope to show by use of the decoupling theorem that expression (6.2) is small but there is no hope to show that expression (6.1) is small.

That time-averaging cannot be written as a Haar measure average over unitaries can be seen as follows. The time-evolution leads to unitaries of the form $e^{-\mathbb{i}Ht}$ acting on density operators by conjugation. Any density operator which is a mixture of energy eigenstates is conserved by such a unitary. On the other hand, a unitary from the Haar measure on the group of unitaries acting on a certain space will with unit probability not leave it invariant.

We will therefore have to show that (6.1) is small in a direct calculation. This has in fact already been done in [LPSW09] for pure initial states. We generalize these calculations to mixed initial states since when dealing with the decoupling theorem we always allow for mixed states. Then by use of the decoupling theorem we give conditions for (6.2) being small, thereby extending the equivalent theorem in [LPSW09] and providing a technically simpler proof.

We take in this chapter a long-term perspective in that we are only concerned with temporal averages (whether the temporal average of the distance of the state of the system from its temporal average is small and whether the time-averaged state is independent of the initial state). If the temporal average of a distance is small we can by Markov's inequality (4.3) conclude that the fraction of times, for which the distance is large, is small. Still this does not enable us to make definite statements

about the state of the system at a given time. Of special interest is the question how long the system needs to reach the equilibrium state or to become independent of its initial state. The results of this chapter do not tell us anything about this. We will be concerned with problems like these in the following three chapters.

We assume in this chapter that the Hamiltonian has *non-degenerate energy gaps* [LPSW09], i.e.

$$E_i - E_j = E_k - E_l \Rightarrow (i = j \wedge k = l) \vee (i = k \wedge j = l) \ . \tag{6.4}$$

This in particular implies that the energy levels are non-degenerate. The assumption is very weak since it only excludes a set of Hamiltonians with zero measure, any arbitrarily small perturbation will lift all degeneracies with unit probability.

6.2 Equilibration

Theorem 6.1 (Generalization of [LPSW09, Theorem 1])**.** *Consider an initial state $\rho_{SE}(0) \in \mathcal{S}_=(\mathcal{H}_S \otimes \mathcal{H}_E)$ which is subject to an evolution governed by a Hamiltonian with non-degenerate energy gaps. Let $\langle \ldots \rangle_t$ denote the temporal average under this evolution. Define $\omega := \langle \rho(t) \rangle_t$. Then,*

$$\langle \| \rho_S(t) - \omega_S \|_1 \rangle_t \leq \sqrt{d_S} 2^{-\frac{1}{2} H_2(E)_\omega} \leq d_S 2^{-\frac{1}{2} H_2(SE)_\omega} \ . \tag{6.5}$$

If we have some knowledge about the initial state stored in a classical reference R, then

$$\langle \| \rho_{SR}(t) - \omega_{SR} \|_1 \rangle_t \leq \sqrt{d_S} 2^{-\frac{1}{2} H^\varepsilon_{\min}(E|R)_\omega} + 4\varepsilon \leq d_S 2^{-\frac{1}{2} H^\varepsilon_{\min}(SE|R)_\omega} + 4\varepsilon \ . \tag{6.6}$$

We leave it as an open problem to generalize the above bounds to the case of a quantum reference.

We will sketch an example which illustrates why conditioning the smooth min-entropy on R is relevant. Let

$$\rho_{SER}(0) = \frac{1}{d_S d_E} \sum_{k=1}^{d_S d_E} |k\rangle\langle k|_{SE} \otimes |k\rangle\langle k|_R \tag{6.7}$$

with $|k\rangle_{SE} = \frac{1}{\sqrt{2}}|E_k\rangle + \frac{1}{\sqrt{2}}|E_{k+1}\rangle$. In this case we have (for $\varepsilon \to 0$) $H_2(SE)_\omega = H^\varepsilon_{\min}(SE)_\omega = \log d_S + \log d_E$ which would predict equilibration for $\log d_E \gtrapprox \log d_S$

if conditioning were not necessary. On the other hand, $H_{\min}^{\varepsilon}(SE|R)_{\omega} = 1$, so we do not predict equilibration. In fact we have

$$\begin{aligned}&\|\rho_{SR}(t) - \omega_{SR}\|_1 \\ &\quad = \frac{1}{2d_S d_E} \sum_{k=1}^{d_S d_E} \left\| e^{-\mathbb{i}(E_k - E_{k+1})t} \operatorname{Tr}_E |E_k\rangle\langle E_{k+1}| + e^{+\mathbb{i}(E_k - E_{k+1})t} \operatorname{Tr}_E |E_{k+1}\rangle\langle E_k| \right\|_1\end{aligned} \tag{6.8}$$

for all times t.

Now assume that the basis $\{|E_k\rangle\}_{k=1,\ldots,d_S d_E}$ is close to a product basis $\{|i\rangle_S \otimes |j\rangle_E\}_{i=1,\ldots,d_S, j=1,\ldots,d_E}$. Assume furthermore that the $|E_k\rangle$ are ordered such that states which look similar on E are grouped together. Then most summands in (6.8) are close to 2 and $\|\rho_{SR}(t) - \omega_{SR}\|_1$ can certainly be made larger than, say, $\frac{1}{2}$ for all times t.

Proof. We first prove the part without an involved reference. Let the Hamiltonian of the joint system be $H_{SE} = \sum_k E_k |E_k\rangle\langle E_k|$ and let $\rho(0) = \sum_{i,j} \rho_{ij} |E_i\rangle\langle E_j|$. The assumption of non-degenerate energy gaps implies non-degenerate energy-levels (i.e. $E_i = E_j \Rightarrow i = j$) so that according to eqn. (2.29) temporal averaging cancels all off-diagonal terms, $\omega_{SE} = \sum_k \rho_{kk} |E_k\rangle\langle E_k|$. We conclude that

$$\rho_S(t) - \omega_S = \sum_{i \neq j} \rho_{ij} e^{-\mathbb{i}(E_i - E_j)t} \operatorname{Tr}_E |E_i\rangle\langle E_j| \; . \tag{6.9}$$

By use of a standard inequality [FvdG99] we have

$$\|\rho_S(t) - \omega_S\|_1 \leq \sqrt{d_S} \, \|\rho_S(t) - \omega_S\|_2 = \sqrt{d_S \operatorname{Tr}_S (\rho_S(t) - \omega_S)^2} \; . \tag{6.10}$$

Using the concavity of the square root function, the time-average yields

$$\langle \|\rho_S(t) - \omega_S\|_1 \rangle_t \leq \sqrt{d_S \operatorname{Tr}_S \langle (\rho_S(t) - \omega_S)^2 \rangle_t} \tag{6.11}$$

where

$$\langle (\rho_S(t) - \omega_S)^2 \rangle_t = \sum_{i \neq j} \sum_{m \neq n} \rho_{ij} \rho_{mn} \langle e^{-\mathbb{i}(E_i - E_j + E_m - E_n)t} \rangle_t \operatorname{Tr}_E |E_i\rangle\langle E_j| \operatorname{Tr}_E |E_m\rangle\langle E_n| \; . \tag{6.12}$$

By use of the assumption, this time-average can only be non-zero if $i = n$ and $j = m$, which simplifies the term in our bound to

$$\operatorname{Tr}_S \langle (\rho_S(t) - \omega_S)^2 \rangle_t = \sum_{i \neq j} \rho_{ij} \rho_{ji} \operatorname{Tr}_S \left(\operatorname{Tr}_E |E_i\rangle\langle E_j| \operatorname{Tr}_E |E_j\rangle\langle E_i| \right) \; . \tag{6.13}$$

We assume that $\{|s\rangle_S\}_s$ and $\{|e\rangle_E\}_e$ are orthonormal bases of S and E, respectively, and introduce an identity operator in the form $\mathbb{1}_S = \sum_{s'} |s'\rangle\langle s'|_S$.

$$
\begin{aligned}
&\mathrm{Tr}_S\langle(\rho_S(t)-\omega_S)^2\rangle_t \\
&\quad= \sum_{i\neq j} \rho_{ij}\rho_{ji} \,\mathrm{Tr}_S\left(\mathrm{Tr}_E\,|E_i\rangle\langle E_j|\,\mathrm{Tr}_E\,|E_j\rangle\langle E_i|\right) \\
&\quad= \sum_{i\neq j} \rho_{ij}\rho_{ji} \sum_{ss'ee'} \langle se|E_i\rangle\langle E_j|s'e\rangle\langle s'e'|E_j\rangle\langle E_i|se'\rangle \\
&\quad= \sum_{i\neq j} \rho_{ij}\rho_{ji} \sum_{ss'ee'} \langle se|E_i\rangle\,\langle E_i|se'\rangle\,\langle s'e'|E_j\rangle\,\langle E_j|s'e\rangle \\
&\quad= \sum_{i\neq j} \rho_{ij}\rho_{ji} \,\mathrm{Tr}_E\left(\mathrm{Tr}_S\,|E_i\rangle\langle E_i|\,\mathrm{Tr}_S\,|E_j\rangle\langle E_j|\right) \\
&\quad= \sum_{i\neq j} \underbrace{(\rho_{ij}\rho_{ji}-\rho_{ii}\rho_{jj})}_{\leq 0} \underbrace{\mathrm{Tr}_E\left(\mathrm{Tr}_S\,|E_i\rangle\langle E_i|\,\mathrm{Tr}_S\,|E_j\rangle\langle E_j|\right)}_{\geq 0} \\
&\quad+ \sum_{i\neq j} \rho_{ii}\rho_{jj} \,\mathrm{Tr}_E\left(\mathrm{Tr}_S\,|E_i\rangle\langle E_i|\,\mathrm{Tr}_S\,|E_j\rangle\langle E_j|\right) \\
&\quad\leq \sum_{i\neq j} \rho_{ii}\rho_{jj} \,\mathrm{Tr}_E\left(\mathrm{Tr}_S\,|E_i\rangle\langle E_i|\,\mathrm{Tr}_S\,|E_j\rangle\langle E_j|\right) \\
&\quad\leq \sum_{i\neq j} \rho_{ii}\rho_{jj} \,\mathrm{Tr}_E\left(\mathrm{Tr}_S\,|E_i\rangle\langle E_i|\,\mathrm{Tr}_S\,|E_j\rangle\langle E_j|\right) + \sum_{i=j} \underbrace{\rho_{ii}}_{\geq 0}\rho_{jj} \,\mathrm{Tr}_E\left(\mathrm{Tr}_S\,|E_i\rangle\langle E_i|\,\mathrm{Tr}_S\,|E_j\rangle\langle E_j|\right) \\
&\quad= \sum_{ij} \rho_{ii}\rho_{jj} \,\mathrm{Tr}_E\left(\mathrm{Tr}_S\,|E_i\rangle\langle E_i|\,\mathrm{Tr}_S\,|E_j\rangle\langle E_j|\right) \\
&\quad= \mathrm{Tr}_E\left(\omega_E^2\right) \\
&\quad= 2^{-H_2(E)_\omega}\ . \qquad (6.14)
\end{aligned}
$$

During the derivation we made use of the following inequalities:

- $\rho_{ii}\rho_{jj} - \rho_{ij}\rho_{ji} \geq 0$: Follows from the positive semi-definiteness of the matrix $\begin{pmatrix} \rho_{ii} & \rho_{ij} \\ \rho_{ji} & \rho_{jj} \end{pmatrix}$ which itself follows from the fact that $\forall \alpha, \beta \in \mathbb{C}$ we have $\left(\bar{\alpha}\langle E_i| + \bar{\beta}\langle E_j|\right)\rho\left(\alpha|E_i\rangle + \beta|E_j\rangle\right) \geq 0$. C.f. [HJ05, Observation 7.1.2].
- $\mathrm{Tr}_E\left(\mathrm{Tr}_S\,|E_i\rangle\langle E_i|\,\mathrm{Tr}_S\,|E_j\rangle\langle E_j|\right) \geq 0$: The trace of the product of two positive semi-definite operators is non-negative. Since $\mathrm{Tr}_S\,|E_i\rangle\langle E_i|$ is a positive semi-definite operator it allows for a spectral decomposition $\mathrm{Tr}_S\,|E_i\rangle\langle E_i| =$

$\sum_r \underbrace{\lambda_r}_{\geq 0} |e_r\rangle\langle e_r|$ and hence

$$\mathrm{Tr}_E \left(\mathrm{Tr}_S |E_i\rangle\langle E_i| \, \mathrm{Tr}_S |E_j\rangle\langle E_j| \right) = \sum_r \lambda_r \underbrace{\langle e_r| \left(\mathrm{Tr}_S |E_j\rangle\langle E_j| \right) |e_r\rangle}_{\geq 0} \geq 0 \ .$$

- $\rho_{ii} \geq 0$: Follows directly from $\langle E_i|\rho|E_i\rangle \geq 0$.

We conclude that

$$\langle \|\rho_S(t) - \omega_S\|_1 \rangle_t \leq \sqrt{d_S \, \mathrm{Tr}_S \langle (\rho_S(t) - \omega_S)^2 \rangle_t} \leq \sqrt{d_S} 2^{-\frac{1}{2} H_2(E)_\omega} \ , \tag{6.15}$$

which concludes the proof of the stronger bound in the first part of the theorem. The weak subadditivity of the Rényi entropies [vDH02] gives

$$H_2(E) \geq H_2(SE) - H_0(S) \geq H_2(SE) - \log d_S \ , \tag{6.16}$$

which proves the weaker bound.

In order to prove the second part of the theorem, first note that since by assumption R is classical, we can according to eqn. (2.18) write

$$\rho_{SER}(0) = \sum_i p_i \rho_{SE}^{(i)} \otimes |i\rangle\langle i|_R \ . \tag{6.17}$$

The time-averaged state is given by

$$\omega_{SR} = \sum_i p_i \omega_S^{(i)} \otimes |i\rangle\langle i|_R \tag{6.18}$$

and hence by use of the triangle inequality and eqn. (6.15) we have

$$\begin{aligned}
\langle \|\rho_{SR}(t) - \omega_{SR}\|_1 \rangle_t &= \left\langle \left\| \sum_i p_i \left(\rho_S^{(i)} - \omega_S^{(i)} \right) \otimes |i\rangle\langle i|_R \right\|_1 \right\rangle_t \\
&\leq \sum_i p_i \left\langle \left\| \left(\rho_S^{(i)} - \omega_S^{(i)} \right) \otimes |i\rangle\langle i|_R \right\|_1 \right\rangle_t \\
&= \sum_i p_i \left\langle \left\| \rho_S^{(i)} - \omega_S^{(i)} \right\|_1 \right\rangle_t \\
&\leq \sum_i p_i \sqrt{d_S} 2^{-\frac{1}{2} H_2(E)_{\omega^{(i)}}} \\
&\leq \sqrt{d_S} \sum_i p_i 2^{-\frac{1}{2} H_{\min}(E)_{\omega^{(i)}}} \\
&= \sqrt{d_S} 2^{-\frac{1}{2} H_{\min}(E|R)_\omega} \ .
\end{aligned} \tag{6.19}$$

For the last equality we used eqn. (3.62). Now let $\tilde{\rho}_{SER}(0) \in \mathcal{B}^{\varepsilon}\left(\rho_{SER}(0)\right)$. Since the purified distance is preserved by the unitary evolution and since it can only decrease under partial traces we have that $P\left(\tilde{\rho}_{SR}(t), \rho_{SR}(t)\right) \leq \varepsilon$ for all t and hence also $P\left(\tilde{\omega}_{SR}, \omega_{SR}\right) \leq \varepsilon$. So using Lemma 3.3 we find

$$\begin{aligned}\langle \|\rho_{SR}(t) - \omega_{SR}\|_1 \rangle_t &\leq \langle \|\rho_{SR}(t) - \tilde{\rho}_{SR}(t)\|_1 \rangle_t + \langle \|\tilde{\rho}_{SR}(t) - \tilde{\omega}_{SR}\|_1 \rangle_t + \langle \|\tilde{\omega}_{SR} - \omega_{SR}\|_1 \rangle_t \\ &\leq \sqrt{d_S} 2^{-\frac{1}{2} H_{\min}(E|R)_{\tilde{\omega}}} + 4\varepsilon \ . \end{aligned} \tag{6.20}$$

We conclude that

$$\langle \|\rho_{SR}(t) - \omega_{SR}\|_1 \rangle_t \leq \sqrt{d_S} 2^{-\frac{1}{2} H^{\varepsilon}_{\min}(E|R)_{\omega}} + 4\varepsilon \ . \tag{6.21}$$

The weaker bound is then obtained by applying Lemma A.5 to each element in $\mathcal{B}^{\varepsilon}(\rho_{SER})$ separately. □

This is a very strong statement. If the unitary evolution leads the joint state of S and E through many different states and thus its time-average has a high entropy (as measured by $H_2(SE)_{\omega}$ or $H^{\varepsilon}_{\min}(SE|R)_{\omega}$), the system will definitely (and not only very likely, as in most other theorems in this thesis) spend most of the time close to its temporal average. If we consider an initial state far from its temporal average we can predict that the distance will become small. Additionally it is shown in [LPSW10] that besides being close to its temporal average for most times, the state $\rho_S(t)$ of the system only fluctuates slowly around it. In [Sho11] it is shown that the expectation value of "realistic" quantum observables will equilibrate, which does not only hold when looking at the smaller part of a bipartite system, but for any quantum mechanical system. This is then used to derive (6.5).

It can be shown that the $H_2(SE)_{\omega}$ term is on average of the order of $\log d_{\Omega} - 1$ when the initial state $\rho_{SE}(0)$ is drawn from the Haar measure on a large enough space $\mathcal{H}_{\Omega} \subseteq \mathcal{H}_S \otimes \mathcal{H}_E$. Combining this with the above theorem yields the statement that if we draw the initial state from a space which is much larger than $\mathcal{H}_S$ the system will most likely be close to its temporal average for most times.

Theorem 6.2 (Generalization of [LPSW09, Theorem 2 (i)])**.** *Let $\mathcal{H}_{\Omega} \subseteq \mathcal{H}_S \otimes \mathcal{H}_E$ and $\rho_{SE}(0) \in \mathcal{S}_{=}(\mathcal{H}_{\Omega})$. Let $\omega^U := \langle U \rho U^{\dagger} \rangle_t$ denote the temporal average under the evolution governed by a non-degenerate Hamiltonian H_{SE}. Then,*

$$\int_{\mathbb{U}(\Omega)} 2^{-H_2(SE)_{\omega^U}} \mathrm{d}U < \frac{2}{d_{\Omega}} \ . \tag{6.22}$$

By use of the convexity of the function $x \mapsto \frac{1}{x}$ we can also conclude from this that

$$\begin{aligned}\int_{\mathbb{U}(\Omega)} 2^{H_2(SE)_{\omega^U}} \mathrm{d}U &= \int_{\mathbb{U}(\Omega)} \frac{1}{2^{-H_2(SE)_{\omega^U}}} \mathrm{d}U \\ &\geq \left(\int_{\mathbb{U}(\Omega)} 2^{-H_2(SE)_{\omega^U}} \mathrm{d}U \right)^{-1} \\ &> \frac{d_\Omega}{2} \ . \end{aligned} \tag{6.23}$$

Proof. We first derive a bound on the averaged purity. In order to separate the integral over U from the rest we define the time-averaging operator $\mathbb{T}$ which for a non-degenerate Hamiltonian is given by

$$\mathbb{T}(\rho) := \langle \rho \rangle_t = \sum_k |E_k\rangle\langle E_k|\rho|E_k\rangle\langle E_k| \ . \tag{6.24}$$

By use of Lemma A.6 we have

$$\begin{aligned}\int_{\mathbb{U}(\Omega)} p(\omega^U) \mathrm{d}U &= \int_{\mathbb{U}(\Omega)} \mathrm{Tr}\left[\omega^U \cdot \omega^U\right] \mathrm{d}U \\ &= \int_{\mathbb{U}(\Omega)} \mathrm{Tr}\left[(\omega^U \otimes \omega^U)\mathbb{S}\right] \mathrm{d}U \\ &= \int_{\mathbb{U}(\Omega)} \mathrm{Tr}\left[(\mathbb{T}\otimes\mathbb{T})(U\rho U^\dagger \otimes U\rho U^\dagger)\mathbb{S}\right] \mathrm{d}U \\ &= \mathrm{Tr}\left[(\mathbb{T}\otimes\mathbb{T})\left\{\int_{\mathbb{U}(\Omega)} U^{\otimes 2}(\rho\otimes\rho)(U^\dagger)^{\otimes 2}\mathrm{d}U\right\}\mathbb{S}\right] \end{aligned} \tag{6.25}$$

where $\mathbb{S} = \mathbb{S}_{ES\leftrightarrow E'S'}$ denotes a "SWAP"-operator as introduced in Lemma A.6. It "swaps" SE and a copy $S'E'$ of it.

In the last equality we made use of the linearity of all operators. The integral is solved in Lemma A.8 and leads to

$$\int_{\mathbb{U}(\Omega)} p(\omega^U)\mathrm{d}U = \mathrm{Tr}\left[(\mathbb{T}\otimes\mathbb{T})\left\{\frac{d_\Omega - p(\rho)}{d_\Omega^3 - d_\Omega}\cdot \mathbb{1}_{\Omega\Omega'} + \frac{p(\rho)d_\Omega - 1}{d_\Omega^3 - d_\Omega}\cdot \mathbb{S}_{\Omega\leftrightarrow\Omega'}\right\}\mathbb{S}_{ES\leftrightarrow E'S'}\right] \tag{6.26}$$

Again, $\mathbb{S}_{\Omega\leftrightarrow\Omega'}$ denotes a "SWAP"-operator as introduced in Lemma A.6. In order to compute the trace we introdue as a shorthand notation $|k\rangle \equiv |E_k\rangle$ and $|kl\rangle \equiv$

$|E_k\rangle \otimes |E_l\rangle$ and find

$$\begin{aligned}
\int_{\mathbb{U}(\Omega)} p(\omega^U)\mathrm{d}U &= \sum_{kl} \mathrm{Tr}\left[|kl\rangle\langle kl| \left\{ \frac{d_\Omega - p(\rho)}{d_\Omega^3 - d_\Omega} \cdot \mathbb{1}_{\Omega\Omega'} + \frac{p(\rho)d_\Omega - 1}{d_\Omega^3 - d_\Omega} \cdot \mathbb{S}_{\Omega\leftrightarrow\Omega'} \right\} |kl\rangle\langle kl| \mathbb{S}_{ES\leftrightarrow E'S'}\right] \\
&= \sum_{kl} \langle kl| \left\{ \frac{d_\Omega - p(\rho)}{d_\Omega^3 - d_\Omega} \cdot \mathbb{1}_{\Omega\Omega'} + \frac{p(\rho)d_\Omega - 1}{d_\Omega^3 - d_\Omega} \cdot \mathbb{S}_{\Omega\leftrightarrow\Omega'} \right\} |kl\rangle\langle kl| \mathbb{S}_{ES\leftrightarrow E'S'} |kl\rangle \\
&= \sum_{kl} \langle kl| \left\{ \frac{d_\Omega - p(\rho)}{d_\Omega^3 - d_\Omega} \cdot \mathbb{1}_{\Omega\Omega'} + \frac{p(\rho)d_\Omega - 1}{d_\Omega^3 - d_\Omega} \cdot \mathbb{S}_{\Omega\leftrightarrow\Omega'} \right\} |kl\rangle\langle kl| lk\rangle \\
&= \sum_{k} \langle kk| \left\{ \frac{d_\Omega - p(\rho)}{d_\Omega^3 - d_\Omega} \cdot \mathbb{1}_{\Omega\Omega'} + \frac{p(\rho)d_\Omega - 1}{d_\Omega^3 - d_\Omega} \cdot \mathbb{S}_{\Omega\leftrightarrow\Omega'} \right\} |kk\rangle \\
&= \frac{(p(\rho)+1)\cdot(d_\Omega - 1)}{d_\Omega^3 - d_\Omega} \cdot \sum_k \langle kk| \mathbb{1}_{\Omega\Omega'} |kk\rangle \\
&= \frac{p(\rho)+1}{d_\Omega(d_\Omega+1)} \cdot \sum_k \left(\langle k| \mathbb{1}_\Omega |k\rangle\right)^2 \\
&\leq \frac{p(\rho)+1}{d_\Omega+1} \\
&< \frac{2}{d_\Omega} \ . \qquad (6.27)
\end{aligned}$$

The first inequality is due to the fact that $\sum_k \langle k|\mathbb{1}_\Omega|k\rangle = d_\Omega$ and that every summand is at most 1. □

A direct corollary of Theorem 6.1 and Theorem 6.2 is the equilibration in the situations we are physically interested in, namely if we put a system in contact with an environment about which only a few macroscopic properties are known.

Corollary 6.3. *Consider a pure initial state of the system $\rho_S(0) \in \mathcal{S}_=(\mathcal{H}_S)$. For an aribtrary state $\rho_E \in \mathcal{S}_=(\mathcal{H}_{\Omega_E})$ with $\mathcal{H}_{\Omega_E} \subseteq \mathcal{H}_E$, assume that the initial state of the environment is given by $U\rho_E U^\dagger$ with $U \in \mathbb{U}(\Omega_E)$. Let the temporal evolution be governed by a Hamiltonian with non-degenerate energy gaps and denote the time-evolved state of the system by $\rho_S(t)$. Then,*

$$\Pr\left[\langle \|\rho_S(t) - \langle\rho_S(t)\rangle_t\|_1\rangle_t > \sqrt{\frac{d_S}{\sqrt{d_{\Omega_E}}}}\right] < \sqrt{2}\sqrt{\frac{d_S}{\sqrt{d_{\Omega_E}}}} \qquad (6.28)$$

where the probability is computed over the choice of U from the Haar measure on $\mathbb{U}(\Omega_E)$.

Instead of understanding this as a statement about a probability when choosing a unitary U from the Haar measure on $\mathbb{U}(\Omega_E)$, we may equivalently understand it as a statement about a probability when choosing a state $\tilde{\rho}_E = U\rho_E U^\dagger$ with the same eigenvalues as ρ_E and Haar distributed eigenstates from $\mathcal{H}_{\Omega_E}$. That is, we chose $\tilde{\rho}_E$ from the (unique) unitarily invariant measure on an orbit of the action of $\mathbb{U}(\Omega_E)$ on $\mathcal{S}_=(\Omega_E)$. Since $\mathcal{S}_=(\Omega_E)$ is the disjoint union of these orbits, (6.28) also holds if we pick the initial state of the environment $\tilde{\rho}_E$ from any unitarily invariant measure on $\mathcal{S}_=(\mathcal{H}_{\Omega_E})$.

Since the dimensions grow exponentially with the number of particles and we are interested in environments with many constituent particles, we assume $\sqrt{\frac{d_S}{\sqrt{d_{\Omega_E}}}}$ to be very small in physical situations. Thus *any initial state of a system equilibrates for almost any initial state of a large enough environment the system is interacting with.* We emphasize that the strength of this statement is mainly due its very weak assumptions and hence its wide applicability. We did not make any assumptions on the Hamiltonian whatsoever with the sole exception of requiring a Hamiltonian with non-degenerate energy gaps which rules out cases in which system and environment do not interact at all. Nor did we make any assumption about the quantum-mechanical systems we call "system" and "environment".

Proof. Since $\rho_S(0)$ is pure it can be written in the form $\rho_S(0) = |\psi\rangle\langle\psi|_S$. For such a $|\psi\rangle_S$ let $\mathcal{H}_\Omega := |\psi\rangle_S \otimes \mathcal{H}_{\Omega_E}$ so $d_\Omega = d_{\Omega_E}$. Let ω^U_{SE} denote the temporal average obtained from the initial state $\rho_S(0)\otimes U\rho_E U^\dagger$. Applying first (6.5) then the concavity of the square root function and finally Theorem 6.2 we find

$$\begin{aligned}\int_{\mathbb{U}(\Omega_E)} \langle \left\| \rho_S(t) - \omega^U_S \right\|_1 \rangle_t \mathrm{d}U &\leq d_S \int_{\mathbb{U}(\Omega_E)} 2^{-\frac{1}{2}H_2(SE)_{\omega^U}} \mathrm{d}U \\ &\leq d_S \sqrt{\int_{\mathbb{U}(\Omega_E)} 2^{-H_2(SE)_{\omega^U}} \mathrm{d}U} \\ &\leq d_S \sqrt{\frac{2}{d_{\Omega_E}}} \,. \end{aligned} \tag{6.29}$$

Markov's inequality (4.3) then tells us that

$$\Pr_U \left[\langle \left\| \rho_S(t) - \omega^U_S \right\|_1 \rangle_t > \sqrt{\frac{d_S}{\sqrt{d_{\Omega_E}}}} \right] < \sqrt{2} \sqrt{\frac{d_S}{\sqrt{d_{\Omega_E}}}} \tag{6.30}$$

where the probability is computed over the choice of U from the Haar measure. □

Theorem 6.2 generalizes [LPSW09, Theorem 2 (i)] which makes a statement about averages of entropy measures for pure initial states. With only the statement about the averaged purity at hand, we had to apply Markov's inequality in order to obtain the above corollary. [LPSW09, Theorem 2 (ii)] shows that for pure initial states the Rényi entropy of order 2 of the temporal average is not only on average higher than $\log d_\Omega - 1$ but also that the probability of it being smaller than $\log d_\Omega - 2$ is exponentially small. We were not able to generalize this to mixed initial states. In order to obtain an exponentially strong statement by use of Lemma A.3 we would have to upper bound the Lipschitz constant of the function $U \mapsto \operatorname{Tr}\left(\left\langle U\rho U^\dagger\right\rangle_t\right)^2$ which seems difficult. A statement showing that the probability of the entropy of the temporal average being smaller than $\log d_\Omega - 2$ is also for a mixed initial state exponentially small would conclude the generalization of the equilibration results of [LPSW09].

6.3 Initial state independence of the temporal average

6.3.1 The relevant measure of entanglement and the equilibrium state

In this section we deal with the question whether the temporal average of the state of the system is the same for different initial states with support restricted to $\mathcal{H}_\Omega \subseteq \mathcal{H}_S \otimes \mathcal{H}_E$. We will find that this is the case if the uncertainty about the initial state is high enough and/or if the relevant energy eigenstates are sufficiently entangled. In order to quantify the latter, we introdue the quantity $\min_k^\varepsilon H_{\min}(S)_{|E_k\rangle\langle E_k|}$. $H_{\min}(S)_{|E_k\rangle\langle E_k|}$ is a measure of the entanglement of the k-th energy eigenstate. It can be seen by use of the Schmidt decomposition (2.12) that it corresponds to the (negative logarithm of the) maximal overlap of $|E_k\rangle$ with a product state. $\min_k H_{\min}(S)_{|E_k\rangle\langle E_k|}$ therefore quantifies the entanglement of the least entangled energy eigenstate. Energy eigenstates which are (close to) orthogonal to $\mathcal{H}_\Omega$ are irrelevant for the evolution of states with initial support restricted to $\mathcal{H}_\Omega$ and should therefore not be considered in a suitable entanglement measure. We therefore use a "smooth minimum"

$$\min_k^\varepsilon H_{\min}(S)_{|E_k\rangle\langle E_k|} := \max_I \min_{k\notin I} H_{\min}(S)_{|E_k\rangle\langle E_k|} \tag{6.31}$$

where the maximum is over all index sets $I \subseteq \{1, \ldots, d_S d_E\}$ with the property that

$$\sum_{k \in I} \langle E_k | \pi_\Omega | E_k \rangle \leq \varepsilon \ . \tag{6.32}$$

This means that we are allowed to neglect energy eigenstates which are not too important (measured by $\langle E_k | \pi_\Omega | E_k \rangle$) for the evolution of initial states from $\mathcal{H}_\Omega$. The minimum is thus only over energy eigenstates which have significant overlap with the space $\mathcal{H}_\Omega$. Note that since $\sum_k \langle E_k | \pi_\Omega | E_k \rangle = 1$ and since there are $d_S d_E$ terms in the sum, most summands are very small. Hence a small ε suffices to "smoothen away" the vast majority of irrelevant energy eigenstates. In fact, since our entanglement measure is only concerned with the least entangled of the remaining eigenstates, it suffices to smoothen away those eigenstates which are untypically poorly entangled. The maximal value $\langle E_k | \pi_\Omega | E_k \rangle$ can take is $\frac{1}{d_\Omega}$. So if d_Ω is large, we can smoothen away all eigenstates which are untypically poorly entangled and thus interpret the "smooth minimum" as the typical entanglement of the relevant energy eigenstates.

Consider the unitary U_{SE} describing the temporal evolution and acting on density operators by conjugation as in eqn. (2.24). While U_{SE} does in general not leave the space of initial states $\mathcal{H}_\Omega$ invariant, it preserves its linear structure. It makes therefore sense to speak of a time-dependent equiprobable state $\pi_\Omega(t)$. We can thus define a time-averaged equiprobable state

$$\Omega_{SE} := \langle \pi_\Omega(t) \rangle_t = \sum_k |E_k\rangle\langle E_k|\pi_\Omega|E_k\rangle\langle E_k| \ . \tag{6.33}$$

The last expression is invariant under an application of U_{SE} to the state π_Ω, it does therefore not matter which $\pi_\Omega(t)$ we insert.

For any $\rho_{SE} \in \mathcal{S}_=(\mathcal{H}_\Omega)$ we have

$$\begin{aligned}
\Omega_{SE} &= \sum_k |E_k\rangle\langle E_k| \left(\int_{\mathbb{U}(\mathcal{H}_\Omega)} U \rho U^\dagger \mathrm{d}U \right) |E_k\rangle\langle E_k| \\
&= \int_{\mathbb{U}(\mathcal{H}_\Omega)} \sum_k |E_k\rangle\langle E_k| \left(U \rho U^\dagger \right) |E_k\rangle\langle E_k| \mathrm{d}U \\
&= \int_{\mathbb{U}(\mathcal{H}_\Omega)} \langle U \rho U^\dagger \rangle_t \mathrm{d}U \ .
\end{aligned} \tag{6.34}$$

That is, besides understanding Ω_{SE} as the temporal average of the averaged state π_Ω, we may as well understand it as an average over temporal averages of states from $\mathcal{S}_=(\mathcal{H}_\Omega)$.

Our theorem provides a sufficient condition for almost all initial states ρ_{SER} (with support restricted do $\mathcal{H}_\Omega \otimes \mathcal{H}_R$) yielding the equilibrium state Ω_S as their local time-average. If this condition is fulfilled, we also predict the local time-averaged state to be decoupled from a reference which may be correlated with the initial state. That is, we predict that with very high probability

$$I(S:R)_{\langle\rho\rangle_t} \approx 0 \tag{6.35}$$

where

$$\mathrm{Tr}_{ER}\langle\rho\rangle_t \approx \Omega_S \ . \tag{6.36}$$

The condition for this to happen is discussed in Section 6.3.3.

6.3.2 Formal version of the theorem

Formally, we have the following theorem.

Theorem 6.4 (Generalization of [LPSW09, Theorem 3])**.** *Assume $\mathcal{H}_\Omega \subseteq \mathcal{H}_S \otimes \mathcal{H}_E$ and consider an initial state $\rho_{SER} \in \mathcal{S}(\mathcal{H}_\Omega \otimes \mathcal{H}_R)$. The SE-part is subject to a temporal evolution governed by a non-degenerate Hamiltonian H_{SE}. Let $\Omega_S = \mathrm{Tr}_E \Omega_{SE}$ where Ω_{SE} is as defined in eqn. (6.33). Then, for any $\varepsilon > 0$,*

$$\int_{\mathbb{U}(\Omega)} \left\| \mathrm{Tr}_E \langle U\rho U^\dagger\rangle_t - \Omega_S \otimes \rho_R \right\|_1 dU \leq \sqrt{\frac{d_S}{d_\Omega}} 2^{-\frac{1}{2}H^\varepsilon_{\min}(\Omega|R)_\rho - \frac{1}{2}\min_k^{\varepsilon^2/2} H_{\min}(S)_{|E_k\rangle\langle E_k|}} + 12\varepsilon \tag{6.37}$$

where $\min_k^{\varepsilon^2/2}$ denotes a "smooth minimum" as introduced above.

Proof. Our proof is based on Theorem 4.1. We consider the channel $\mathcal{T}$ from Ω to S given by first taking the temporal average and then tracing out the environment E. According to eqn. (2.29) the channel is therefore given by $\mathcal{T}_{\Omega\to S}(\rho_{SE}) = \mathrm{Tr}_E\left(\sum_k |E_k\rangle\langle E_k|\rho|E_k\rangle\langle E_k|\right)$ if the energy-levels of H_{SE} are non-degenerate. Note that this is indeed a CPTPM. By use of the definition of Ω_S and eqn. (4.2) we see that the state τ_S appearing in eqn. (4.1) is given by Ω_S. Theorem 4.1 therefore yields

$$\int_{\mathbb{U}(\Omega)} \left\| \mathrm{Tr}_E \langle U\rho U^\dagger\rangle_t - \Omega_S \otimes \rho_R \right\|_1 dU \leq 2^{-\frac{1}{2}H^\varepsilon_{\min}(\Omega|R)_\rho - \frac{1}{2}H^\varepsilon_{\min}(\Omega'|S)_\tau} + 12\varepsilon \ . \tag{6.38}$$

In order to examine the entropy-term specifying the channel $\mathcal{T}_{\Omega\to S}$ we make use of the chain rule (3.7) and obtain

$$H^\varepsilon_{\min}(\Omega'|S)_\tau \geq H^\varepsilon_{\min}(\Omega' S)_\tau - \log d_S \ . \tag{6.39}$$

The $\Omega' S$-term makes an evaluation of $\tau_{\Omega' S}$ necessary.

$$\begin{aligned}
\tau_{\Omega' S} &= \mathcal{T}_{\Omega \to S}\left(\Psi_{\Omega' \Omega}\right) \\
&= \mathrm{Tr}_E \left(\sum_k |E_k\rangle\langle E_k|_{SE} \Psi_{\Omega' \Omega} |E_k\rangle\langle E_k|_{SE} \right) \\
&= \sum_{i,j,k} \frac{1}{d_\Omega} \langle E_k|_{SE} |i\rangle_\Omega \langle j|_\Omega |E_k\rangle_{SE} |i\rangle\langle j|_{\Omega'} \otimes \mathrm{Tr}_E |E_k\rangle\langle E_k|_{SE} \\
&= \sum_k p_k \rho^{(k)}_{\Omega'} \otimes \mathrm{Tr}_E |E_k\rangle\langle E_k| \qquad (6.40)
\end{aligned}$$

where $p_k = \langle E_k|\pi_\Omega|E_k\rangle$ and

$$\rho^{(k)}_{\Omega'} = \sum_{i,j} \frac{\langle E_k|i\rangle\langle j|E_k\rangle}{\langle E_k|\mathbb{1}_\Omega|E_k\rangle} |i\rangle\langle j|_{\Omega'} \in \mathcal{S}_=\left(\mathcal{H}_{\Omega'}\right) \ . \qquad (6.41)$$

Note that

$$\begin{aligned}
\tau_{\Omega'} &= \sum_k p_k \rho^{(k)}_{\Omega'} \\
&= \frac{1}{d_\Omega} \sum_{i,j,k} \langle j|E_k\rangle\langle E_k|i\rangle|i\rangle\langle j|_{\Omega'} \\
&= \pi_{\Omega'} \ . \qquad (6.42)
\end{aligned}$$

Given an index set $I \subseteq \{1, \ldots, d_S d_E\}$ let

$$\tilde{\tau}_{\Omega' S} := \sum_{k \notin I} p_k \rho^{(k)}_{\Omega'} \otimes \mathrm{Tr}_E |E_k\rangle\langle E_k| \ . \qquad (6.43)$$

Then

$$\begin{aligned}
\tilde{\tau}_{\Omega' S} &\leq \tilde{\tau}_{\Omega'} \otimes \max_{l \notin I} \lambda_{\max}(\mathrm{Tr}_E |E_l\rangle\langle E_l|) \mathbb{1}_S \\
&\leq \tau_{\Omega'} \otimes \max_{l \notin I} \lambda_{\max}(\mathrm{Tr}_E |E_l\rangle\langle E_l|) \mathbb{1}_S \\
&= \pi_{\Omega'} \otimes \max_{l \notin I} \lambda_{\max}(\mathrm{Tr}_E |E_l\rangle\langle E_l|) \mathbb{1}_S \ . \qquad (6.44)
\end{aligned}$$

Since $A \leq B$ implies $\lambda_{\max}(A) \leq \lambda_{\max}(B)$ and since $\lambda_{\max}(A \otimes B) = \lambda_{\max}(A) \otimes \lambda_{\max}(B)$ (6.44) implies

$$\lambda_{\max}(\tilde{\tau}_{\Omega' S}) \leq \frac{1}{d_\Omega} \max_{l \notin I} \lambda_{\max}(\mathrm{Tr}_E |E_l\rangle\langle E_l|) \qquad (6.45)$$

and hence

$$\begin{aligned} H_{\min}^{\varepsilon}(\Omega' S)_{\tau} &\geq H_{\min}(\Omega' S)_{\tilde{\tau}} \\ &\geq \log d_{\Omega} + \min_{l \notin I} H_{\min}(S)_{\mathrm{Tr}_E |E_l\rangle\langle E_l|} \end{aligned} \tag{6.46}$$

as long as the index set I is such that $P(\tau_{\Omega' S}, \tilde{\tau}_{\Omega' S}) \leq \varepsilon$. We note that

$$\begin{aligned} \left\| \tau_{\Omega' S} - \tilde{\tau}_{\Omega' S} \right\|_1 &= \left\| \sum_{k \in I} p_k \rho_{\Omega'}^{(k)} \otimes \mathrm{Tr}_E |E_k\rangle\langle E_k| \right\|_1 \\ &= \sum_{k \in I} p_k \end{aligned} \tag{6.47}$$

so by use of Lemma 3.3

$$P(\tau_{\Omega' S}, \tilde{\tau}_{\Omega' S}) \leq \sqrt{2 \sum_{k \in I} p_k} \ . \tag{6.48}$$

We thus require that the index set I be such that $\sum_{k \in I} p_k \leq \frac{\varepsilon^2}{2}$. Using the notation introduced in Section 6.3.1, we denote the maximization over all such index sets by $\max_I^{\varepsilon^2/2}$. We combine this with (6.39) to obtain

$$H_{\min}^{\varepsilon}(\Omega'|S)_{\tau} \geq \log d_{\Omega} + \max_{I}^{\varepsilon^2/2} \min_{l \notin I} H_{\min}(S)_{\mathrm{Tr}_E |E_l\rangle\langle E_l|} - \log d_S \tag{6.49}$$

which yields the theorem. □

6.3.3 Discussion

Given that the conditions of the corresponding theorems are fulfilled, almost all states in $\mathcal{H}_{\Omega}$ are locally close to π_S^{Ω} (as shown in Theorem 5.1) and almost all states starting their temporal evolution in $\mathcal{H}_{\Omega}$ are locally almost always close to Ω_S (as shown in Theorems 6.1, 6.2 and 6.4). These two states are in general not identical. While π_S^{Ω} depends only on $\mathcal{H}_{\Omega}$, the state Ω_S depends on both $\mathcal{H}_{\Omega}$ and the energy eigenstates. Ω_{SE} is identical to π_{Ω} if the temporal evolution preserves the space $\mathcal{H}_{\Omega}$ which is the case if and only if there is an index set I such that

$$\mathcal{H}_{\Omega} = \mathrm{span}_{\mathbb{C}} \left\{ |E_k\rangle \right\}_{k \in I} \ . \tag{6.50}$$

With the interpretation of $\mathcal{H}_{\Omega}$ given in eqn. (3.59) this is in particular the case if the operators A and H_{SE} commute.

Theorem 6.4 provides the following sufficient condition for almost all states which are only known up to a unitary on a certain space (e.g. all pure states on a certain space) yielding the same time-averaged local state Ω_S (for $\varepsilon \to 0$):

$$\begin{aligned} \log d_S \lessapprox & \underbrace{\log d_\Omega}_{\substack{\text{uncertainty about eigenstates of the initial state}}} \\ & + \underbrace{H_{\min}^{\varepsilon}(\Omega|R)_\rho}_{\substack{\text{uncertainty due to mixture of the initial state}}} \\ & + \underbrace{\min_k{}^{\varepsilon^2} H_{\min}(S)_{|E_k\rangle\langle E_k|}}_{\substack{\text{typical entanglement of the relevant energy eigenstates}}} \end{aligned} \tag{6.51}$$

Roughly speaking, the less we know about the initial state and the more entangled the relevant energy eigenstates are between the system and the environment, the higher is the fraction of initial states which yield approximately the same time-averaged state on the system. We compare this to [LPSW09, Theorem 3] which gives (in our notation) the condition

$$\log d_S \lessapprox \log d_\Omega - \log\left(\sum_k \langle E_k|\pi_\Omega|E_k\rangle 2^{-H_2(S)_{|E_k\rangle\langle E_k|}}\right) \tag{6.52}$$

Note that in [LPSW09, Theorem 3] it is only shown that this condition is sufficient for pure initial states. While it seems not immediately clear which entanglement measure is better suited for which kind of problem, our result is more general in that it allows for mixed initial states and takes their possible correlations to a reference into account.

Here, we want to reproduce two crucial statements derived in [LPSW09]. We consider that we put the system in contact with the environment at $t = 0$ so that the initial state is a product. We first investigate under what conditions the time-averaged state is independent of the initial state of the environment. Let the initial state be a product of a fixed pure state ϕ_S of the system and a generic state ρ_E whose eigenstates are restricted to a subspace $\mathcal{H}_{\Omega_E} \subseteq \mathcal{H}_E$ of the environment. Since we are not interested in our initial correlations with the environment state ρ_E we chose the reference to be trivial in this case. The subspace of the environment describes the set of states following a certain macroscopic constraint like having a given temperature. Note that $\mathcal{H}_{\Omega_E}$ is not a subspace of $\mathcal{H}_S \otimes \mathcal{H}_E$, but $\mathcal{H}_\Omega = |\phi\rangle_S \otimes \mathcal{H}_{\Omega_E}$ is. We have $d_\Omega = d_{\Omega_E}$ and $H_{\min}^{\varepsilon}(\Omega)_{\phi\otimes\rho} = H_{\min}^{\varepsilon}(\Omega_E)_\rho$. Our condition thus turns into

$$\log d_S \lessapprox \log d_{\Omega_E} + H_{\min}^{\varepsilon}(\Omega_E)_\rho + \min_k{}^{\varepsilon^2} H_{\min}(S)_{|E_k\rangle\langle E_k|} \ . \tag{6.53}$$

We have $H_{\min}^{\varepsilon}(\Omega_E)_\rho \geq 0$. Since also $\min_k^{\varepsilon^2} H_{\min}(S)_{|E_k\rangle\langle E_k|} \geq 0$ and since we assume that $\log d_S \lessapprox \log d_{\Omega_E}$ in a physical situation we conclude that *if we put any pure state of a system in contact with an environment, then almost any state of the environment following some macroscopic constraint yields the same time-averaged state of the system.* Similarly as in Corollary 6.3 "almost any" here means "all but exponentially few from any unitarily invariant measure on $\mathcal{S}_=(\mathcal{H}_{\Omega_E})$ according to Theorem 4.2". It will be shown in Chapter 8 that the statement also holds for mixed initial states of the system.

Independence of the initial state of the system turns out to be a much more subtle issue, as already pointed out in Section 4.1. It does not hold in general that almost all initial states of the system yield the same time-averaged state. Trivially, we cannot become independent if there are operators acting on the system which commute with the Hamiltonian thus giving rise to conserved quantities on the system. But even requiring that there be no conserved quantities on the system is not sufficient. This can be seen by looking at an example which is given in [LPSW09]. Consider that the differences between the energy levels of the system Hamiltonian dominate by far over the differences between the energy levels in the environment and in the interaction between system and environment. Then simply due to reasons of energy conservation the joint evolution cannot turn one of the eigenstates of the system Hamiltonian into another. Applying our theorem to this problem generalizes the corresponding result of [LPSW09] to mixed initial states which may initially be correlated to a reference. If we take $\mathcal{H}_\Omega = \mathcal{H}_S \otimes |\phi\rangle_E$ the condition becomes

$$\log d_S \lessapprox \log d_S + H_{\min}^{\varepsilon}(S|R)_\rho + \min_k^{\varepsilon^2} H_{\min}(S)_{|E_k\rangle\langle E_k|} \tag{6.54}$$

so that the dimensional terms cancel. Hence for any pure initial state of the environment almost all states of SR which are connected by unitaries on S yield the same temporal average on S if the conditional entropy about them given access to R is high enough and/or the relevant energy eigenstates are sufficiently entangled.

This result may also be understood in the context of Theorems 5.4 and 5.5. Let S denote a set of qubits with initial correlations to R. Letting the qubits interact for a time which grows polynomially in the number of qubits has according to Theorem 5.4 the same effect as applying a Haar mesure random unitary (as far as we are concerned with the first two moments of the measure). This then approximates according to Theorem 5.5 the predictions of the decoupling theorem which we used to derive Theorem 6.4. Hence, if we first let the isolated qubits interact for a quadratic time and put them consecutively in contact with the environment this allows to translate the above statement from a statement about different initial states of the system to a statement about different qubit evolutions.

Finally, consider an isolated system $\mathcal{H}_S$ with Hamiltonian H_S which is not interacting with any environment E. In this scenario there is no notion of the "entanglement of energy eigenstates". For a non-degenerate Hamiltonian, time-averaging is formally just a measurement in the energy eigenbasis (c.f. (2.29)). The conditional entropy of the Choi-Jamiołkowski isomorphism of an orthogonal measurement is zero (c.f. Table 4.1)). So whether we obtain decoupling and an identical time-average on S for almost all initial states from SR which are connected by unitaries from $\mathbb{U}(S)$ solely depends on the sign of $H_{\min}^{\varepsilon}(S|R)_\rho$. For pure initial states of S, this conditional entropy is zero so that we can neither apply the decoupling theorem nor its converse. We can neither predict that most pure states lead to the same time-average nor the opposite. This is *not* due to the difference between min- and max-entropies which coincide in this case. So in a certain sense time-averaging is decoupling-wise "critical". However, if we add a partial trace over a large part of the system almost all time-averages will be identical on the remaining part.

Chapter 7

On the times which are necessary for entropy changes

In the previous chapters of this thesis we discussed necessary conditions for the thermalization of the system, namely equilibration to the time-averaged state and initial state independence of the time-averaged state. In this chapter we would like to address the time-scales which are necessary in order to change different entropy measures by a given amount. Since the unitary time evolution does not change the entropy of the states on which it acts, entropy changes can only occur if we consider one part of a bipartite system and an interactive Hamiltonian governing the joint evolution of the system.

7.1 The time needed to change Rényi entropies with $\alpha > 1$

For notational convenience we express the following Theorem not in terms of the Rényi entropies but in terms of the *Schatten α-norms* which are closely related to the former. For $\alpha > 0$ we define

$$\|A\|_\alpha := (\operatorname{Tr} |A|^\alpha)^{1/\alpha} \tag{7.1}$$

with $|A| = \sqrt{A^\dagger A}$. For $\rho \in \mathcal{S}_=(\mathcal{H}_A)$ and $\alpha \neq 1$ we have

$$H_\alpha(A)_\rho = \frac{\alpha}{1-\alpha} \log \left(\|\rho_A\|_\alpha \right) \ . \tag{7.2}$$

Theorem 7.1. *Let $\alpha > 1$. Consider a state $\rho_{SE}(0) \in \mathcal{S}_{=}(\mathcal{H}_S \otimes \mathcal{H}_E)$ evolving under a Hamiltonian H_{SE} with interaction strength $\Delta(H_{int})$ as introduced in (2.35). Then for all times t*

$$\left| \frac{d}{dt} \left\| \rho_S(t) \right\|_\alpha \right| \leq \frac{1}{\bar{\bar{T}}} \tag{7.3}$$

where

$$\bar{T}\left(H_{SE}, \rho_{SE}(0)\right) := \left(\min\left\{2\Delta(H_{int}), \left\| [H_{SE}, \rho_{SE}(0)] \right\|_1\right\}\right)^{-1} \ . \tag{7.4}$$

It is worth noting that the above bound is symmetric under an exchange of S and E. In particular it does not matter whether we want to change the entropy of the larger or the smaller part of the joint system SE. We will find the time $\bar{T}$ to be of fundamental importance when lower-bounding the times which are needed for entropy changes. It only depends on the Hamiltonian and the initial state and is diverging if either the system does not interact with the environment or if the initial state commutes with the Hamiltonian. In these cases no changes of the local entropy are possible. In the latter case the initial state does not evolve at all, as can be seen by the von Neumann equation (2.23). We need a long time to change local entropies if the interactive part of the Hamiltonian is weak or if the initial state is close to a mixture of energy eigenstates of the Hamiltonian.

Following the discussion in Section 2.5, Theorem 7.1 may be strengthened if we replace $\Delta\left(H_{int}\right)$ in the definition of $\bar{T}$ by $\Delta\left(\tilde{H}_{int}\right)$ or $\Delta\left(\hat{H}_{int}\right)$, where $\Delta\left(\tilde{H}_{int}\right)$ and $\Delta\left(\hat{H}_{int}\right)$ are as defined in equations (2.37) and (2.40), respectively.

In the limit of $\alpha \to \infty$ we find

$$\left| \frac{\mathrm{d}}{\mathrm{d}t} \lambda_{\max}\left(\rho_S(t)\right) \right| \leq \frac{1}{\bar{\bar{T}}} \ . \tag{7.5}$$

In the special cases of $\alpha = 2$ we find

$$\left| \frac{\mathrm{d}}{\mathrm{d}t} \sqrt{p(\rho_S(t))} \right| \leq \frac{1}{\bar{\bar{T}}} \tag{7.6}$$

or in integrated form a minimal time of

$$\bar{T} \cdot \left| \sqrt{p(i)} - \sqrt{p(f)} \right| \tag{7.7}$$

to change the purity from $p(i)$ to $p(f)$. This improves and corrects an existing bound [Gog10b].[1]

Proof. Up to eqn. (7.11) the proof is due to [RKIA11] and reproduced here for completeness. By use of the von Neumann equation (2.23) we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Tr}_S\left\{\rho_S^\alpha(t)\right\} &= -\,\mathbb{i}\operatorname{Tr}_S\left\{\alpha\rho_S^{\alpha-1}(t)\operatorname{Tr}_E\left[H_{SE},\rho_{SE}(t)\right]\right\}\\ &= -\,\mathbb{i}\,\alpha\operatorname{Tr}_{SE}\left\{\left(\rho_S^{\alpha-1}(t)\otimes\mathbb{1}_E\right)\left[H_{SE},\rho_{SE}(t)\right]\right\}\ . \end{aligned} \tag{7.8}$$

Using the cyclic property of the trace we have

$$\begin{aligned}\operatorname{Tr}_{SE}\left\{\left(\rho_S^{\alpha-1}(t)\otimes\mathbb{1}_E\right)\left[H_S\otimes\mathbb{1}_E,\rho_{SE}(t)\right]\right\} &= \operatorname{Tr}_{SE}\left\{\left[\rho_S^{\alpha-1}(t)\otimes\mathbb{1}_E,H_S\otimes\mathbb{1}_E\right]\rho_{SE}(t)\right\}\\ &= \operatorname{Tr}_S\left\{\left[\rho_S^{\alpha-1}(t),H_S\right]\operatorname{Tr}_E\rho_{SE}(t)\right\}\\ &= \operatorname{Tr}_S\left\{\left[\operatorname{Tr}_E\rho_{SE}(t),\rho_S^{\alpha-1}(t)\right]H_S\right\}\\ &= 0\ . \end{aligned} \tag{7.9}$$

Furthermore,

$$\begin{aligned}\operatorname{Tr}_{SE}\left\{\left(\rho_S^{\alpha-1}(t)\otimes\mathbb{1}_E\right)\left[\mathbb{1}_S\otimes H_E,\rho_{SE}(t)\right]\right\} &= \operatorname{Tr}_{SE}\left\{\left[\rho_S^{\alpha-1}(t)\otimes\mathbb{1}_E,\mathbb{1}_S\otimes H_E\right]\rho_{SE}(t)\right\}\\ &= 0\ . \end{aligned} \tag{7.10}$$

We conclude that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Tr}_S\left\{\rho_S^\alpha(t)\right\} &= -\,\mathbb{i}\,\alpha\operatorname{Tr}_{SE}\left\{\left(\rho_S^{\alpha-1}(t)\otimes\mathbb{1}_E\right)\left[H_{int},\rho_{SE}(t)\right]\right\}\\ &= -\,\mathbb{i}\,\alpha\operatorname{Tr}_{SE}\left\{H_{int}\left[\rho_{SE}(t),\rho_S^{\alpha-1}(t)\otimes\mathbb{1}_E\right]\right\}\ . \end{aligned} \tag{7.11}$$

We introduce the notation $\rho_{cor}(t) := \rho_{SE}(t) - \rho_S(t)\otimes\rho_E(t)$ [GHH07] to find

$$\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Tr}_S\left\{\rho_S^\alpha(t)\right\} = -\,\mathbb{i}\,\alpha\operatorname{Tr}_{SE}\left\{H_{int}\left[\rho_{cor}(t),\rho_S^{\alpha-1}(t)\otimes\mathbb{1}_E\right]\right\}\ . \tag{7.12}$$

We bound the absolute value of this derivative by use of the inequality

$$\left|\operatorname{Tr}(AB)\right| \le \operatorname{Tr}\left|AB\right| = \left\|AB\right\|_1 \le \left\|A\right\|_1\left\|B\right\|_\infty \tag{7.13}$$

[1] Besides minor flaws, in the derivation of (2.6.40) in [Gog10b] the term $\|\rho_{cor}\|_1$ is (combining (2.6.28), (2.6.33) and (2.6.39)) upper-bounded by $2\sqrt{\log d_S}$. This is pointless since the trace-distance $\|\rho_{cor}\|_1$ is upper-bounded by 2 anyway.

and the triangle inequality which yields

$$\begin{aligned}\left|\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Tr}_S\left\{\rho_S^\alpha(t)\right\}\right| &\leq \alpha\left\|H_{int}\right\|_\infty \left\|\left[\rho_{cor}(t),\rho_S^{\alpha-1}(t)\otimes\mathbb{1}_E\right]\right\|_1 \\ &\leq 2\alpha\left\|H_{int}\right\|_\infty \left\|\rho_{cor}(t)\right\|_1 \left\|\rho_S^{\alpha-1}(t)\otimes\mathbb{1}_E\right\|_\infty \\ &= 2\alpha\left\|H_{int}\right\|_\infty \left\|\rho_{cor}(t)\right\|_1 \left\|\rho_S^{\alpha-1}(t)\right\|_\infty \ . \end{aligned} \tag{7.14}$$

We address each of the three norms individually. The term $\left\|H_{int}\right\|_\infty$ can be optimized as follows. It can be seen directly from (7.11) that adding a constant times the identity to H_{int} is irrelevant. Following the discussion in Section 2.5 we can therefore replace the term $\left\|H_{int}\right\|_\infty$ by $\frac{1}{2}\Delta(H_{int})$ where

$$\Delta(H_{int}) := \lambda_{\max}(H_{int}) + \lambda_{\max}(-H_{int}) \ . \tag{7.15}$$

The term $\left\|\rho_{cor}(t)\right\|_1$ is a trace distance and hence upper-bounded by 2. Since $\alpha > 1$ we have

$$\left\|\rho_S^{\alpha-1}(t)\right\|_\infty = \lambda_{\max}(\rho_S)^{\alpha-1} \ . \tag{7.16}$$

Note that this point is the reason why the proof does not work for $0 < \alpha < 1$. For $0 < \alpha < 1$ we would obtain a negative power of the *smallest* non-zero eigenvalue, for which it seems impossible to find a general upper bound. In conclusion,

$$\begin{aligned}\left|\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Tr}_S\left\{\rho_S^\alpha(t)\right\}\right| &\leq 2\alpha\cdot\Delta(H_{int})\cdot\lambda_{\max}(\rho_S)^{\alpha-1} \\ &\leq 2\alpha\cdot\Delta(H_{int})\cdot\left(\operatorname{Tr}_S\left\{\rho_S^\alpha(t)\right\}\right)^{\frac{\alpha-1}{\alpha}} \ . \end{aligned} \tag{7.17}$$

From the last inequality we see that our bound should be the better, the larger α is. Another upper bound for $\left|\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Tr}_S\left\{\rho_S^\alpha(t)\right\}\right|$ can be obtained from eqn. (7.8). We define $U_{SE} := e^{-\mathbb{i}H_{SE}t}$. Since U_{SE} commutes with the Hamiltonian we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Tr}_S\left\{\rho_S^\alpha(t)\right\} &= -\mathbb{i}\,\alpha\operatorname{Tr}_{SE}\left\{\left(\rho_S^{\alpha-1}(t)\otimes\mathbb{1}_E\right)U_{SE}\left[H_{SE},\rho_{SE}(0)\right]U_{SE}^\dagger\right\} \\ &= -\mathbb{i}\,\alpha\operatorname{Tr}_{SE}\left\{U_{SE}^\dagger\left(\rho_S^{\alpha-1}(t)\otimes\mathbb{1}_E\right)U_{SE}\left[H_{SE},\rho_{SE}(0)\right]\right\} \end{aligned} \tag{7.18}$$

we have again by (7.13)

$$\begin{aligned}\left|\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Tr}_S\left\{\rho_S^\alpha(t)\right\}\right| &\leq \alpha\left\|U_{SE}^\dagger\left(\rho_S^{\alpha-1}(t)\otimes\mathbb{1}_E\right)U_{SE}\right\|_\infty \left\|\left[H_{SE},\rho_{SE}(0)\right]\right\|_1 \\ &\leq \alpha\left\|\rho_S^{\alpha-1}(t)\right\|_\infty \left\|\left[H_{SE},\rho_{SE}(0)\right]\right\|_1 \\ &= \alpha\lambda_{\max}\left(\rho_S(t)\right)^{\alpha-1}\left\|\left[H_{SE},\rho_{SE}(0)\right]\right\|_1 \\ &\leq \alpha\cdot\left(\operatorname{Tr}_S\left\{\rho_S^\alpha(t)\right\}\right)^{\frac{\alpha-1}{\alpha}}\cdot\left\|\left[H_{SE},\rho_{SE}(0)\right]\right\|_1 \ . \end{aligned} \tag{7.19}$$

Defining

$$\bar{T}\left(H_{SE},\rho_{SE}(0)\right) := \left(\min\left\{2\Delta(H_{int}), \left\|[H_{SE},\rho_{SE}(0)]\right\|_1\right\}\right)^{-1} \tag{7.20}$$

the differential equations (7.17) and (7.19) can be combined to

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Tr}_S\left\{\rho_S^\alpha(t)\right\}\right| \leq \frac{\alpha}{\bar{T}}\cdot\left(\operatorname{Tr}_S\left\{\rho_S^\alpha(t)\right\}\right)^{\frac{\alpha-1}{\alpha}} \quad . \tag{7.21}$$

By use of Lemma A.9 we obtain that for all times t

$$\operatorname{Tr}_S\left\{\rho_S^\alpha(t)\right\} \in \left[\left(\left(\operatorname{Tr}_S\left\{\rho_S^\alpha(0)\right\}\right)^{\frac{1}{\alpha}} - \frac{t}{\bar{T}}\right)^\alpha, \left(\left(\operatorname{Tr}_S\left\{\rho_S^\alpha(0)\right\}\right)^{\frac{1}{\alpha}} + \frac{t}{\bar{T}}\right)^\alpha\right] \quad . \tag{7.22}$$

The time needed to change from $\operatorname{Tr}_S\left\{\rho_S^\alpha(i)\right\}$ to $\operatorname{Tr}_S\left\{\rho_S^\alpha(f)\right\}$ is therefore at least

$$\bar{T}\cdot\left|\left(\operatorname{Tr}_S\left\{\rho_S^\alpha(i)\right\}\right)^{\frac{1}{\alpha}} - \left(\operatorname{Tr}_S\left\{\rho_S^\alpha(f)\right\}\right)^{\frac{1}{\alpha}}\right| = \bar{T}\cdot\left|\left\|\rho_S(i)\right\|_\alpha - \left\|\rho_S(f)\right\|_\alpha\right| \quad . \tag{7.23}$$

If it were possible to have $\left|\frac{\mathrm{d}}{\mathrm{d}t}\left\|\rho_S(t)\right\|_\alpha\right| > \frac{1}{\bar{T}}$, (7.23) could be violated for an infinitesimally small change of $\frac{\mathrm{d}}{\mathrm{d}t}\left\|\rho_S(t)\right\|_\alpha$, and hence the assertion. □

7.1.1 The time needed for arbitrarily high changes of entropy

Combining eqn. (7.2) and Theorem 7.1 we find that the time which is needed to change a Rényi-entropy with $\alpha > 1$ from $H_\alpha(i)$ to $H_\alpha(f)$ is at least

$$\bar{T}\cdot\left|2^{\frac{1-\alpha}{\alpha}H_\alpha(i)} - 2^{\frac{1-\alpha}{\alpha}H_\alpha(f)}\right| \quad . \tag{7.24}$$

The second factor is always smaller than 1, so this bound allows arbitrarily high changes of entropy to happen in time $\bar{T}$. This may come a bit as a surprise. We will therefore show in this section that (up to a factor of at most π) this is also achievable.

In order to see how arbitrarily high entropies can be achieved in a fixed time, consider the initial state $|00\rangle_{SE} \equiv |0\rangle_S \otimes |0\rangle_E$ which has zero local entropy. We evolve it to $|\psi_r\rangle_{SE} = \frac{1}{\sqrt{r}}\sum_{i=1}^r |i\rangle_S \otimes |i\rangle_E$, the maximally entangled state of rank r. This state has local entropy $\log r$ (for any α). As a brief calculation shows, the Hamiltonian $H_{SE} = E_-|-\rangle\langle -|_{SE}$ with $|-\rangle_{SE} = \frac{1}{\sqrt{2}}|00\rangle_{SE} - \frac{1}{\sqrt{2}}|\psi_r\rangle_{SE}$ evolves $|00\rangle_{SE}$ into $|\psi_r\rangle_{SE}$ in time $\frac{\pi}{E_-}$. This achieves the Margolus-Levitin bound [ML97] for this scenario, which provides a lower bound for the time needed to turn a state into an orthogonal state.

Note, however, that the minimal time needed to achieve an arbitrarily high entropy may be lower than the Margolus-Levitin bound. The necessary time could actually be lowered by choosing one of the states $|i\rangle$ to be identical to $|0\rangle$.

What does our bound give for this case? A brief calculation reveals that

$$[H_{SE}, \rho_{SE}(0)] = \frac{E_-}{2} \left(|00\rangle\langle\psi_r|_{SE} - |\psi_r\rangle\langle 00|_{SE}\right) \tag{7.25}$$

so $\|[H_{SE}, \rho_{SE}(0)]\|_1 = E_-$. As for the $2\Delta(H_{int})$ term, we define $\pi_{X,r} := \frac{1}{r}\sum_{i=1}^{r} |i\rangle\langle i|_X$ and find by use of (2.36)

$$\begin{aligned} H_{int} =& \frac{E_-}{2} \left(|00\rangle\langle 00| + |\psi_r\rangle\langle\psi_r| - |00\rangle\langle\psi_r| - |\psi_r\rangle\langle 00|\right. \\ & \left. -|0\rangle\langle 0| \otimes \pi_E - \pi_{S,r} \otimes \pi_E - \pi_S \otimes |0\rangle\langle 0|_E - \pi_S \otimes \pi_{E,r}\right) \ . \end{aligned} \tag{7.26}$$

We have $\langle\psi_r|H_{int}|\psi_r\rangle = \frac{E_-}{2}\left(1 - \frac{1}{rd_E} - \frac{1}{rd_S}\right)$ and $\langle 22|H_{int}|22\rangle = \frac{E_-}{2}\left(\frac{1}{r} - \frac{1}{rd_E} - \frac{1}{rd_S}\right)$. Since $\Delta(H_{int})$ is according to (2.35) the difference between the largest and the smallest eigenvalue of H_{int} we have $2\Delta(H_{int}) \geq E_-(1 - \frac{1}{r})$ and thus for large r we obtain $\bar{T} = \frac{1}{E_-}$. Hence arbitrarily high entropies can be achieved starting from a pure state in time $\pi\bar{T}$.

7.1.2 The time needed to reach the Boltzmann distribution

An entropy which is particularly interesting in the context of thermalization is the entropy of the Boltzmann state. The time which is needed to reach this entropy gives a lower bound on the thermalization time.

Corollary 7.2. *Consider a quantum mechanical system interacting with an environment such that its equilibrium state follows the Boltzmann distribution. Let E_0 denote the ground state energy of the system, β the inverse temperature of its equilibrium state, $Z(\beta)$ the corresponding partition function and $\Delta(H_{int})$ the strength of the interaction between the system and its environment (as defined in (2.35)). Then the time needed to reach thermal equilibrium starting from any pure state is at least*

$$\frac{1}{2\Delta(H_{int})}\left(1 - \frac{e^{-\beta E_0}}{Z(\beta)}\right) \ . \tag{7.27}$$

This lower bound on the thermalization time holds for any pure initial state of the system and arbitrary initial states of the environment. Note that higher equilibrium temperatures yield a higher lower bound on the thermalization time. A stronger bound can be obtained if we replace $\Delta(H_{int})$ by $\Delta\left(\tilde{H}_{int}\right)$ or $\Delta\left(\hat{H}_{int}\right)$, as discussed in Section 2.5.

Proof. Integrating (7.5) yields that the time which is needed to change the maximal eigenvalue from $\lambda_{\max}(i)$ to $\lambda_{\max}(f)$ is at least

$$\bar{T} \cdot |\lambda_{\max}(i) - \lambda_{\max}(f)| \ .$$

The assertion then follows directly from noting that the maximal eigenvalue of the Boltzmann state $\frac{1}{Z}e^{-\beta H_S}$ is $\frac{1}{Z}e^{-\beta E_0}$ and that by definition $\bar{T} \geq \frac{1}{2\Delta(H_{int})}$. It is easy to see that applying the corresponding bound for another entropy measure H_α with $\alpha > 1$ yields a weaker bound. □

7.2 The time needed to change the von Neumann entropy

We mentioned during the proof of Theorem 7.1 why our theorem does not hold for $\alpha < 1$. In the limit of $\alpha \to 1$, that is for the von Neumann entropy (3.4), Theorem 7.1 becomes trivial. This is why we try another approach for this entropy. We can do a series expansion of the logarithm and apply eqn. (7.11) to each power separately. This yields the equality [RKIA11]

$$\frac{\mathrm{d}}{\mathrm{d}t} H(S)_{\rho(t)} = -\,\mathbb{i}\,\mathrm{Tr}_{SE}\left\{H_{int}\left[\log(\rho_S(t)) \otimes \mathbb{1}_E, \rho_{SE}(t)\right]\right\} \tag{7.28}$$

where the logarithm is only applied to non-zero eigenvalues. From this we see that a state ρ_{SE} has a vanishing derivative of the von Neumann enropy in S under any Hamiltonian H_{SE} if and only if

$$[\rho_S \otimes \mathbb{1}_E, \rho_{SE}] = 0 \ . \tag{7.29}$$

States with this property are called "lazy states" [RKIA11].

The absolute value of the derivative in (7.28) cannot straightforwardly be upper-bounded as it was possible for $\alpha > 1$. The difficulties arise due to the $\log(\rho_S(t))$ term whose entries may have an arbitrarily high absolute value. Since the logarithm is only applied to non-zero eigenvalues, this term is not even stable against small perturbations of ρ. Still, the commutator in (7.28) renders the absolute value of the derivative of the von Neumann entropy finite for any finite-dimensional system S (c.f. [Bra07], for example, a result we will discuss in more detail in this section). We derive here three simple bounds on the rate of change of the local von Neumann entropy an will see later on that the weakest of them is achievable up to a small factor.

Theorem 7.3. *Let the joint evolution of a system S in contact with an environment E be governed by a joint Hamiltonian H_{SE} with interaction strength $\Delta(H_{int})$. Let $\{p_i\}_{i=1,\dots,d_S}$ be the eigenvalues of the state $\rho_S(t)$ of the system at time t. Then the following upper bounds on the absolute value of the temporal derivative of the von Neumann entropy in S apply:*

$$\begin{aligned}\left|\frac{d}{dt}H(S)_{\rho(t)}\right| &\leq \frac{1}{2}\Delta(H_{int})\left\|\sum_{j,k=1}^{d_S}\left(\log p_j - \log p_k\right)\sqrt{p_j}\sqrt{p_k}|j\rangle\langle k|_S\right\|_1 \\ &\leq 2\Delta(H_{int})\sqrt{\sum_{i=1}^{d_S} p_i\left(\log p_i\right)^2} \\ &\leq 2\Delta(H_{int})\log d_S \ . \end{aligned} \tag{7.30}$$

We can learn two immediate things from the first bound. While the entries of $\log\rho_S(t)$ cannot be bounded by any finite value, the expression given in the first bound can. We conclude that the temporal derivative of the von Neumann entropy is finite for all times, which is as we would expect. Furthermore, the temporal derivative of the von Neumann entropy is zero if the distribution described by the p_j's is flat, i.e. if $\rho_S(t)$ is proportional to a projector. Such states are lazy states.

Since the eigenvalues p_i depend on the time t, only the last bound can help us to lower-bound the time we need to change the von Neumann entropy from $H(i)$ to $H(f)$. In contrast to the bounds on the rate of change of the Schatten norms we now have for the first time an explicit dependence on the size of the system. This is no fundamental difference, however, since the entropic quantities which scale with the size of the system are actually the logarithms of the quantities whose derivative we bounded in Section 7.1. In fact, the factor $\log d_S$ in the last bound makes sure that our lower bound on the time which is needed to turn the von Neumann entropy from 0 to $\log d_S$ or vice versa does not depend on the size of S. This is as we would expect following the discussion in Section 7.1.1. Our lower bound for the time which is needed to reach a fully mixed state starting from a pure one is now $\frac{1}{2\Delta(H_{int})}$. This reproduces what we found when dealing with the Rényi entropies with $\alpha > 1$, but we have lost the dependence on the relation between the initial state and the Hamiltonian.

Proof. Since we did not impose any restrictions on the Hamiltonian whatsoever, we can formally extend the environment with a purifying system P and extend the Hamiltonian to $H_{SEP} = H_{SE} \otimes \mathbb{1}_P$. By use of (2.36) we find that H_{int} gets an additional factor $\mathbb{1}_P$ so that the quantities $\|H_{int}\|_\infty$ and $\Delta(H_{int})$ are invariant under this extension.

Let $\rho_{SEP}(t) = |\mu\rangle\langle\mu|_{SEP}$. Then by use of (7.28) and (7.13)

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}H(S)_{\rho(t)}\right| \leq \|H_{int}\|_\infty \, \|[\log(\rho_S(t)) \otimes \mathbb{1}_{EP}, |\mu\rangle\langle\mu|_{SEP}]\|_1 \ . \tag{7.31}$$

It can be seen directly from (7.28) that we can add an arbitrary constant times the identity to H_{int} and thus replace $\|H_{int}\|_\infty$ by $\frac{1}{2}\Delta(H_{int})$. Now let $|\nu\rangle_{S\tilde{P}}$ denote a purification of ρ_S. Since both $|\nu\rangle_{S\tilde{P}}$ and $|\mu\rangle_{SEP}$ are purifications of ρ_S, there is an isometry $V_{\tilde{P}\to EP}$ with $V_{\tilde{P}\to EP}|\nu\rangle_{S\tilde{P}} = |\mu\rangle_{SEP}$. Hence,

$$\begin{aligned}
\left|\frac{\mathrm{d}}{\mathrm{d}t}H(S)_\rho\right| &\leq \frac{1}{2}\Delta(H_{int}) \left\| \left[\log(\rho_S) \otimes \left(V_{\tilde{P}\to EP}\mathbb{1}_{\tilde{P}}V_{\tilde{P}\to EP}^\dagger\right), V_{\tilde{P}\to EP}|\nu\rangle\langle\nu|_{S\tilde{P}}V_{\tilde{P}\to EP}^\dagger\right]\right\|_1 \\
&= \frac{1}{2}\Delta(H_{int}) \left\| V_{\tilde{P}\to EP}\left[\log(\rho_S) \otimes \mathbb{1}_{\tilde{P}}, |\nu\rangle\langle\nu|_{S\tilde{P}}\right] V_{\tilde{P}\to EP}^\dagger \right\|_1 \\
&= \frac{1}{2}\Delta(H_{int}) \, \|[\log(\rho_S) \otimes \mathbb{1}_{\tilde{P}}, |\nu\rangle\langle\nu|_{S\tilde{P}}]\|_1 \ .
\end{aligned} \tag{7.32}$$

The commutator may therefore be calculated for an arbitrary purification $|\nu\rangle_{S\tilde{P}}$ of $\rho_S(t)$. Let $\rho_S(t) = \sum_{i=1}^{d_S} p_i|i\rangle\langle i|_S$ and $|\nu\rangle_{S\tilde{P}} = \sum_{i=1}^{d_S}\sqrt{p_i}|i\rangle_S|i\rangle_{\tilde{P}}$. Then,

$$\begin{aligned}
\|[\log(\rho_S) \otimes \mathbb{1}_{\tilde{P}}, |\nu\rangle\langle\nu|_{S\tilde{P}}]\|_1 &= \left\| \sum_{j,k=1}^{d_S} (\log p_j - \log p_k)\sqrt{p_j}\sqrt{p_k}|j\rangle\langle k|_S \otimes |j\rangle\langle k|_{\tilde{P}} \right\|_1 \\
&= \left\| \sum_{j,k=1}^{d_S} (\log p_j - \log p_k)\sqrt{p_j}\sqrt{p_k}|j\rangle\langle k|_S \right\|_1 \ .
\end{aligned} \tag{7.33}$$

For the second equality we used that the $d_S \cdot (d_S - 1)$ states $|j\rangle_S|k\rangle_{\tilde{P}}$ with $j \neq k$ yield a zero eigenvalue.

The operator $\mathbb{i}\,[\log(\rho_S) \otimes \mathbb{1}_{\tilde{P}}, |\nu\rangle\langle\nu|_{S\tilde{P}}]$ is Hermitian and has vanishing trace, so its eigenvalues are real and sum up to zero. The operator $\Pi_{S\tilde{P}}$ which is the projection

onto the eigenstates with positive eigenvalues therefore allows to write

$$\begin{aligned}
\left\|\mathrm{i}\left[\log(\rho_S)\otimes\mathbb{1}_{\tilde{P}},|\nu\rangle\langle\nu|_{S\tilde{P}}\right]\right\|_1 &= 2\,\mathrm{Tr}\left\{\Pi_{SP}\,\mathrm{i}\left[\log(\rho_S)\otimes\mathbb{1}_{\tilde{P}},|\nu\rangle\langle\nu|_{S\tilde{P}}\right]\Pi_{SP}\right\}\\
&= 2\,\mathrm{i}\,\mathrm{Tr}\left\{\left[\Pi,\log(\rho)\otimes\mathbb{1}\right]|\nu\rangle\langle\nu|\right\}\\
&= 2\,\mathrm{i}\langle\nu|\left[\Pi,\log\rho\otimes\mathbb{1}\right]|\nu\rangle\\
&\leq 4\left|\langle\nu|\Pi\left(\log\rho\otimes\mathbb{1}\right)|\nu\rangle\right|\\
&\leq 4\sqrt{\langle\nu|\Pi\Pi^\dagger|\nu\rangle}\sqrt{\langle\nu|\left(\log\rho\otimes\mathbb{1}\right)^\dagger\left(\log\rho\otimes\mathbb{1}\right)|\nu\rangle}\\
&\leq 4\sqrt{\langle\nu|\left(\log\rho\otimes\mathbb{1}\right)^2|\nu\rangle}\\
&= 4\sqrt{\sum_{i=1}^{d_S}p_i\left(\log p_i\right)^2}\\
&\leq 4\log d_S\ . \qquad (7.34)
\end{aligned}$$

The second inequality is due to an application of Cauchy-Schwarz, the last one can be proved by use of a Lagrange multiplier. □

Our bound agrees well with the result of [Bra07] where it is shown that for pure global states $|\phi\rangle_{SE}$ the optimal rate with which the local entropy can be increased (optimized over pure states *and* Hamiltonians) is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}H(S)_\phi = \|H_{SE}\|_\infty\cdot 2\max_{\frac{1}{2}\leq\lambda\leq 1}\sqrt{\lambda(1-\lambda)}\log\left(\frac{\lambda(d-1)}{1-\lambda}\right) \qquad (7.35)$$

where $d=\min\{d_S,d_E\}$. For $d=2$ the optimal rate is approximately $1.9123\cdot\|H_{SE}\|_\infty$ and for large d it is approximately $\log d\cdot\|H_{SE}\|_\infty$ [Bra07].

Using this result, we can slightly improve the weakest bound in Theorem (7.3) if a specific d_S is given. From (7.28) we know that only the interactive part of the Hamiltonian is relevant for changes of the local entropy, allowing us to replace $\|H_{SE}\|_\infty$ in (7.35) by $\|H_{int}\|_\infty$. Again, it can be seen from (7.28) that we can add an arbitrary constant times the identity to H_{int} and thus replace $\|H_{int}\|_\infty$ by $\frac{1}{2}\Delta(H_{int})$. Multiplying all energy levels by -1 inverts the time-evolution, so an upper bound on the rate with which the entropy can increase which only involves $\Delta(H_{int})$ is also an upper bound on the rate with which it may decrease. Finally, we may add a purifying system P to SE, formally include it into the environment and replace $d=\min\{d_S,d_Ed_P\}$ in (7.35) by d_S. Our adjusted version of (7.35) then becomes

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}H(S)_\rho\right| \leq \Delta(H_{int})\cdot\max_{\frac{1}{2}\leq\lambda\leq 1}\sqrt{\lambda(1-\lambda)}\log\left(\frac{\lambda(d_S-1)}{1-\lambda}\right) \qquad (7.36)$$

which is achievable if $d_S \leq d_E d_P$. This improves the weakest bound in Theorem (7.3) by a factor of approximately 2 for $d_S = 2$ and approximately 4 for large d_S.

An even stronger bound is proved in [CLV03] for the case of a product Hamiltonian $H_{SE} = K_S \otimes K_E$ and a pure global state $|\phi\rangle_{SE}$. The optimal rate with which the entropy can be increased by a Hamiltonian of this form is given by

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} H(S)_\phi &= \frac{\Delta(K_S)\Delta(K_E)}{4} \cdot 2 \max_{0 \leq \lambda \leq 1} \sqrt{\lambda(1-\lambda)} \log\left(\frac{\lambda}{1-\lambda}\right) \\ &\approx \frac{\Delta(K_S)\Delta(K_E)}{4} \cdot 1.9123 \ . \end{aligned} \tag{7.37}$$

Notably, this bound is independent of the dimension of S. So any product Hamiltonian can (up to a change of the time-scale through higher energies) increase local entropies with the same optimal rate as the two-qubit Ising interaction $\sigma_z \otimes \sigma_z$ [CLV03].

As discussed in Section 2.5 we may improve all bounds in this section if we replace $\Delta\left(H_{int}\right)$ by $\Delta\left(\tilde{H}_{int}\right)$ or $\Delta\left(\hat{H}_{int}\right)$. This is because the derivation of eqn. (7.28) works for every decomposition

$$H_{SE} = P_S \otimes \mathbb{1}_E + \mathbb{1}_S \otimes Q_E + \hat{H}_{int} \ . \tag{7.38}$$

7.3 The time needed to change the max-entropy

While we were not able to derive a general lower bound for the time needed to change the entropy H_α for $0 < \alpha < 1$, we obtain a corollary for situations in which there is one eigenvalue which is dominant.

Corollary 7.4. *In the scenario of Theorem 7.1 let $0 < \alpha < 1$. Consider a pure initial state on S. After a time t we have*

$$H_\alpha(S) \leq \frac{1}{1-\alpha} \log\left((1 - \frac{t}{\bar{T}})^\alpha + d_S^{1-\alpha} \cdot \left(\frac{t}{\bar{T}}\right)^\alpha\right) \ . \tag{7.39}$$

We note that again our bound allows to achieve arbitrarily high entropies in time $\bar{T}$.

Proof. Let $\lambda := \lambda_{\max}(\rho_S(t))$. For a fixed λ, the eigenvalues which maximize $H_\alpha(S)$ are given by $\left(\lambda, \frac{1-\lambda}{d_S-1}, \ldots, \frac{1-\lambda}{d_S-1}\right)$. This can be seen by use of a Lagrange multiplier.

Hence

$$H_\alpha(S) \leq \frac{1}{1-\alpha} \log \left(\lambda^\alpha + (d_S - 1) \cdot \left(\frac{1-\lambda}{d_S - 1} \right)^\alpha \right)$$
$$= \frac{1}{1-\alpha} \log \left(\lambda^\alpha + d_S^{1-\alpha} \cdot (1-\lambda)^\alpha \right) \ . \tag{7.40}$$

Integrating (7.5) and using that at $t = 0$ we have by assumption $\lambda_{\max}(\rho_S(0)) = 1$ we find

$$\lambda_{\max}(\rho_S(t)) \geq 1 - \frac{t}{\bar{T}} \ . \tag{7.41}$$

Since (7.40) is (in the relevant parameter interval) monotously decreasing in λ, we may according to (7.41) replace λ in (7.40) by $1 - \frac{t}{\bar{T}}$ and find

$$H_\alpha(S) \leq \frac{1}{1-\alpha} \log \left((1 - \frac{t}{\bar{T}})^\alpha + d_S^{1-\alpha} \cdot \left(\frac{t}{\bar{T}} \right)^\alpha \right) \ . \tag{7.42}$$

□

Since $H_{\max} \equiv H_{\frac{1}{2}}$ (and all α-entropies with $0 \leq \alpha < 1$ for that matter) are constructed so as to be sensible even to the small eigenvalues of the state for which it is calculated it is generally hard to find find upper bounds for it. In the above corollary we had to upper-bound the entropy after a certain time by the "worst-case" scenario in which all the probability weight which is not concentrated in the dominant eigenvalue is uniformly distributed among the remaining eigenvalues. This made the dimensional term in our bound necessary. Note that the above bound diverges proportionally to $\log d_S$ and thus can become arbitrarily large (and arbitrarily larger than $H_{\min}$) for any $\lambda_{\max}(\rho(t)) < 1$, as long as the dimension $\log d_S$ is high enough.

Fortunately the smoothing procedure provides a way out by allowing us to smoothen away the small eigenvalues and thus lose the dependence on the dimension. The price we have to pay for this is that the smoothing parameter becomes time-dependent.

Corollary 7.5. *In the scenario of Theorem 7.1 consider a pure initial state on S. After a time t we have for $\varepsilon \geq \sqrt{2\frac{t}{\bar{T}}}$*

$$H_{\max}^\varepsilon(S)_\rho \leq -\log \frac{1}{1 - t/\bar{T}} = -\frac{1}{\ln 2} \cdot \frac{t}{\bar{T}} + \mathcal{O}\left((\frac{t}{\bar{T}})^2 \right) \ . \tag{7.43}$$

So while we were not able to derive a useful upper bound on the max-entropy of the actual state $\rho_S(t)$, we can find an alternate state which is at most $\sqrt{2\frac{t}{\bar{T}}}$ away (in purified distance) an for which we can show that the max-entropy is even negative.

Proof. It follows directly from the definition of the purified distance (3.33) that a normalized state ρ has a purified distance $\sqrt{1-\lambda_{\max}(\rho)^2}$ to the subnormalized state which only consists of the eigenvalue $\lambda_{\max}(\rho)$ and the projector onto the corresponding eigenstate. Thus,

$$H_{\max}^{\sqrt{1-\lambda_{\max}(\rho_S)^2}}(S)_\rho \leq -\log \frac{1}{\lambda_{\max}(\rho_S)} \ . \tag{7.44}$$

We conclude from (7.41) that

$$\sqrt{1-\lambda_{\max}\left(\rho_S\right)^2} \leq \sqrt{2\frac{t}{\bar{T}}} \ . \tag{7.45}$$

Since a larger smoothing parameter leads to a smaller smooth max-entropy we conclude that

$$H_{\max}^{\sqrt{2t/\bar{T}}}(S)_\rho \leq -\log \frac{1}{1-t/\bar{T}} \ . \tag{7.46}$$

□

Chapter 8

Independence of the initial state of the environment

In contrast to the previous chapters where we have been dealing with temporal averages we will now make statements about the state at a given time t. Such statements are stronger than the ones concerned with temporal averages. If some quantity is small for all times, so is its temporal average, while the converse does in general not hold. In this chapter we are concerned with whether the system is independent of the initial microstate of the environment at any give time. In the following chapter we deal with the analogous question for independence of the initial state of the system.

8.1 Simple version

We have already seen in Section 6.3.3 that if we take the initial state to be a product of a pure state on the system and a state drawn from a subspace of the environment which is much larger than the system (e.g. all states of the environment with a fixed temperature) then the overwhelming majority of environment states will lead to the same time-averaged state. We will show that in this scenario not only the time-averaged state but also the actual state of the system is for all times highly unlikely to depend on the individual initial state of the environment chosen. The evolved state at a given time may therefore solely depend on the Hamiltonian, the macroscopic constraint which is described by the space the environment state is drawn from and the initial state of the system.

We first obtain a statement which is completely independent of the details of the Hamiltonian. Consider that we put a state $\rho_S(0)$ in contact with the environment

which is in a generic state $\rho_E(0)$. The initial state of the environment $\rho_E(0)$ is subject to a constraint (e.g. its temperature is given) which we describe by a subspace $\mathcal{H}_{\Omega_E} \subseteq \mathcal{H}_E$. We let the initial product state evolve and ask ourselves after a time t whether the state of the system depends on the initial state of the environment. Our answer is that this will for almost all initial states of the environment not be the case if the uncertainty about the initial state of the environment is high enough, i.e. if

$$H_{\min}^{\varepsilon}(E)_{\rho} + \log d_{\Omega_E} \gtrapprox 2 \log d_S \ . \tag{8.1}$$

Theorem 8.1. *Consider an initial product state*

$$\rho_S(0) \otimes \rho_E(0) \tag{8.2}$$

where the support of $\rho_E(0)$ is restricted to $\mathcal{H}_{\Omega_E} \subseteq \mathcal{H}_E$. Let $\rho_S\left(\rho_E(0); t\right)$ denote the state of the system at time t when the environment was initially in the state $\rho_E(0)$, so

$$\rho_S\left(\rho_E(0); t\right) = \mathrm{Tr}_E\left[e^{-\mathbb{i} H_{SE} t}\left(\rho_S(0) \otimes \rho_E(0)\right) e^{+\mathbb{i} H_{SE} t}\right] \ . \tag{8.3}$$

Then for all times t there is a state $\tau_S(t)$ such that

$$\int_{\mathbb{U}(\mathcal{H}_{\Omega_E})} \left\| \rho_S\left(U \rho_E(0) U^{\dagger}; t\right) - \tau_S(t) \right\|_1 \mathrm{d}U \leq \frac{d_S}{\sqrt{d_{\Omega_E}}} \cdot 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(E)_{\rho}} + 12\varepsilon \tag{8.4}$$

where the state $\tau_S(t)$ does depend on Ω_E but not on $\rho_E(0)$.

Proof. We apply Theorem 4.1 for the channel describing the dependence $\rho_E(0) \mapsto \rho_S\left(\rho_E(0); t\right)$ which is given by

$$\mathcal{T}_{\Omega_E \to S}(\rho_E(0)) := \mathrm{Tr}_E\left[U_{SE}\left(\rho_S(0) \otimes \rho_E(0)\right) U_{SE}^{\dagger}\right] \tag{8.5}$$

where $U_{SE} = e^{-\mathbb{i} H_{SE} t}$. Since we are not interested in our relation to the initial state of the environment, we take the reference to be trivial and thus have $H_{\min}^{\varepsilon}(\Omega_E|R)_{\rho} = H_{\min}^{\varepsilon}(\Omega_E)_{\rho} \leq \log d_{\Omega_E}$.[1] Theorem 4.1 then predicts

$$\int_{\mathbb{U}(\Omega_E)} ||\mathcal{T}_{\Omega_E \to S}(U \rho_E(0) U^{\dagger}) - \tau_S||_1 \mathrm{d}U \leq 2^{-\frac{1}{2} H_{\min}^{\varepsilon}(\Omega_E)_{\rho(0)} - \frac{1}{2} H_{\min}^{\varepsilon}(\Omega'_E|S)_{\tau}} + 12\varepsilon \tag{8.6}$$

[1] For the case of an environment which is initially correlated to a reference the proof works in exactly the same way as presented here.

where $\tau_{S\Omega'_E} = J\left(\mathcal{T}_{\Omega_E \to S}\right)$. Let ϕ_{SP} be a purification of ρ_S. We define

$$\tau_{SP\Omega_E\Omega'_E} = U_{SE}\left(\phi_{SP} \otimes \Psi_{\Omega_E\Omega'_E}\right) U_{SE}^{\dagger} \tag{8.7}$$

which purifies the Choi-Jamiołkowski isomorphism $\tau_{S\Omega'_E}$ of the channel $\mathcal{T}_{\Omega_E \to S}$. Note that for all times t we have

$$H_{\min}^{\varepsilon}(SP\Omega_E)_\tau \geq \log d_{\Omega_E} \ . \tag{8.8}$$

By use of Lemma 3.7, Lemma 3.4 and the strong subadditivity (3.43) we find

$$\begin{aligned} H_{\min}^{\varepsilon}(\Omega'_E|S)_\tau &\geq H_{\min}^{\varepsilon}(S\Omega'_E)_\tau - \log d_S \\ &= H_{\min}^{\varepsilon}(P\Omega_E)_\tau - \log d_S \\ &\geq H_{\min}^{\varepsilon}(P\Omega_E|S)_\tau - \log d_S \\ &\geq H_{\min}^{\varepsilon}(P\Omega_E S)_\tau - 2\log d_S \\ &= \log d_{\Omega_E} - 2\log d_S \end{aligned} \tag{8.9}$$

which concludes the proof. □

Since the entropy-term in (8.1) is non-negative it follows directly that if the restricted environment is more than twice as large as the system there is no Hamiltonian and no time for which the system depends on the initial microstate of the environment. This holds up to a fraction of initial states of the environment $\rho_E(0)$ which is exponentially small.

Corollary 8.2. *For every initial state of the system $\rho_S(0)$ there is a state $\tau_S(t)$ such that*

$$\Pr_{\rho_E(0)}\left[\left\|\rho_S(t) - \tau_S(t)\right\|_1 > \frac{d_S}{\sqrt{d_{\Omega_E}}} + d_{\Omega_E}^{-1/3}\right] < e^{-d_{\Omega_E}^{1/3}/16} \tag{8.10}$$

where the probability is computed over the choice of the initial state of the environment $\rho_E(0)$ from any unitarily invariant measure on $\mathcal{S}_=(\mathcal{H}_{\Omega_E})$ and where $\tau_S(t)$ does not depend on $\rho_E(0)$.

Proof. This follows directly from reproducing the proof of Theorem 8.1 for the decoupling theorem in the form of Theorem 4.2 and setting $\varepsilon = 0$ and $\delta = d_{\Omega_E}^{-1/3}$. The change from the statement about the choice of a unitary from the Haar measure on $\mathbb{U}(\mathcal{H}_{\Omega_E})$ to a statement about the choice of a state from any unitarily invariant measure on $\mathcal{S}_=(\mathcal{H}_{\Omega_E})$ is as in the discussion after Corollary 6.3. □

The state of the system at any time may therefore only depend on the space $\mathcal{H}_{\Omega_E}$ but not on the eigenvalues and eigenstates of $\rho_E(0)$. In physical terms, it may depend on some macroscopic properties of the environment but not on its detailed microstate. This statement literally holds for all Hamiltonians, all times t and all initial states of the system.

A sufficient condition for only exponentially few initial states on Ω_E leading to a distinct evolved state of the system is according to the above corollary given by

$$\log d_{\Omega_E} \gtrapprox \max\{12, 2\log d_S\} \tag{8.11}$$

which we assume to be the case in physical situations.

Note that the exponentially small fraction of environment states which may lead to a distinct evolved state of the system is not just a technicality. Let $\rho_S(t)$ be a cup of coffee which at $t = 0$ is wonderfully hot and $\mathcal{H}_{\Omega_E}$ denote the space of environment states at room temperature. The overwhelming majority of initial states with support restricted to $\mathcal{H}_{\Omega_E}$ will just cool down the coffee. However, if $\rho_E(0)$ contains a human being with the intention of putting sugar into the coffee, this will lead to a distinct state $\rho_S(t)$. This state $\rho_S(t)$ is distinguishable from $\tau_S(t)$, the cooling unsweetened state of the cup of coffee, with a simple measurement known as tasting. Hence by use of (3.25) $\rho_S(t)$ has a trace distance close to 2 from $\tau_S(t)$. With E containing many moles of particles, the fraction $e^{-d_{\Omega_E}^{1/3}/16}$ is vanishingly small. The example thus shows that "almost all environment states" in a Haar measure/Hilbert space sense may not be the same as what we colloquially understand by it. In fact, most states from the Haar measure on $\mathcal{H}_{\Omega_E}$ do not contain living beings at all.

8.2 Essentially tight version

In the following we restrict ourselves to pure initial states of the system. This allows us to derive a bound for independence of the initial state of the environment which involves the Hamiltonian and the time and is essentially tight. By this we mean that it is tight up to differences between smooth min- and max-entropies and small correction terms. It will show that the behavior of the initial state π_{Ω_E} teaches us a lot about arbitrary initial states with support restricted to $\mathcal{H}_{\Omega_E}$. Let $\tau_{SE}(t)$ be the evolved state of $\tau_{SE}(0) = \phi_S \otimes \pi_{\Omega_E}$. If at a given time we have for $\varepsilon \to 0$

$$H_{\max}^{\varepsilon}(S)_\tau - H_{\min}^{\varepsilon}(E)_\tau \lessapprox 0 \tag{8.12}$$

almost any initial state of the environment will lead to the same state of the system at that time, namely $\tau_S(t)$. This condition is obviously fulfilled for small enough t.

If, on the other hand,

$$H^{\varepsilon}_{\min}(S)_\tau - H^{\varepsilon}_{\max}(E)_\tau \gtrapprox 0 \tag{8.13}$$

there is *cum grano salis* a substantial fraction of initial environment states which lead to different states of the system.

Intuitively, for $t = 0$ the entropy of the state $\tau_{SE}(t)$ is entirely in the environment, $H(E) = \log d_{\Omega_E}$ and $H(S) = 0$. The unitary evolution will leave the entropy of the joint system unchanged for all times, $H(SE) = \log d_{\Omega_E}$. The unitary has two effects. First, it transfers entropy from the environment into the system so that $H(S) \geq 0$. Second, it will construct correlations between the system and the environment so that $H(S) + H(E) \geq H(SE)$. Only if the unitary is at some time specifically such that it can transfer enough entropy from the environment to the system that $H(S) \gtrapprox H(E)$ the initial state of the environment has a significant influence on the state of the system at that time. This is not possible if the conserved entropy $H(SE) = \log d_{\Omega_E}$ is more than twice the maximal value $H(S)$ can take. In this case, (8.12) is fulfilled for all times. This can be seen as follows. If $\log d_{\Omega_E} > 2 \log d_S$ we have by use of (3.43) and (3.7)

$$\begin{aligned} H^{\varepsilon}_{\min}(E)_\tau &\geq H^{\varepsilon}_{\min}(E|S)_\tau \\ &\geq H^{\varepsilon}_{\min}(ES)_\tau - \log d_S \\ &\geq \log d_{\Omega_E} - \log d_S \\ &> \log d_S \\ &\geq H^{\varepsilon}_{\max}(S)_\tau \ . \end{aligned} \tag{8.14}$$

If we are interested in the more general case of mixed initial states of S, we may formally extend S with a purifying system S' and extend the Hamiltonian to $\mathbb{1}_{S'} \otimes H_{SE}$. In this case, the sign of $H(S'S)_\tau - H(E)_\tau$ is decisive for independence of the initial state of E. As can be seen from eqn. (7.8) (with S and E interchanged) for example, the extension of S to SS' does not affect how the entropies in E develop.

Theorem 8.3. *Consider an initial product state $\phi_S \otimes \rho_E(0)$ where the support of $\rho_E(0)$ is restricted to $\mathcal{H}_{\Omega_E} \subseteq \mathcal{H}_E$. Let $\rho_S(\rho_E(0); t)$ denote the evolved state of the system. Then for all times t we have*

$$\begin{aligned} &\int_{\mathbb{U}(\mathcal{H}_{\Omega_E})} \left\| \rho_S\left(U \rho_E(0) U^\dagger; t\right) - \tau_S(t) \right\|_1 \mathrm{d}U \\ &\qquad \leq 2^{-\frac{1}{2} H^{\varepsilon}_{\min}(E)_\rho + \frac{1}{2} H^{\frac{\varepsilon}{2}}_{\max}(S)_\tau - \frac{1}{2} H^{\frac{\varepsilon}{2}}_{\min}(E)_\tau + \log \frac{24}{\varepsilon^2}} + 12\varepsilon \end{aligned} \tag{8.15}$$

where the state $\tau_{SE}(t)$ does not depend on $\rho_E(0)$ and is given by

$$\tau_{SE}(t) = e^{-\mathbb{i}\, H_{SE} t} \left(\phi_S \otimes \pi_{\Omega_E}\right) e^{+\mathbb{i}\, H_{SE} t} \ . \tag{8.16}$$

Conversely, if for any $\varepsilon' > 0$ and ε'', $\varepsilon''' \geq 0$ we have

$$H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(E)_\rho - H_{\min}^{\varepsilon'''}(S)_\tau + H_{\max}^{\varepsilon''}(E)_\tau + \log \frac{2}{\varepsilon'^2} < 0 \tag{8.17}$$

then for every $\omega_S \in \mathcal{S}_=(\mathcal{H}_S)$

$$\int_{\mathbb{U}(\mathcal{H}_{\Omega_E})} \left\| \rho_S \left(U \rho_E(0) U^\dagger; t\right) - \omega_S \right\|_1 \mathrm{d}U > \frac{\varepsilon}{2} \ . \tag{8.18}$$

Proof. The channel describing the dependence $\rho_E(0) \mapsto \rho_S(t)$ is given by

$$\mathcal{T}_{\Omega_E \to S}(\rho_E(0)) := \mathrm{Tr}_E \left[U_{SE} \left(\phi_S \otimes \rho_E(0)\right) U_{SE}^\dagger \right] \tag{8.19}$$

where $U_{SE} = e^{-\mathbb{i}\, H_{SE} t}$. We apply Theorem 4.1 to this channel and thus need to evaluate the entropic terms $H_{\min}^{\varepsilon}(\Omega_E|R)_\rho$ and $H_{\min}^{\varepsilon}(\Omega'_E|S)_\tau$. Since we are not interested in our relation to the initial state of the environment, we take the reference to be trivial and thus have $H_{\min}^{\varepsilon}(\Omega_E|R)_\rho = H_{\min}^{\varepsilon}(\Omega_E)_\rho$. Defining

$$\tau_{\Omega'_E \Omega_E S} = U_{SE} \left(\phi_S \otimes \Psi_{\Omega'_E \Omega_E}\right) U_{SE}^\dagger \tag{8.20}$$

we have by use of Lemma 3.8 and Lemma 3.4 that

$$\begin{aligned} H_{\min}^{\varepsilon}(\Omega'_E|S)_\tau &\geq H_{\min}^{\frac{\varepsilon}{2}}(\Omega'_E S)_\tau - H_{\max}^{\frac{\varepsilon}{2}}(S)_\tau - 2 \cdot \log \frac{24}{\varepsilon^2} \\ &= H_{\min}^{\frac{\varepsilon}{2}}(\Omega_E)_\tau - H_{\max}^{\frac{\varepsilon}{2}}(S)_\tau - 2 \cdot \log \frac{24}{\varepsilon^2} \ . \end{aligned} \tag{8.21}$$

We notice that

$$\begin{aligned} \tau_{\Omega_E S} &= \mathrm{Tr}_{\Omega'_E} \left[U_{SE} \left(\phi_S \otimes \Psi_{\Omega'_E \Omega_E}\right) U_{SE}^\dagger \right] \\ &= U_{SE} \left(\phi_S \otimes \pi_{\Omega_E}\right) U_{SE}^\dagger \end{aligned} \tag{8.22}$$

which concludes the proof of the first part of the theorem. The second part follows from direct application of Theorem 4.4 and using that due to the purity of $\tau_{\Omega'_E \Omega_E S}$ and Lemma 3.4 we have

$$H_{\max}^{\varepsilon''}(\Omega'_E S)_\tau = H_{\max}^{\varepsilon''}(\Omega_E)_\tau \ . \tag{8.23}$$

□

Chapter 9

Independence of the initial state of the system

The problem of deriving rigorous bounds for the time scales which are needed for thermalization has recently been called the most important open problem in the project of justifying statistical physics from first principles of quantum mechanics [LPSW10, HGJ11]. It is our goal to get the solution of this problem a little closer. We recall the four conditions for thermalization which are equilibration, environment state independence, system state independence and Boltzmann form of the equilibrium state. In principal, each of these conditions may serve to derive a lower bound on the equilibration time. In Corollary 7.2 we derived a bound which makes use of the Boltzmann distribution of the equilibrium state. In the previous chapter we saw that independence of the initial state of the environment is fulfilled at $t = 0$ and for all further times which is why this criterion cannot give us a lower bound for the thermalization time. In this chapter we try to find out which times are necessary for a system to become independent of its initial state. More precisely, we discuss the following questions.

- In Sections 9.1.1 and 9.1.2 we consider a given Hamiltonian and a given initial state of the environment and discuss conditions for whether different initial states of the system have already become indistinguishable at some given time.
- In Section 9.1.3 we ask ourselves how long we can guarantee that different pure initial states have not yet evolved to states which are indistinguishable.
- In Section 9.1.4 we find that the times obtaines in Section 9.1.3 can be improved if we look at a large number of systems undergoing an i.i.d. interaction with their respective environment.

- In Section 9.2.2 we find that the times obtained in Section 9.1.3 can be improved if we are willing to neglect an exponentially small fraction of initial states of the environment.
- In Section 9.2.3 we combine the improvements found in Sections 9.1.4 and 9.2.2.
- In Section 9.3 we find sufficient conditions for a system staying close to its initial state for all times (and thus never becoming independent of it).

9.1 For generic initial states of the system and specific initial states of the environment

9.1.1 Informal version

In the last chapter we could show that for any initial state of the system it is highly unlikely to depend on the initial state of the environment at any time in a physical situation where even a restricted environment is dimension-wise much larger than the system. Trivially, an analogue statement for independence of the initial state of the system cannot hold since the system definitely does depend on its initial state for some small enough time. What is more, independence of the initial state is a more subtle issue than independence of the environment for all timescales. We have already seen in Section 6.3 that we can only guarantee that the temporal average of the system is independend of its initial state if the relevant energy eigenstates are sufficiently entangled.

We will find analogously to the last section that the evolution of the initial state π_S teaches us a lot about how generic initial states of the system develop. In order to obtain an essentially tight criterion, we restrict ourselves to pure initial states of the environment. Let the initial state of the environmet be ϕ_E and denote by $\tau_{SE}(t)$ the evolved state of the initial state $\pi_S \otimes \phi_E$. For this state, $H(S)$ is initially maximal (for all entropy measures) while $H(E)$ is zero. For pure initial states of the system, as long as

$$H^{\varepsilon}_{\min}(S)_\tau \gtrapprox H^{\varepsilon}_{\max}(E)_\tau \tag{9.1}$$

the system will not have "forgotten" about its initial state. If at any time

$$H^{\varepsilon}_{\max}(S)_\tau \lessapprox H^{\varepsilon}_{\min}(E)_\tau \tag{9.2}$$

the system will be independent of its initial state.

In the case of mixed initial states ρ_S of the system which may initially be correlated to a reference, we have to add a term $H_{\min}^{\varepsilon}(S|R)_\rho$ to the right hand side of the above conditions. In this case we predict decoupling between S and R if and only if we predict that S has become independent of its initial state. So a system becoming independent of its initial state is closely related to a loss of an observer's initial knowledge about the system.

Again, the above criterion is tight up to differences between smooth min- and max-entropies. In the informal criterion given above the gap between the two entropy measures converts to a critical time-interval for which we do not know whether the system is already independent of its initial state.

Similarly as in the last chapter, we may formally extend the environment with a purifying system E' and the Hamiltonian to $H_{SE} \otimes \mathbb{1}_{E'}$ if we are interested in mixed initial states of the environment. In this case, the sign of $H(S) - H(EE')$ is of interest.

An immediate consequence is that if the system under interest is more than twice as large as the "environment" it is interacting with, it will for all times retain some memory about its (pure) initial state. Using (3.43), Lemma 3.7 and the definition of the smooth max-entropy (3.41) we have for all times

$$\begin{aligned} H_{\min}^{\varepsilon}(S)_\tau - H_{\max}^{\varepsilon}(E)_\tau &\geq H_{\min}^{\varepsilon}(S|E)_\tau - H_{\max}^{\varepsilon}(E)_\tau \\ &\geq H_{\min}^{\varepsilon}(SE)_\tau - \log d_E - H_{\max}^{\varepsilon}(E)_\tau \\ &\geq \log d_S - 2\log d_E \ . \end{aligned} \tag{9.3}$$

9.1.2 Formal version

Theorem 9.1. *Consider an initial state $\rho_{SR}(0)$ which is put in contact with a pure state of the environment ϕ_E and then evolves under a joint Hamiltonian H_{SE}. Let $\mathcal{T}_{S\to S}^t$ denote an "evolution operator" which evolves $\rho_S(0)$ to $\rho_S(t)$. Let $\tau_{SE}(t)$ denote the evolved state of the initial state $\tau_{SE}(0) = \pi_S \otimes \phi_E$. Then,*

$$\begin{aligned} &\int_{\mathbb{U}(S)} ||\mathcal{T}_{S\to S}^t\left(U\rho_{SR}(0)U^\dagger\right) - \mathcal{T}_{S\to S}^t\left(\pi_S\right)\otimes\rho_R||_1 \mathrm{d}U \\ &\quad \leq 2^{-\frac{1}{2}H_{\min}^{\varepsilon}(S|R)_\rho - \frac{1}{2}H_{\min}^{\frac{\varepsilon}{2}}(E)_{\tau(t)} + \frac{1}{2}H_{\max}^{\frac{\varepsilon}{2}}(S)_{\tau(t)} + \log\frac{24}{\varepsilon^2}} + 12\varepsilon \ . \end{aligned} \tag{9.4}$$

Conversely, as long as for any $\varepsilon' > 0$ and ε'', $\varepsilon''' \geq 0$ we have

$$H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(S|R)_\rho + H_{\max}^{\varepsilon''}(E)_{\tau(t)} - H_{\min}^{\varepsilon'''}(S)_{\tau(t)} + \log\frac{2}{\varepsilon'^2} < 0 \tag{9.5}$$

there is no state $\omega_S \in \mathcal{S}_=(\mathcal{H}_S)$ *such that*

$$\int_{\mathbb{U}(S)} \left\| \mathcal{T}^t_{S\to S}\left(U\rho_{SR}(0)U^\dagger\right) - \omega_S \otimes \rho_R \right\|_1 \mathrm{d}U \leq \frac{\varepsilon}{2} \ . \tag{9.6}$$

Proof. Let $U_{SE} := e^{-\mathbb{i}\, H_{SE} t}$. We are interested in the channel

$$\mathcal{T}_{S\to S} : \rho_S \mapsto \mathrm{Tr}_E \left[U_{SE}\left(\rho_S \otimes \phi_E\right) U_{SE}^\dagger \right] \ . \tag{9.7}$$

From Theorem 4.1 we have that

$$\int_{\mathbb{U}(S)} ||\mathcal{T}(U\rho_{SR}U^\dagger) - \tau_S \otimes \rho_R||_1 \mathrm{d}U \leq 2^{-\frac{1}{2}H^\varepsilon_{\min}(S|R)_\rho - \frac{1}{2}H^\varepsilon_{\min}(S'|S)_\tau} + 12\varepsilon \tag{9.8}$$

where $\tau_{SS'}$ is the Choi-Jamiołkowski isomorphism of the channel $\mathcal{T}$ which is purified by the state

$$\tau_{SS'E} = U_{SE}\left(\Psi_{SS'} \otimes \phi_E\right) U_{SE}^\dagger \ . \tag{9.9}$$

By use of Lemma 3.4 and 3.8 we have

$$\begin{aligned} H^\varepsilon_{\min}(S'|S)_\tau &\geq H^{\frac{\varepsilon}{2}}_{\min}(SS')_\tau - H^{\frac{\varepsilon}{2}}_{\max}(S)_\tau - 2\cdot\log\frac{24}{\varepsilon^2} \\ &= H^{\frac{\varepsilon}{2}}_{\min}(E)_\tau - H^{\frac{\varepsilon}{2}}_{\max}(S)_\tau - 2\cdot\log\frac{24}{\varepsilon^2} \end{aligned} \tag{9.10}$$

which proves the first part of the theorem. The second part of the theorem follows directly from the converse decoupling theorem, Theorem 4.4 and applying that again by Lemma 3.4

$$H^{\varepsilon''}_{\max}(S'S)_\tau = H^{\varepsilon''}_{\max}(E)_\tau \ . \tag{9.11}$$

The relevant state for the entropic quantities is therefore

$$\tau_{SE} = U_{SE}\left(\pi_S \otimes \phi_E\right) U_{SE}^\dagger \ . \tag{9.12}$$

□

We recall that any statement about averaged distances obtained from Theorem 4.2 can be converted into a statement about exponentially small probabilities which

is based on Theorem 4.2 and proven analogously. Instead of (9.4) we obtain that for every $\delta > 0$

$$\begin{aligned}
&\Pr_U \left\{ \left\| \mathcal{T}^t_{S\to S}\left(U\rho_{SR}(0)U^\dagger\right) - \mathcal{T}^t_{S\to S}\left(\pi_S\right)\otimes\rho_R \right\|_1 \geq \right. \\
&\qquad \left. 2^{-\frac{1}{2}H^\varepsilon_{\min}(S|R)_\rho - \frac{1}{2}H^{\frac{\varepsilon}{2}}_{\min}(E)_{\tau(t)} + \frac{1}{2}H^{\frac{\varepsilon}{2}}_{\max}(S)_{\tau(t)} + \log\frac{24}{\varepsilon^2}} + 12\varepsilon + \delta \right\} \\
&\leq 2e^{-d_S\delta^2/16}
\end{aligned} \tag{9.13}$$

where the probability is computed over the choice of U from the Haar measure on $\mathbb{U}(S)$.

9.1.3 Quantitative version

We saw in the previous two sections that comparing the local entropies of the state

$$\tau_{SE}(t) = e^{-\mathrm{i}\,H_{SE}t}\left(\pi_S\otimes\phi_E\right)e^{+\mathrm{i}\,H_{SE}t}\ . \tag{9.14}$$

determines whether generic initial states of the system have on average already evolved close to some specific state or not. It is not the case as long as (9.5) is fulfilled. In Chapter 7 we discussed how fast different entropy measures can be changed. We apply these results in this section to investigate how long we can guarantee that (9.5) is fulfilled and thus how long we can guarantee that different initial states have on average not yet evolved to the inside of a ball with a certain radius in trace distance. The times we obtain this way only depend on the radius, the Hamiltonian and the initial state of the environment. These times provide lower bounds on the thermalization time. We discuss the limits of a sufficiently large system and of a small radius separately.

Theorem 9.2. *Consider a system S which at $t = 0$ is put in contact with an environment which is in a pure state ϕ_E. Let the joint evolution be governed by a Hamiltonian H_{SE}. Let $\bar{T} = \bar{T}\left(H_{SE}, \pi_S\otimes\phi_E\right)$ as introduced in (7.4). Let $\rho^\phi_S(t)$ denote the state of the system at time t given that its state at $t = 0$ was ϕ_S.*

For a system S with $\underline{\log d_S \gtrapprox 10}$ for

$$t < \begin{cases} 9.5\cdot 10^{-3}\cdot\bar{T} \\ 1.9\cdot 10^{-2}\cdot\bar{T} \\ 2.2\cdot 10^{-2}\cdot\bar{T} \\ 2.4\cdot 10^{-2}\cdot\bar{T} \end{cases} \tag{9.15}$$

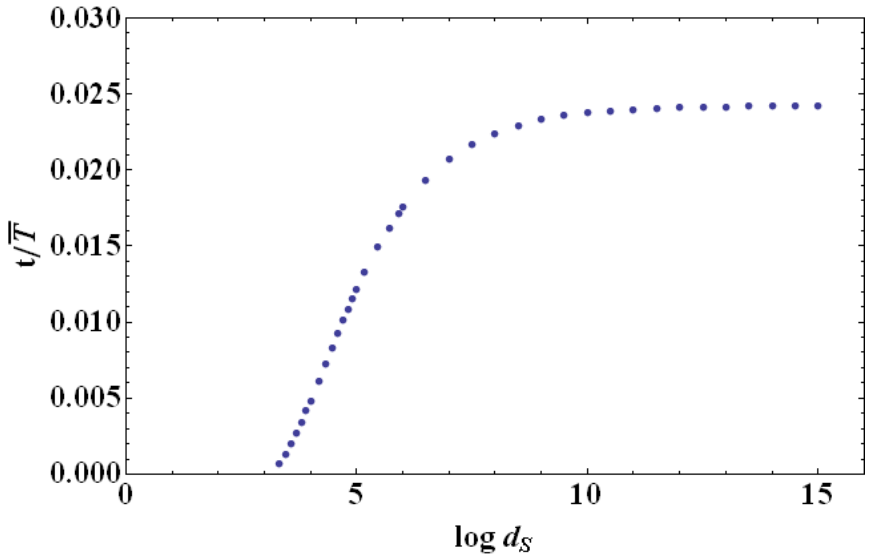

Figure 9.1: The times for which we can guarantee that different pure initial states of S have not yet evolved to the same state as a function of $\log d_S$. The times are measured in units of $\bar{T}\left(H_{SE}, \pi_S \otimes \phi_E\right)$.

there is no state $\omega_S \in \mathcal{S}_=(\mathcal{H}_S)$ *such that*

$$\left\langle \left\| \rho_S^{\phi}(t) - \omega_S \right\|_1 \right\rangle_{\phi_S} < \frac{\varepsilon}{2} \tag{9.16}$$

where

$$\begin{cases} \varepsilon = 10^{-1} \\ \varepsilon = 10^{-2} \\ \varepsilon = 10^{-3} \\ \varepsilon \to 0 \end{cases} . \tag{9.17}$$

For a specific ϕ_E, H_{SE} *and* ω_S *and large enough* d_S *the l.h.s. of (9.16) can be made arbitrarily small in time* $\pi\bar{T}$.

In the limit of $\varepsilon \to 0$ *the times for which we can guarantee that there is no state* $\omega_S \in \mathcal{S}_=(\mathcal{H}_S)$ *which fulfills (9.16) are given in Figure 9.1 as a function of* $\log d_S$.

Our lower bounds are therefore (for small ε and large d_S) achievable up to factor of order of magnitude 10^2.

Combining the above theorem with Corollary 7.2 we find our final lower bound on the equilibration time. It is given by

$$\max\left\{\frac{1}{2\Delta(H_{int})}\left(1-\frac{e^{-\beta E_0}}{Z(\beta)}\right), \frac{c}{2\Delta(H_{int})}, \frac{c}{\|[H_{SE}, \pi_S\otimes\phi_E]\|_1}\right\} \tag{9.18}$$

with a constant $c \approx 2\cdot 10^{-2}$. This bound holds for all pure initial states from a quantum mechanical system S with $\log d_S \gtrapprox 7$ (c.f. Figure 9.1). It only depends on the Hamiltonian H_{SE} and on the initial state of the environment ϕ_E. We think of the inverse equilibrium temperature β as being a function of these two operators. The first of the three terms in our bound achieves the maximum unless one of the following two cases applies:

- The system is in equilibrium with high probability in its ground state, i.e. the equilibrium temperature is much lower than the gap between the system's ground state and its first excited state. In this case, the first term becomes smaller than the second one.

- $\pi_S\otimes\phi_E$ is close to commuting with H_{SE}, i.e. all partial traces of energy eigenstates are either close to ϕ_E or have eigenstates which are close to orthogonal to it. In this case, the third term becomes larger than the first one.

The proof of the above theorem and its discussion constitute the remainder of this section.

According to Theorem 9.1 as long as

$$H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(S|R)_\rho + H_{\max}^{\varepsilon''}(E)_\tau - H_{\min}^{\varepsilon'''}(S)_\tau + \log\frac{2}{\varepsilon'^2} < 0 \tag{9.19}$$

holds the system is not yet $\frac{\varepsilon}{2}$-independent of its initial state. A first difficulty arises due to the term

$$H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(S|R)_\rho\ . \tag{9.20}$$

In order to make the dependence on the epsilons analytically manageable, we resctrict ourselves to pure initial states of the system which allows to apply Lemma A.10. This is not a severe restriction since pure initial states minimize the conditional entropy term (as long as there is no quantum mechanical entanglement between initial state and reference) and thus remain distinguishable for the longest time anyway.

The state $\tau_{SE}(t) = e^{-\mathbb{i}H_{SE}t}(\pi_S\otimes\phi_E)e^{+\mathbb{i}H_{SE}t}$ starts with maximal entropy in S and zero entropy in E. In the previous chapter we derived results which answer exactly the question how fast different entropy measures can be changed. For example,

we have shown that the maximal rate with which $\lambda_{\max}$ can be changed is a quantity we called $\frac{1}{\overline{T}}$. In our case it is given by

$$\frac{1}{\overline{T}} = \min\left\{2\Delta(H_{int}), \|[H_{SE}, \pi_S \otimes \phi_E]\|_1\right\} \ . \tag{9.21}$$

Integrating (7.5) and using that $\lambda_{\max}\left(\tau_S(0)\right) = \frac{1}{d_S}$ we find

$$\lambda_{\max}\left(\tau_S(t)\right) \leq \frac{1}{d_S} + \frac{t}{\overline{T}} \ . \tag{9.22}$$

Setting $\varepsilon''' = 0$ we have

$$H_{\min}^{\varepsilon'''}(S)_\tau \geq -\log\left(\frac{1}{d_S} + \frac{t}{\overline{T}}\right) \ . \tag{9.23}$$

Note that setting $\varepsilon''' = 0$ is the only possibility to make the term $H_{\min}^{\varepsilon'''}(S)_\tau$ analytically manageable since we have no results concerning the evolution of the second largest eigenvalue of τ_S.

A much deeper problem is that we have to upper bound the term $H_{\max}^{\varepsilon''}(E)_\tau$. In Corollaries 7.4 and 7.5 we derived bounds for how fast $H_{\max}$ can increase if it is initially zero which is precisely what we are interested in. Corollary 7.4 only highlights the difficulties: it depends on d_E which we actually do not want in our bound and, even worse, diverges proportionally to its logarithm and thus becomes useless in physical situations with a large environment. This is why we have to apply Corollary 7.5 involving a time-dependent smoothing parameter $\varepsilon'' = \sqrt{2\frac{t}{\overline{T}}}$, which then also appears in the conditional entropy term in (9.19).

Since we have now bounded all three entropic terms in (9.19) in the desired direction we put all this together and find that the system is not yet $\frac{\varepsilon}{2}$-independent of its initial state as long as

$$-\log\left(1 - \left(\varepsilon' + 2\sqrt{2\frac{t}{\overline{T}}} + \sqrt{\varepsilon}\right)^2\right) + \log\left(1 - \frac{t}{\overline{T}}\right) + \log\left(\frac{1}{d_S} + \frac{t}{\overline{T}}\right) + \log\frac{2}{\varepsilon'^2} < 0 \ . \tag{9.24}$$

The remaining free parameter ε' can be chosen so as to optimize the condition. This allows us to numerically find the time for which we can guarantee that different pure initial states of the system have on average not yet evolved to the inside of a ball with radius $\frac{\varepsilon}{2}$.

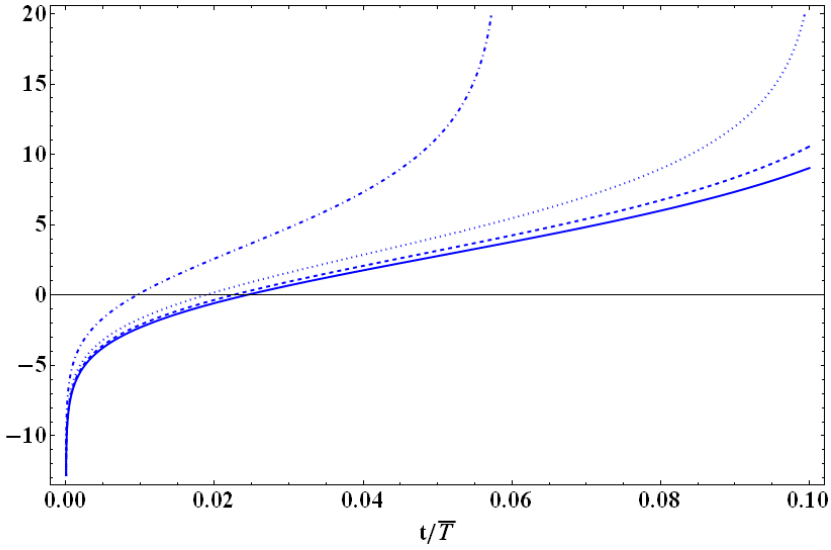

Figure 9.2: The left hand side of (9.24) as a function of t (measured in units of $\bar{T}$) in the limit of large d_S for various values of ε. For $\varepsilon = 10^{-1}$ (dashed-dotted line) we can guarantee that the system is not yet independent of its initial state up to time $9.5 \cdot 10^{-3} \cdot \bar{T}$, for $\varepsilon = 10^{-2} \cdot \bar{T}$ (dotted line) up to time $1.9 \cdot 10^{-2} \cdot \bar{T}$, for $\varepsilon = 10^{-3}$ (dahed line) up to time $2.2 \cdot 10^{-2}$ and in the limit $\varepsilon \to 0$ (bold line) up to time $2.4 \cdot 10^{-2} \cdot \bar{T}$. As discussed we used the values $\varepsilon'' = \sqrt{2\frac{t}{\bar{T}}}$, $\varepsilon''' = 0$ and the optimal value for ε'.

This time only depends on ε, d_S and $\bar{T}$, and through $\bar{T}$ on the initial state of the environment ϕ_E and on the Hamiltonian. The dependence on ε and d_S is negligible for small ε and large d_S. There is an implicit dependence on d_S through the terms in $\bar{T}$, however. The resulting times for which we can guarantee that the system still "memorizes" its initial state are of the order $2 \cdot 10^{-2} \cdot \bar{T}$ for small ε and large d_S, as illustrated in Figure 9.2.

Achievability

We now investigate the question of how tight the lower bounds in Theorem 9.2 are. That is, how fast can we achieve

$$H_{\max}^{\varepsilon}(S)_{\tau(t)} \lessapprox H_{\min}^{\varepsilon}(E)_{\tau(t)} \tag{9.25}$$

which by Theorem 9.1 allows to predict that almost all pure initial states of S have evolved to $\tau_S(t)$. To this end, we analyze a particular combination of a Hamiltonian

H_{SE} and initial state of the environment ϕ_E. We will find that the bound $2 \cdot 10^{-2} \cdot \bar{T}$ is tight up to a factor of approximately 10^2.

As we have seen in Theorem 9.1, the initial state $\tau_{SE}(0) = \pi_S \otimes |1\rangle\langle 1|_E$ determines widely the behavior of generic initial states of the form $\phi_S \otimes |1\rangle\langle 1|_E$. For this initial state we consider an environment which is a copy of the system and an evolution governed by the Hamiltonian

$$H_{SE} = \sum_{i=2}^{d_S} |E_i\rangle\langle E_i| \tag{9.26}$$

with

$$|E_i\rangle = \frac{1}{\sqrt{2}}|i\rangle_S \otimes |1\rangle_E - \frac{1}{\sqrt{2}}|1\rangle_S \otimes |i\rangle_E \ . \tag{9.27}$$

A straightforward calculation shows that the effect of this Hamiltonian is to interchange the contents of S and E in time π with the exception of the first coefficient. Formally,

$$\tau_S(t) = \frac{1-\cos(t)}{2}|1\rangle\langle 1|_S + \frac{1+\cos(t)}{2}\pi_S \tag{9.28}$$

and

$$\tau_E(t) = \frac{1+\cos(t)}{2}|1\rangle\langle 1|_E + \frac{1-\cos(t)}{2}\pi_E \ . \tag{9.29}$$

This seems not too far from optimal in order to reach the condition 9.25 which allows us to guarantee initial state independence.[1]

In order to calculate the quantity $\bar{T}$ for this combination of Hamiltonian and initial state we first calculate by use of (2.36)

$$H_{int} = \sum_{i=2}^{d_S} \left(|E_i\rangle\langle E_i| - \frac{1}{2}\left(|i\rangle\langle i|_S + |1\rangle\langle 1|_S\right) \otimes \pi_E - \pi_S \otimes \frac{1}{2}\left(|1\rangle\langle 1|_E + |i\rangle\langle i|_E\right) \right) \ . \tag{9.30}$$

[1] Obviously, this Hamiltonian does not thermalize the system since it does not lead to equilibration. In order to make it thermalizing we could add many more copies of the system S to the environment so that $E = E_1 \dots E_N$ and $\mathcal{H}_S \cong \mathcal{H}_{E_i}$. One could then extend the Hamiltonian in such a manner that it does not move the contens of E_1 back to S but further to E_2 and so on and finally from E_N back to S. Choosing a large N one could make the fraction of time for which S is not close to its temporal average arbitrarily small.

Since

$$\langle 11|H_{int}|11\rangle_{SE} = -1 + \frac{1}{d_S} \tag{9.31}$$

and

$$\langle E_2|H_{int}|E_2\rangle_{SE} = 1 - \frac{1}{d_S} \tag{9.32}$$

we find

$$\Delta(H_{int}) \geq 2(1 - \frac{1}{d_S}) \ . \tag{9.33}$$

Furthermore,

$$\begin{aligned}\lambda_{\max}(H_{int}) &\leq \lambda_{\max}\left(\sum_{i=2}^{d_S} |E_i\rangle\langle E_i|\right) + \lambda_{\max}\left(\sum_{i=2}^{d_S} -\frac{1}{2}\left(|i\rangle\langle i|_S + |1\rangle\langle 1|_S\right)\otimes \pi_E\right) \\ &\quad + \lambda_{\max}\left(\sum_{i=2}^{d_S} -\pi_S \otimes \frac{1}{2}\left(|1\rangle\langle 1|_E + |i\rangle\langle i|_E\right)\right) \\ &= 1 - \frac{1}{2d_S} - \frac{1}{2d_S}\end{aligned} \tag{9.34}$$

and

$$\begin{aligned}\lambda_{\max}(-H_{int}) &\leq \lambda_{\max}\left(\sum_{i=2}^{d_S} -|E_i\rangle\langle E_i|\right) + \lambda_{\max}\left(\sum_{i=2}^{d_S} \frac{1}{2}\left(|i\rangle\langle i|_S + |1\rangle\langle 1|_S\right)\otimes \pi_E\right) \\ &\quad + \lambda_{\max}\left(\sum_{i=2}^{d_S} \pi_S \otimes \frac{1}{2}\left(|1\rangle\langle 1|_E + |i\rangle\langle i|_E\right)\right) \\ &= 0 + \frac{1}{2}\frac{d_S - 1}{d_S} + \frac{1}{2}\frac{d_S - 1}{d_S}\end{aligned} \tag{9.35}$$

so

$$\lambda_{\max}(H_{int}) + \lambda_{\max}(-H_{int}) \leq 2(1 - \frac{1}{d_S}) \ . \tag{9.36}$$

We conclude that

$$\Delta(H_{int}) = 2(1 - \frac{1}{d_S}) \ . \tag{9.37}$$

This expression for the "interaction strength" can (for $d_S \geq 2$) be improved by calculating $\Delta\left(\tilde{H}_{int}\right)$ as defined in (2.37) instead of $\Delta\left(H_{int}\right)$. We simply find that $\tilde{H}_{int} = H_{SE}$ so that

$$\Delta\left(\tilde{H}_{int}\right) = 1 \ . \tag{9.38}$$

So, as in the example in Section 2.5, the "interaction strength" obtained from $\Delta\left(\tilde{H}_{int}\right)$ is not only lower but also simpler.

We will see that in order to calculate $\bar{T}$ it does not matter whether we use $\Delta\left(H_{int}\right)$ or $\Delta\left(\tilde{H}_{int}\right)$. However, it *will* make a difference in Section 9.1.4.

A brief calculation shows that

$$[H_{SE}, \tau_{SE}(0)] = \frac{1}{2d_S} \cdot \sum_{i=2}^{d_S} |i1\rangle\langle 1i| - |1i\rangle\langle i1| \ . \tag{9.39}$$

The operator $\sum_{i=2}^{d_S} |i1\rangle\langle 1i| - |1i\rangle\langle i1|$ has $(d_S-1)^2+1$ eigenstates $|ij\rangle$ with $i \neq 1 \neq j$ or $i = 1 = j$ which yield a zero eigenvalue. Their orthogonal complement has dimension $2(d_S-1)$. The states $|1i\rangle \pm \mathbb{i}\,|i1\rangle$ with $2 \leq i \leq d_S$ form an orthonormal basis for this orthogonal complement and yield eigenvalues $\pm\,\mathbb{i}$. So

$$\left\| \sum_{i=2}^{d_S} |i1\rangle\langle 1i| - |1i\rangle\langle i1| \right\|_1 = 2(d_S - 1) \tag{9.40}$$

and

$$\bar{T} = \left(\min\left\{2\Delta(H_{int}), \|[H_{SE}, \tau_{SE}(0)]\|_1\right\}\right)^{-1} = \frac{d_S}{d_S - 1} \tag{9.41}$$

for this combination of Hamiltonian and initial state.

For large d_S and a small smoothing parameter we need a time close to π to fulfill (9.25) and thus guarantee initial state indepence, as is illustrated in Figure 9.4. The formally precise condition for initial state independence, however, is that the r.h.s. of (9.4) be small. Since we deal with pure initial states we use $H_{\min}^{\varepsilon}(S|R)_\rho \geq 0$. We evaluate the r.h.s. of (9.4) at $t = \pi$. According to (9.28) and (9.29) we have $\tau_S(\pi) = |1\rangle\langle 1|_S$ and $\tau_E(\pi) = \pi_E$. We find therefore for the r.h.s. of (9.4)

$$2^{-\frac{1}{2}H_{\min}^{\varepsilon}(S|R)_\rho - \frac{1}{2}H_{\min}^{\frac{\varepsilon}{2}}(E)_{\tau(t)} + \frac{1}{2}H_{\max}^{\frac{\varepsilon}{2}}(S)_{\tau(t)} + \log\frac{24}{\varepsilon^2}} + 12\varepsilon \leq \frac{1}{\sqrt{d_S}} \cdot \frac{24}{\varepsilon^2} + 12\varepsilon \ . \tag{9.42}$$

For $\varepsilon = d_S^{-1/6}$ this becomes $36 d_S^{-1/6}$ which becomes arbitrarily small for large enough d_S. Hence for large enough d_S it is possible that all but exponentially few pure initial states of S evolve to the same state in time $\pi = \pi \cdot \frac{d_S - 1}{d_S} \cdot \bar{T}$.

Theorem 9.2 guarantees that we are not yet independent of the initial state up to times of order $2 \cdot 10^{-2} \cdot \bar{T}$ for large enough d_S. We conclude that up to a factor of at most 10^2 our lower bound on the time which is necessary for initial state independence is also achievable.

The gap which is expressed by the factor 10^2 is not only due to the estimates we applied but fundamentally due to the differences between min- and max entropies and thus due to the basic techniques on which our attempt is based. This is illustrated in the limit of a vanishing smoothing parameter ε in Figures 9.3 and 9.4. A non-zero smoothing parameter allows to reduce the gap between the two entropy measures by decreasing untypically large eigenvalues (in the case of $H_{\min}$) or by ignoring untypically small eigenvalues (in the case of $H_{\max}$). Because we were restricted to smoothing parameters which render the entropic quantities analytically evaluable, we were in general not able to chose the optimal smoothing parameters. Consequently, our best attempt above still led to results with a gap of two orders of magnitude between times for which we can guarantee that the system still "remembers" its initial state and times for which we can show that it has "forgotten" about it.

So far in this section we have been interested in finding times for which we can guarantee that $H_{\min}^{\varepsilon}(S)_\tau \gtrapprox H_{\max}^{\varepsilon}(E)_\tau$. We know that the system has *not yet* become independent of its initial state (or, more generally, not yet thermalized) as long as this is fulfilled. In Chapter 6 we saw that sufficient entanglement of the relevant energy eigenstates will eventually lead the system to independence of its initial state, but did not make any predictions about which times are sufficient for this. As we saw in this section, the condition $H_{\max}^{\varepsilon}(S)_\tau \lessapprox H_{\min}^{\varepsilon}(E)_\tau$ provides in principle times which are sufficient. It seems outside the scope of our techniques to obtain a general formula for times which are sufficient to reach this condition which only depends on the initial state and the Hamiltonian.

9.1.4 The erasure time of a quantum memory

In an i.i.d. scenario the gap between smooth min- and max-entropies can be closed according to Theorem 3.5 which allows to replace the smooth entropies by the von Neumann entropy. A little more work is required to rigorously prove the following theorem. It states that if we apply an i.i.d. channel $\mathcal{T}_{A\to B}^{\otimes n}$ to a pure state of a tensor product space $\mathcal{H}_A^{\otimes n}$, then for large enough n the sign of $H(A|B)_\tau$ provides a tight

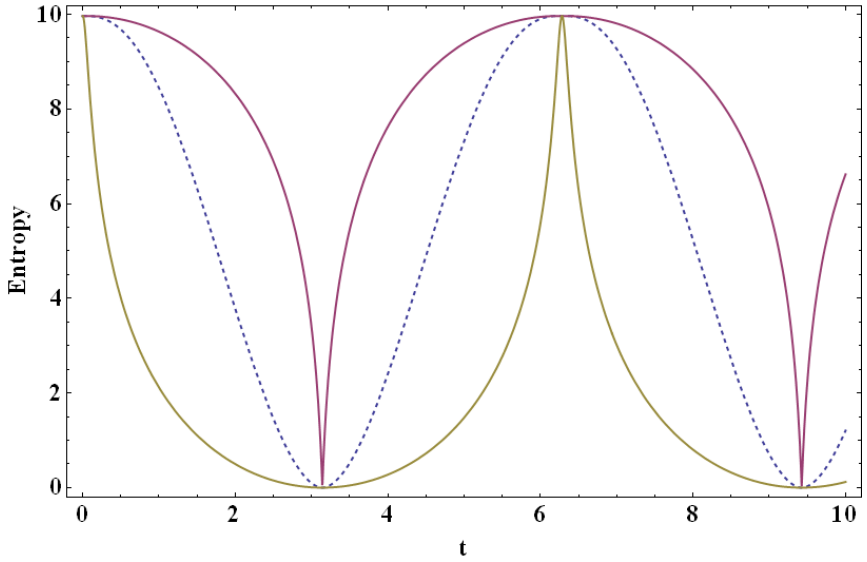

Figure 9.3: $H_{\min}^{\varepsilon}(S)_{\tau(t)}$ and $H_{\max}^{\varepsilon}(S)_{\tau(t)}$ in the limit $\varepsilon \to 0$ for $d_S = 10^3$. The maximal difference between $H_{\min}(S)$ and $H_{\max}(S)$ grows like $\log d_S - 1$ and can thus become arbitrarily large. The dashed line shows the von Neumann entropy $H_1(S)_{\tau(t)}$.

criterion for whether different input states yield the same output. For large enough n, if $H(A|B)_\tau > 0$ almost all states $|\phi\rangle_{A^n} \in \mathcal{H}_A^{\otimes n}$ yield the same output. This is not the case if $H(A|B)_\tau < 0$.

Theorem 9.3. *Let $\mathcal{T}_{A\to B}$ be a CPTPM with Choi-Jamiołkowski representation $\tau_{A'B} = J(\mathcal{T})$. Then*

$$\begin{aligned}\Pr_{\phi}\left\{\left\|\mathcal{T}^{\otimes n}\left(\phi_{A^n}\right) - \tau_B^{\otimes n}\right\|_1 > \exp\left(-n\cdot\frac{\ln 2}{2}\cdot\left(H(A|B)_\tau - c\right)\right) + \frac{1}{n} + d_A^{-n/3}\right\} \\ \leq 2\exp\left(-d_A^{n/3}/16\right)\end{aligned} \tag{9.43}$$

where the probability is computed over the choice of ϕ_{A^n} from the Haar measure on $\mathcal{H}_A^{\otimes n}$. The constant c is

$$c \in \mathcal{O}\left(\sqrt{\frac{\log n}{n}}\cdot\log d_A\right) \ . \tag{9.44}$$

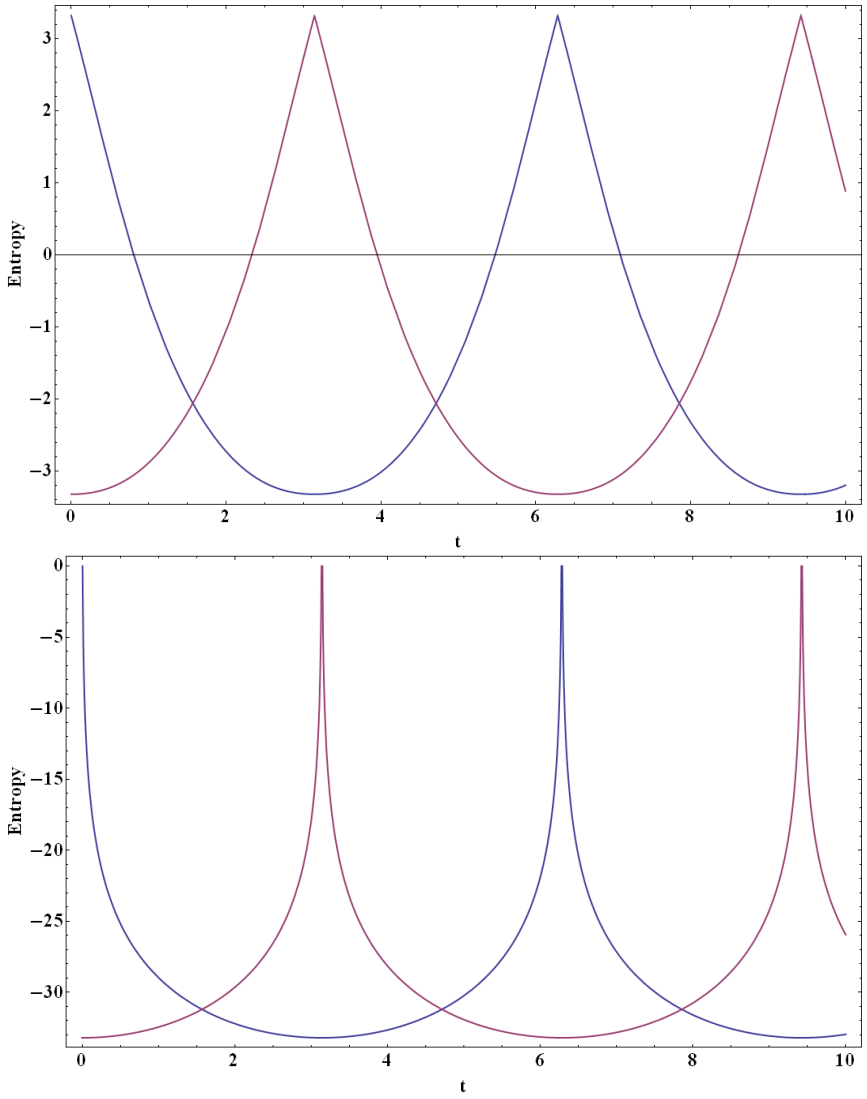

Figure 9.4: The functions $H^{\varepsilon}_{\min}(S)_{\tau(t)} - H^{\varepsilon}_{\max}(E)_{\tau(t)}$ and $H^{\varepsilon}_{\min}(S)_{\tau(t)} - H^{\varepsilon}_{\max}(E)_{\tau(t)}$ in the limit $\varepsilon \to 0$. For $d_S = 10^1$ (upper figure) one of the two functions is positive for approximately half of all times. For $d_S = 10^{10}$ (lower figure) the fraction of times for which one of the two functions is positive is vanishing.

Conversely, $\forall \varepsilon, \varepsilon' > 0$ $\exists N$ *such that* $\forall n \in \mathbb{N}$, $n > N$ *if* $H(A'|B)_\tau < -\varepsilon$ *then* $\nexists \omega_{B^n} \in \mathcal{S}_=(\mathcal{H}_B^{\otimes n})$ *such that*

$$\left\langle \left\| \mathcal{T}^{\otimes n}\left(\phi_{A^n}\right) - \omega_{B^n} \right\|_1 \right\rangle_{\phi_{A^n}} \leq \frac{1}{2} - \varepsilon' \tag{9.45}$$

and in particular

$$\left\langle \left\| \mathcal{T}^{\otimes n}\left(\phi_{A^n}\right) - \tau_B^{\otimes n} \right\|_1 \right\rangle_{\phi_{A^n}} > 1 - 2\varepsilon' \tag{9.46}$$

where the average is computed over the Haar measure on $\mathcal{H}_A^{\otimes n}$.

Proof. Theorem 4.2 gives for the channel $\mathcal{T}_{A^n \to B^n}^{\otimes n}$ that

$$\begin{aligned} &\Pr_{\phi_{A^n}} \left\{ \left\| \mathcal{T}_{A^n \to B^n}^{\otimes n}(\phi_{A^n}) - \tau_B^{\otimes n} \right\|_1 \geq 2^{-\frac{1}{2} H_{\min}^\varepsilon(A^n)_{\phi_{A^n}} - \frac{1}{2} H_{\min}^\varepsilon(A'^n|B^n)_{\tau^{\otimes n}}} + 12\varepsilon + \delta \right\} \\ &\quad \leq 2e^{-d_A^n \delta^2/16} \ . \end{aligned} \tag{9.47}$$

For the exponent we find with Lemma A.11 and (3.12)

$$\begin{aligned} &-\frac{1}{2} H_{\min}^\varepsilon(A^n)_{\phi_{A^n}} - \frac{1}{2} H_{\min}^\varepsilon(A'^n|B^n)_{\tau^{\otimes n}} \\ &\quad \leq \frac{n}{2} \left\{ -\frac{1}{n} H_{\min}^\varepsilon(A'^n|B^n)_{\tau^{\otimes n}} \right\} \\ &\quad \leq \frac{n}{2} \left\{ -H(A|B)_\tau + \frac{1}{\sqrt{n}} \cdot 4\sqrt{\log\left(\frac{2}{\varepsilon^2}\right)} \cdot \log\left(2^{-\frac{1}{2} H_{\min}(A|B)_\tau} + 2^{\frac{1}{2} H_{\max}(A|B)_\tau} + 1 \right) \right\} \\ &\quad \leq \frac{n}{2} \left\{ -H(A|B)_\tau + \underbrace{\frac{1}{\sqrt{n}} \cdot 4\sqrt{\log\left(\frac{2}{\varepsilon^2}\right)} \cdot \log\left(2\sqrt{d_A} + 1 \right)}_{=:c} \right\} \ . \end{aligned} \tag{9.48}$$

Defining $\varepsilon = \frac{1}{12n}$ and $\delta = d_A^{-n/3}$ yields the first assertion.

The condition of the converse, Theorem 4.4, requires for pure input states ϕ_{A^n}

$$H_{\min}^{\varepsilon' + 2\varepsilon'' + \varepsilon''' + \sqrt{\varepsilon}}(A^n)_{\phi_{A^n}} + \log \frac{2}{\varepsilon'^2} + H_{\max}^{\varepsilon''}(A'^n B^n)_{\tau^{\otimes n}} - H_{\min}^{\varepsilon'''}(B^n)_{\tau^{\otimes n}} < 0 \ . \tag{9.49}$$

The l.h.s. of this condition is by use of Lemma A.10, Lemma A.12, Lemma A.11 and (3.12)

$$
\begin{aligned}
&H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(A^n)_{\phi_{A^n}} + \log\frac{2}{\varepsilon'^2} + H_{\max}^{\varepsilon''}(A'^nB^n)_{\tau^{\otimes n}} - H_{\min}^{\varepsilon'''}(B^n)_{\tau^{\otimes n}} \\
&\quad\le n\left\{\frac{1}{n}\log\frac{1}{1-\left(\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}\right)^2} + \frac{1}{n}\log\frac{2}{\varepsilon'^2} + \frac{1}{n}H_{\max}^{\varepsilon''}(A'^nB^n)_{\tau^{\otimes n}} - \frac{1}{n}H_{\min}^{\varepsilon'''}(B^n)_{\tau^{\otimes n}}\right\} \\
&\quad\le n\left\{\frac{1}{n}\log\frac{1}{1-\left(\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}\right)^2} + \frac{1}{n}\log\frac{2}{\varepsilon'^2}\right. \\
&\qquad\left. + H(A'B)_\tau + \frac{1}{\sqrt{n}}\cdot 4\sqrt{\log\frac{2}{\varepsilon''^2}}\cdot\log\left(\sqrt{d_Ad_B}+2\right) - \frac{1}{n}H_{\min}^{\varepsilon'''}(B^n)_{\tau^{\otimes n}}\right\} \\
&\quad\le n\left\{\frac{1}{n}\log\frac{1}{1-\left(\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}\right)^2} + \frac{1}{n}\log\frac{2}{\varepsilon'^2}\right. \\
&\qquad + H(A'B)_\tau + \frac{1}{\sqrt{n}}\cdot 4\sqrt{\log\frac{2}{\varepsilon''^2}}\cdot\log\left(\sqrt{d_Ad_B}+2\right) \\
&\qquad\left. - H(B)_\tau + \frac{1}{\sqrt{n}}\cdot 4\sqrt{\log\left(\frac{2}{\varepsilon'''^2}\right)}\cdot\log\left(2^{-\frac{1}{2}H_{\min}(B)_\tau} + 2^{\frac{1}{2}H_{\max}(B)_\tau}+1\right)\right\} \\
&\quad\le n\left\{\frac{1}{n}\log\frac{1}{1-\left(\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}\right)^2} + \frac{1}{n}\log\frac{2}{\varepsilon'^2}\right. \\
&\qquad + H(A'B)_\tau + \frac{1}{\sqrt{n}}\cdot 4\sqrt{\log\frac{2}{\varepsilon''^2}}\cdot\log\left(\sqrt{d_Ad_B}+2\right) \\
&\qquad\left. - H(B)_\tau + \frac{1}{\sqrt{n}}\cdot 4\sqrt{\log\left(\frac{2}{\varepsilon'''^2}\right)}\cdot\log\left(\sqrt{d_B}+1\right)\right\} \\
&\quad=: n\left\{H(A'|B)_\tau + \delta(n)\right\}\ . \qquad (9.50)
\end{aligned}
$$

Hence if $H(A'|B)_\tau < \delta(n)$ condition (9.49) is fulfilled. Let $\varepsilon' = \varepsilon'' = \varepsilon''' = \frac{1}{n}$. ε may take any value for which the term $\log\frac{1}{1-\left(\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}\right)^2}$ is still well-defined. So for large enough n, ε may take any value of the form $1-2\delta'$ with $\delta'>0$. The quantity $\delta(n)$ becomes arbitrarily small for large enough n. So after relabling $\delta(n)\mapsto\varepsilon$ and $\delta'\mapsto\varepsilon'$ the assertion follows from Theorem 4.4. □

This has a direct physical interpretation. For instance consider the input space $\mathcal{H}_S^{\otimes n}$ to be a data storage and consider n to be large enough that we are allowed to

chose small parameters $\varepsilon, \varepsilon'$. Every factor is then subject to the same interaction with the same environment, i.e. the channel is

$$\mathcal{T}_{S\to S}(\sigma_S) := \mathrm{Tr}_E\left[e^{-\mathbb{i}\,H_{SE}t}\,(\sigma_S\otimes\rho_E)\,e^{+\mathbb{i}\,H_{SE}t}\right] \tag{9.51}$$

By the n-fold tensor product $\mathcal{T}_{S\to S}^{\otimes n}$ we model the i.i.d. characteristics of noise. At $t = 0$ we have $H(S'|S)_\tau = -\log d_S$. According to Theorem 9.3 as long as $H(S'|S)_\tau < 0$ different initial states of the storage $\mathcal{H}_S^{\otimes n}$ have not yet evolved to the inside of a ball with radius $\frac{1}{2}$ in trace distance. The storage still "remembers" its initial state. The storage has been erased by the noise if at some time $H(S'|S)_\tau > 0$. Almost all (pure) initial states *of the product space* $\mathcal{H}_S^{\otimes n}$ will then have evolved to the same state $\tau_S^{\otimes n}$. The channel (9.51) and thus also τ_S are in general time-dependent.

If $\log d_S > 4\log d_E$ (e.g. the noise consists of only a few photons) the noise can never fully erase the storage. To see this, let ϕ_{EP} be a purification of ρ_E, so

$$\tau_{S'SEP} := e^{-\mathbb{i}\,H_{SE}t}\,(\Psi_{S'S}\otimes\phi_{EP})\,e^{+\mathbb{i}\,H_{SE}t} \tag{9.52}$$

is a purification of $\tau_{S'S}$. For all times t we have $H(SEP)_\tau = \log d_S$. Then,

$$\begin{aligned} H(S'|S)_\tau &= H(EP)_\tau - H(S)_\tau \\ &\leq H(EP)_\tau - H(S|EP)_\tau \\ &= 2H(EP)_\tau - H(SEP)_\tau \\ &\leq 4\log d_E - \log d_S \\ &< 0\ . \end{aligned} \tag{9.53}$$

In order to calculate the minimal erasure time we want to know the time at which we have $H(S'|S)_\tau = 0 \Leftrightarrow H(S)_\tau = H(S'S)_\tau$ for the first time. For this we need to upper-bound the rate with which $H(S'S)_\tau$ can increase starting from 0 as well as the rate with which $H(S)_\tau$ can decrease starting from $\log d_S$. Theorem 7.3 tells us that the sum of the two rates is upper-bounded by $6\log d_S \cdot \Delta(H_{int})$, so that the time needed to reach $H(S'|S)_\tau = 0$ is for any S at least $\frac{1}{6}\cdot\frac{1}{\Delta(H_{int})}$. Slightly better results can be obtained from (7.36) for concrete d_S. The erasure time is for $d_S = 2$ at least $0.414\cdot\frac{1}{\Delta(H_{int})}$ and for large d_S at least $\frac{2}{3}\cdot\frac{1}{\Delta(H_{int})}$. Following the discussion in Section 2.5 we may improve the above bounds if we replace $\Delta(H_{int})$ by $\Delta\left(\tilde{H}_{int}\right)$ as a measure of the interaction strength.

We consider an example to discuss the achievability of these bounds. Recall the Hamiltonian with interaction strength $\Delta\left(\tilde{H}_{int}\right) = 1$ discussed in combination with the initial state $|1\rangle\langle 1|_E$ of the environment in Section 9.1.3. Since the state

$\tau_{S'SE} := e^{-\mathbb{i}H_{SE}t}\,(\Psi_{S'S}\otimes|1\rangle\langle 1|_E)\,e^{+\mathbb{i}H_{SE}t}$ is then pure, we have $H(S'S)_\tau = H(E)_\tau$. The time which is necessary to reach $H(S'|S)_\tau = 0$ is therefore equal to the time which is necessary to reach $H(S)_\tau = H(E)_\tau$. This time is independently of d_S given by $\frac{\pi}{2} = \frac{\pi}{2}\cdot\frac{1}{\Delta(\tilde{H}_{int})}$, as can be seen from (9.28), (9.29) and (9.38). Our lower bound on the erasure time is therefore for large d_S tight up to a factor of $\frac{3}{2}\cdot\frac{\pi}{2}\approx 2.356$. The corresponding values for other local dimensions can be found in Table 9.1.

	$d_S = 2$	$d_S = 3$	$d_S = 2^2$	$d_S = 2^3$	$d_S \to \infty$
Minimal erasure time	0.414	0.486	0.521	0.574	$\frac{2}{3}$

Table 9.1: Lower bounds on the times which are necessary to erase a storage S^n for large n through i.i.d. interaction with an environment in terms of the local dimension d_S. The times are measured in units of $\frac{1}{\Delta(\tilde{H}_{int})}$, the inverse of the interaction strength. Erasure is achievable for any local dimension $d_S \geq 2$ in time $\frac{\pi}{2}\cdot\frac{1}{\Delta(\tilde{H}_{int})} \approx 1.571\cdot\frac{1}{\Delta(\tilde{H}_{int})}$.

This bounds hold for any state E may initially be in and for any interaction between S and E. The only assumption we made is that the interaction between S and E may be described as i.i.d. on the length-scale of S. If there are still correlations in the system's interaction with the environment on the length-scale of one copy of S we may group a number of factors S together which is large enough such that the correlations in the interaction decay over the new enhanced length-scale. This obviously reduces n, the number of factors, correspondingly.

9.1.5 Local interactions

So far in this thesis we have dealt with the most general system and interaction compatible with the laws of quantum mechanics. In particular we allowed for cases in which "all of" the environment is interacting with "all of" the system. A more sensible model of the interaction would include a notion of locality and only allow for local interactions. In such a model we hope to find a better bound on the rate with which $H(E)_\tau$ can increase and $H(S)_\tau$ decrease than the one in Theorem 7.3. This bound includes the dimension of the whole system (or environment), which seems to be pessimistic in cases where only parts of it are interacting with the environment (or system).

Consider that the joint system is a set of sites each of which has a space $\mathbb{C}^d$ attached to it. The joint Hilbert space is then the tensor product of all those spaces.

Assume that the strength of the terms in the Hamiltonian which act on multiple sites decreases exponentially with the distance between those sites measured in a suitable distance measure. In such a system, the *Lieb-Robinson velocity* defines an effective light cone for causal interactions (c.f. [Has10] for a concise introduction). The influence of sites which are outside of the past light cone of a certain region defined by the Lieb-Robinson velocity is exponentially suppressed.

This supports the intuition that information can effectively propagate only with a certain maximal velocity if only short-range interactions are allowed. Still, it does not quite solve our problem which is how fast entropies in a certain region can be changed if we only allow for local interactions. Let the system be of the form $\mathcal{H}_S = \left(\mathbb{C}^d\right)^{\otimes n_S}$ and the environment $\mathcal{H}_E = \left(\mathbb{C}^d\right)^{\otimes n_E}$. Let the Hamiltonian be $H_{SE} = H_S \otimes \mathbb{1}_E + \mathbb{1}_S \otimes H_E + \sum_{j=1}^{r} K^{(j)}$. Geometrically, we expect the number of interaction terms r to scale with the surface of the n_S sites in S. In spatial dimension D this would imply a relation

$$r \propto n_S^{(D-1)/D} \ . \tag{9.54}$$

From eqn. (7.28) we know that only the interactive part of the Hamiltonian is relevant for changes of the local entropy. We have

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} H(S)_{\rho(t)} &= -\,\mathbb{i}\,\mathrm{Tr}_{SE}\left\{H_{int}\left[\log(\rho_S(t)) \otimes \mathbb{1}_E, \rho_{SE}(t)\right]\right\} \\ &= \sum_{j=1}^{r} -\,\mathbb{i}\,\mathrm{Tr}_{SE}\left\{K^{(j)}\left[\log(\rho_S(t)) \otimes \mathbb{1}_E, \rho_{SE}(t)\right]\right\} \ . \end{aligned} \tag{9.55}$$

The absolute value of each summand could now be upper-bounded using Theorem 7.3, that is by $2\Delta\left(K^{(j)}\right)\log d_S$. We would actually expect to find a better bound due to the assumed locality of the interaction. Let $K^{(j)}$ act non-trivially only on $w_S(j)$ tensor factors of $\mathcal{H}_S$ and $w_E(j)$ tensor factors of $\mathcal{H}_E$. We would then expect that there is an upper bound on the absolute value of each summand which is of the form

$$\left|-\,\mathbb{i}\,\mathrm{Tr}_{SE}\left\{K^{(j)}\left[\log(\rho_S(t)) \otimes \mathbb{1}_E, \rho_{SE}(t)\right]\right\}\right| \leq c \cdot \Delta\left(K^{(j)}\right) \cdot \min\left\{w_S(j), w_E(j)\right\} \cdot \log d \tag{9.56}$$

with a constant $c \in \mathcal{O}(1)$ which does not depend on n_S. This bound is symmetric in S and E and would thus also apply to entropy changes in E. Whether such an upper bound exists is actually an open problem (c.f. [Bra07] where an equivalent problem is discussed under different nomenclature). Numerical maximization results and the

fact that such a bound can be derived under special assumptions on the eigenvalues of $\rho(t)$ [Bra07] suggest that there is a general upper bound of this form. Assuming that there is such a bound we would have

$$\begin{aligned}\left|\frac{\mathrm{d}}{\mathrm{d}t}H(S)_{\rho(t)}\right| &\leq c\log d\sum_{j=1}^{r}\Delta\left(K^{(j)}\right)\min\left\{w_S(j),w_E(j)\right\}\\ &\leq c\tilde{K}r\log d \end{aligned} \tag{9.57}$$

with $\tilde{K}=\max_j\left(\Delta\left(K^{(j)}\right)\min\left\{w_S(j),w_E(j)\right\}\right)$. This bound scales with r which we expect to scale with the surface of the system. Formally we expect (9.54) to hold. The maximal entropy in S scales with n_S. Together, this would imply that the minimal time which is needed to reach a high-entropy state on S starting from a pure state scales with $n_S^{1/D}$, i.e. with the *diameter* of the system. A time of the same order would be needed to reach $H(S)=H(E)$ starting from $H(S)=\log d_S$ and $H(E)=0$ and thus to reach initial state independence in the i.i.d. scenario.

Every bipartite Hermitian operator like H_{int} may be decomposed into a sum of products of Hermitian operators. From (7.37) we know that for product Hamiltonians the maximal achievable rate of change of the entropy in S does *not* depend on $\log d_S$. From (9.55) we know that local changes of the von Neumann entropy are additive. The maximal rate with which the entropy in S can change therefore scales with the number of product terms in the decomposition of H_{int}. The maximal number of product terms in the decomposition scales polynomially with the local dimension d_S [BHLS03].

A bound on the rate of change of the logarithmic quantity $H(S)$ which grows polynomially in d_S may in general not be of much use for our purposes. Still, we may derive useful bounds with this method if we bound the number of product terms in the decomposition of H_{int} by use of the assumed locality of the interaction.

If the number of product terms in the decomposition of H_{int} scales with the surface of S so does the maximal rate of change of $H(S)$. Under this assumption, the fact, that the minimal time to reach a high-entropic state scales with the diameter of S, may be derived without having to rely on unproved conjectures.

Assume, for example, that the summands $K^{(j)}$ describe interactions between two qubits. Then there can only be a small number of product terms in the decomposition of each $K^{(j)}$. The number of product terms in the decomposition of $H_{int}=\sum_{j=1}^{r}K^{(j)}$ is then at most a small number times r, the number of qubits in S which are interacting with qubits in E. Every such product term leads to a maximal entropy change which according to (7.37) is independent of any local dimensions. From (9.55) we know that the entropy changes induced by different summands of the Hamiltonian

are additive. We conclude that in this scenario the maximal rate with which the local entropy can be changed scales with r, the number of qubits in S which are interacting with qubits in E.

9.2 For generic initial states of the system and generic initial states of the environment

9.2.1 Maximal entropy does not decrease

In the previous section we were dealing with a specific pure initial state of the environment. If we remind ourselves of our results about independence of the initial state of the environment, this is not really necessary if we are ready to neglect an exponentially small fraction of initial states of the environment. Applying our results about independence of the initial state of the *environment* will allow us to progress further on the issue of independence of the initial state of the system.

We are still interested in how long we can guarantee that different initial states of S have not yet evolved to the same state. In Section 9.1 we saw that in order to answer this question it is relevant to know how fast the entropy in S can decrease if it is initially maximal. The bounds we applied hold for *any* initial state of E. Yet, for generic initial states of E they are rather pessimistic. In fact, almost all initial states of E do not lead to a significant decrease of the entropy in S, if S is initially fully mixed and E is sufficiently larger than S. This allows to improve the bounds from Theorem 9.2 for almost all initial states of E.

Lemma 9.4. *Consider an initial product state $\pi_S \otimes \rho_E(0)$. Let $\rho_S(t)$ denote the state of S at time t. Then for all times t*

$$\Pr_{\rho_E}\left[\|\rho_S(t) - \pi_S\|_1 > \frac{d_S}{\sqrt{d_E}} + d_E^{-1/3}\right] < e^{-d_E^{1/3}/16} \tag{9.58}$$

where the proability is computed over the choice of $\rho_E(0)$ from any unitarily invariant measure on $\mathcal{S}_=(\mathcal{H}_E)$.

Proof. This follows directly from Corollary 8.2 where according to (8.7)

$$\begin{aligned}\tau_S &= \mathrm{Tr}_{EE'}\left[e^{-\mathbb{i}\,H_{SE}t}\,(\pi_S \otimes \Psi_{EE'})\,e^{+\mathbb{i}\,H_{SE}t}\right] \\ &= \mathrm{Tr}_E\left[e^{-\mathbb{i}\,H_{SE}t}\,(\pi_S \otimes \pi_E)\,e^{+\mathbb{i}\,H_{SE}t}\right] \\ &= \pi_S\ .\end{aligned} \tag{9.59}$$

□

According to Lemma 9.4, an initial product state which is fully mixed on S will with high probability be close to fully mixed at any time if $\log d_E \gtrapprox \max\{2\log d_S, 12\}$. It then follows from Fannes' inequality, Theorem 3.1, that the von Neumann entropy of such a state will be close to maximal. Similarly, the smooth min-entropy will at any time be large with high probability. We have for all times $\|\rho_S(t) - \pi_S\|_1 \leq \frac{d_S}{\sqrt{d_E}} + d_E^{-1/3}$ except with a probability of less then $e^{-d_E^{1/3}/16}$. If this is the case, we find with Lemma 3.3 and the definition of the smooth min-entropy (3.40) that $H_{\min}^{\varepsilon}(S)_{\rho(t)} \geq \log d_S$ if $\varepsilon \geq \sqrt{2\left(\frac{d_S}{\sqrt{d_E}} + d_E^{-1/3}\right)}$.

9.2.2 Quantitative version

In Section 9.1.2 we saw that the question whether the system is already independent of its initial state is equivalent to the question whether the entropy in E of the state $\tau_{SE}(t)$ is already larger than the entropy in S. In Section 9.1.3 we investigated how long we can guarantee that this is not the case and thus considered the maximal rate with which the entropy in S can decrease. The result in the previous section shows that this is not necessary for most initial states of the environment because S stays close to fully mixed.

The times for which we can guarantee that different initial states of S have on average not yet evolved to the inside of a ball with radius $\frac{\varepsilon}{2}$ depend on d_S and ε. For simplicity, we discuss the two limits of large d_S and small ε.

Corollary 9.5. *Consider a system S which at $t = 0$ is put in contact with an environment which is in a pure state ϕ_E. Let the joint evolution be governed by a Hamiltonian H_{SE} with interaction strength $\Delta\left(\tilde{H}_{int}\right)$ as defined in (2.37). Let $\rho_S^{\phi}(t)$ denote the state of the system at time t given that its state at $t = 0$ was ϕ_S. Let $\log d_E$ be sufficiently larger than $2\log d_S$. Then the following two statements hold for all initial states of the environment ϕ_E except a fraction of less than $e^{-d_E^{1/3}/16}$ from the Haar measure.*

1. For $\log d_S \gtrapprox 25$ as long as

$$t < \begin{cases} 0.058 \cdot \frac{1}{\Delta\left(\tilde{H}_{int}\right)} \\ 0.101 \cdot \frac{1}{\Delta\left(\tilde{H}_{int}\right)} \\ 0.117 \cdot \frac{1}{\Delta\left(\tilde{H}_{int}\right)} \\ 0.125 \cdot \frac{1}{\Delta\left(\tilde{H}_{int}\right)} \end{cases} \tag{9.60}$$

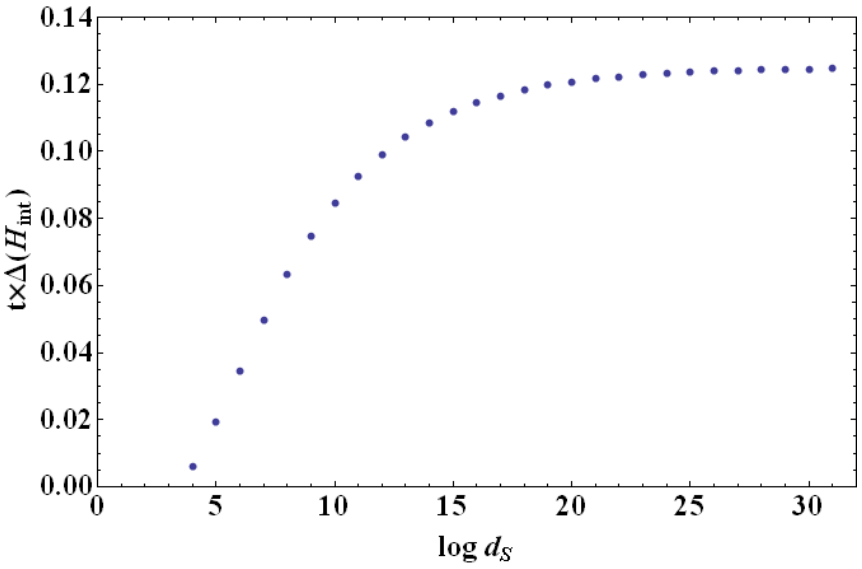

Figure 9.5: The times for which we can guarantee that different pure initial states of S have not yet evolved to the same state as a function of $\log d_S$. The times are measured in units of $\frac{1}{\Delta\left(\tilde{H}_{int}\right)}$. We assumed that E is larger than two copies of S. The statement then holds for all initial states of E from the Haar measure except a fraction which is exponentially small in $d_E^{1/3}$.

there is no state $\omega_S \in \mathcal{S}_=(\mathcal{H}_S)$ *such that*

$$\left\langle \left\| \rho_S^{\phi}(t) - \omega_S \right\|_1 \right\rangle_{\phi_S} < \frac{\varepsilon}{2} \tag{9.61}$$

where

$$\begin{cases} \varepsilon = 10^{-1} \\ \varepsilon = 10^{-2} \\ \varepsilon = 10^{-3} \\ \varepsilon \to 0 \end{cases} . \tag{9.62}$$

2. In the limit of $\varepsilon \to 0$ *the times for which we can guarantee that there is no state* $\omega_S \in \mathcal{S}_=(\mathcal{H}_S)$ *which fulfills (9.61) are given in Figure 9.5.*

We recall that we showed in Section 9.1.3 that initial state independence in the sense of (9.61) can be achieved (for any ε and large enough d_S) in time $\frac{\pi}{\Delta\left(\tilde{H}_{int}\right)}$.

Comparing the above results with the ones obtained in Theorem 9.2, we see that neglecting an exponentially small fraction of initial states of E allowed us to improve the lower bounds on the times which are necessary for initial state independence by a factor of around 5. Since we are now dealing with generic initial states of the environment, we have lost the dependence on the relation between the initial state of the environment and the Hamiltonian which is expressed in Theorem 9.2 with the term $\|[H_{SE}, \pi_S \otimes \phi_E]\|_1$.

Proof. We are interested in the channel

$$\phi_S \mapsto \rho_S^\phi(t) = \mathcal{T}_{S\to S}(\phi_S) = \mathrm{Tr}_E\left[e^{-\mathbb{i}H_{SE}t}(\phi_S \otimes \phi_E)\, e^{+\mathbb{i}H_{SE}t}\right] . \tag{9.63}$$

We introduce the state

$$\tau_{S'SE} := e^{-\mathbb{i}H_{SE}t}(\Psi_{S'S} \otimes \phi_E)\, e^{+\mathbb{i}H_{SE}t} \tag{9.64}$$

which purifies the Choi-Jamiołkowski representation of $\mathcal{T}_{S\to S}$. From Theorem 4.4 we know that if

$$H_{\min}^{\varepsilon'+2\varepsilon''+\varepsilon'''+\sqrt{\varepsilon}}(S)_\phi + H_{\max}^{\varepsilon''}(S'S)_\tau - H_{\min}^{\varepsilon'''}(S)_\tau + \log\frac{2}{\varepsilon'^2} < 0 \tag{9.65}$$

there is no state $\omega_S \in \mathcal{S}_=(\mathcal{H}_S)$ such that

$$\left\langle \left\| \rho_S^\phi(t) - \omega_S \right\|_1 \right\rangle_{\phi_S} < \frac{\varepsilon}{2} . \tag{9.66}$$

The first summand can be bounded by use of Lemma A.10. Using that $\tau_{S'SE}$ is pure we have by Lemma 3.4

$$H_{\max}^{\varepsilon''}(S'S)_\tau = H_{\max}^{\varepsilon''}(E)_\tau \tag{9.67}$$

which can then be bounded by use of Corollary 7.5 and the fact that $\bar{T} \geq \frac{1}{2\Delta(\tilde{H}_{int})}$. As for the third summand, τ_S is at $t = 0$ fully mixed which allows to apply Lemma 9.4. We conclude that except for a fraction of states ϕ_E from the Haar measure which is smaller than $e^{-d_E^{1/3}/16}$ we have

$$H_{\min}^{\varepsilon'''}(S)_\tau \geq \log d_S . \tag{9.68}$$

Here, according to Lemma 9.4 and by use of Lemma 3.3, it is sufficient to choose

$$\varepsilon''' = \sqrt{2\left(\frac{d_S}{\sqrt{d_E}} + d_E^{-1/3}\right)} . \tag{9.69}$$

Since we assumed that $\log d_E$ is sufficiently larger than $2\log d_S$, we view ε''' as negligible. Finally, the parameter ε' can be chosen so as to minimize the l.h.s. of (9.65). Inserting the relevant values of ε and d_S and solving (9.65) for the time for which equality is obtained for the first time yields the assertions. □

9.2.3 In the i.i.d. scenario

In Corollary 9.5 we made use of the insight that if the initial state of S is fully mixed the system will at all later times be so for almost all initial states of the environment. This can also be applied in the i.i.d. scenario. In Section 9.1.4 we had to upper-bound the rates with which $H(S)_\tau$ can decrease from $\log d_S$ and $H(S'S)_\tau$ can increase from 0 (with τ as defined in (9.52)). From Lemma 9.4 and Fannes' inequality, Theorem 3.1, we know that in fact $H(S)_\tau$ will stay close to $\log d_S$ for all times and for all initial states of E except a fraction $e^{-d_E^{1/3}/16}$ from the Haar measure if $\log d_E$ is sufficiently larger than $2 \log d_S$. If we are willing to neglect this exponentially small fraction of environment states, we only have to bound the rate with which $H(S'S)_\tau$ can increase. This allows to improve the bounds from Table 9.1 by a factor of approximately $\frac{3}{2}$ to the ones given in Table 9.2.

	$d_S = 2$	$d_S = 3$	$d_S = 2^2$	$d_S = 2^3$	$d_S \to \infty$
Minimal erasure time	0.414	0.486	0.521	0.574	$\frac{2}{3}$
Minimal typical erasure time	0.685	0.792	0.841	0.909	1

Table 9.2: The first line gives lower bounds on the times which are necessary to erase a storage S^n for large n through interaction with any i.i.d. environment in terms of the local dimension d_S. The second line gives lower bounds if we are willing to neglect a Haar measure fraction of less than $\exp\left(-d_E^{1/3}/16\right)$ of the states the environment might initially be in. The times are measured in units of $\frac{1}{\Delta\left(\tilde{H}_{int}\right)}$, the inverse of the interaction strength. Erasure is achievable for any local dimension $d_S \geq 2$ in time $\frac{\pi}{2} \cdot \frac{1}{\Delta\left(\tilde{H}_{int}\right)} \approx 1.571 \cdot \frac{1}{\Delta\left(\tilde{H}_{int}\right)}$.

9.3 For specific initial states of the system and generic initial states of the environment

In this section, we discuss sufficient conditions under which a system does for most initial states of the environment *not* become independent of a particular initial state. We have already seen in Theorem 6.4 that the time-averaged state of the system is independent of its initial state if the energy eigenstates are sufficiently entangled. In [GME11] a converse statement is proven. Assume that there is a basis $\{|i\rangle_S\}_i$ of $\mathcal{H}_S$ which is such that for all energy eigenstates the partial trace $\mathrm{Tr}_E |E_k\rangle\langle E_k|$ is close to one of the basis states $|i\rangle\langle i|_S$. This implies that all energy eigenstates are close

to product and thus are poorly entangled. Consider two pure initial product states whose S-part is one of the basis elements. Then the distance of their time-averages on S is unlikely to be much smaller than the initial distance of their S-parts. Formally, we have the following theorem.

Theorem 9.6 ([GME11, Theorems 1 and 2])**.** *Consider two pure initial product states $\phi_{SE}^{(i)}(0) = \phi_S^{(i)}(0) \otimes \phi_E^{(i)}(0)$, $i \in \{1,2\}$ evolving under a non-degenerate Hamiltonian H_{SE}, yielding temporal averages $\omega_{SE}^{(i)}$. We define the quantity*

$$R(\phi^{(i)}(0)) = \sum_k \left|\left\langle E_k|\phi^{(i)}\right\rangle_{SE}\right|^2 \left\|\mathrm{Tr}_E\left|E_k\right\rangle\!\left\langle E_k\right| - \phi_S^{(i)}(0)\right\|_1 \tag{9.70}$$

which is small if all energy eigenstates either are close to orthogonal to $\phi_{SE}^{(i)}(0)$ or locally resemble $\phi_S^{(i)}(0)$. Then,

$$\left\|\omega_S^{(1)} - \omega_S^{(2)}\right\|_1 \geq \left\|\phi_S^{(1)}(0) - \phi_S^{(2)}(0)\right\|_1 - R(\phi^{(1)}(0)) - R(\phi^{(2)}(0)) \ . \tag{9.71}$$

The quantity $R(\phi^{(i)}(0))$ is small for almost all $\phi_E^{(i)}(0)$ if $\phi_S^{(i)}(0)$ is an element of a basis $\{|l\rangle_S\}_{l=1}^{d_S}$ of $\mathcal{H}_S$ which is such that the partial trace of each energy eigenstate is close to one of the basis elements. Formally,

$$\left\langle R\left(\phi_S^{(i)}(0) \otimes \phi_E^{(i)}(0)\right)\right\rangle_{\phi_E^{(i)}(0)} \leq \delta d_S \tag{9.72}$$

where

$$\delta := \max_k \min_l \left\|\mathrm{Tr}_E\left|E_k\right\rangle\!\left\langle E_k\right| - \left|l\right\rangle\!\left\langle l\right|_S\right\|_1 \ . \tag{9.73}$$

This disproves the long-held conjecture that all non-integrable systems thermalize (since they do not necessarily become independent of their initial state). We refer to [GME11] for different definitions of integrability. A possible definition is to require the existence of d_S mutually commuting and linearly independent conserved operators on S. Developing the result of the decoupling theorem in the form of Theorems 8.1 and 8.3 a little further, we are able to strengthen this result. Instead of making a statement about the temporal average of S we make a statement about the time-evolved state itself. We do not only prove that the distance between pure initial states which locally resemble an energy eigenstate does not decrease but that such states actually stay close to their initial states for all times. In contrast to statement (9.72) which is a statement about an averaged distance, we obtain an exponentially strong statement. We do not require that all energy eigenstates be close

to product but only the ones which are most relevant for the particular initial state of the system. Furthermore, we allow for mixed initial states of the environment. The most important improvement, however, is the following: Inequality (9.72) gives only a non-trivial statement if $\delta < \frac{2}{d_S}$. This upper bound decreases exponentially with the number of constituent particles of the system. For a large system S we therefore basically need no interaction at all. On the other hand, the bound we obtain is non-trivial independently of the size of S.

Theorems 8.1 and 8.3 do not only tell us that the time-evolved state of the system $\rho_S(t)$ will very likely be close to a state $\tau_S(t)$ which is independent of the initial state of the environment, but also what this state looks like. If $\rho_S(0) = \phi_S$ and the initial state of the environment is drawn from all of $\mathcal{H}_E$ it is given by

$$\tau_S(t) = \mathrm{Tr}_E \left[e^{-\mathbb{i} H_{SE} t} \left(\phi_S \otimes \pi_E \right) e^{+\mathbb{i} H_{SE} t} \right] \ . \tag{9.74}$$

Saying that the time-evolved state becomes independent of the initial state of S is equivalent with saying that $\tau_S(t)$ is independent of ϕ_S. We will show that this does not happen, but that in fact $\tau_S(t) \approx \phi_S$ for all times, if the energy eigenstates which, in a certain sense, locally are "most similar" to ϕ_S, are close to product states.

In order to make this precise, consider a basis $\{|i\rangle_S\}_{i=1,\ldots,d_S}$ of $\mathcal{H}_S$ and a basis $\{|j\rangle_E\}_{j=1,\ldots,d_E}$ of $\mathcal{H}_E$. To each energy eigenstate $|E_k\rangle$ we assign the pair (i,j) of basis states which is closest to $|E_k\rangle$. More precisely, we assign to each $|E_k\rangle$ the pair (i,j) which maximizes the expression $F\left(|E_k\rangle, |i\rangle_S |j\rangle_E\right)$. We denote the pair (i,j) which achieves this maximum by $\left(\xi(k), \hat{\xi}(k)\right)$ and define

$$f_k := F\left(|E_k\rangle, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right) \ . \tag{9.75}$$

This gives a mapping

$$\begin{array}{ccc} \{1,\ldots,d_S d_E\} & \longrightarrow & \{1,\ldots,d_S\} \times \{1,\ldots,d_E\} \\ k & \longmapsto & \left(\xi(k), \hat{\xi}(k)\right) \ . \end{array} \tag{9.76}$$

We will assume that this mapping is injective, that is, every energy eigenstate is best approximated by a different element of the product basis. Since the sets $\{1,\ldots,d_S d_E\}$ and $\{1,\ldots,d_S\} \times \{1,\ldots,d_E\}$ have the same cardinality, this also implies that the mapping is bijective.

The relevant measure of entanglement is then for $\phi \in \{1,\ldots,d_S\}$

$$\delta(\phi) := \min\left\{ f_k : \xi(k) = \phi \right\} \ . \tag{9.77}$$

Lemma 9.7. *With the notation introduced above, assume that the mapping* $k \mapsto \left(\xi(k), \hat{\xi}(k)\right)$ *is injective and consider an initial state* $|\phi\rangle_S$ *with* $\phi \in \{1, \ldots, d_S\}$. *Assume that* $\delta(\phi) > \frac{1}{\sqrt{2}}$. *Then with* $\tau_S(t)$ *as defined in (9.74) we have*

$$\|\tau_S(t) - \phi_S\|_1 \leq 4\delta(\phi)\sqrt{1-\delta(\phi)^2} \tag{9.78}$$

for all times t.

We can always obtain an upper bound on $\|\tau_S(t) - \phi_S\|_1$ which is close to 0 if $\delta(\phi)$ is close to 1. By definition (9.77) the requirement $\delta(\phi) > \frac{1}{\sqrt{2}}$ requires that $f_k > \frac{1}{\sqrt{2}}$ if $\xi(k) = \phi$. If this condition is fulfilled, the r.h.s. of (9.78) is smaller than 2 and thus non-trivial. We can make sure that the mapping is injective as assumed by requiring that $f_k > \frac{1}{\sqrt{2}}$ for all k.

Combining Lemma 9.7 with Corollary 8.2 directly yields the statement that for all times there is an exponentially small probability that the initial state of the environment was such that the system is further away from its initial state than a certain distance.

Theorem 9.8. *With the notation introduced above, assume that the mapping* $k \mapsto \left(\xi(k), \hat{\xi}(k)\right)$ *is injective and consider an initial state* $|\phi\rangle_S$ *with* $\phi \in \{1, \ldots, d_S\}$. *Assume that* $\delta(\phi) > \frac{1}{\sqrt{2}}$. *Let* $\rho_S(t)$ *denote the evolved state of the system. Then for all times* t

$$\Pr_{\rho_E(0)}\left[\|\rho_S(t) - \phi_S\|_1 > 4\delta(\phi)\sqrt{1-\delta(\phi)^2} + \frac{d_S}{\sqrt{d_E}} + d_E^{-1/3}\right] < e^{-d_E^{1/3}/16} \tag{9.79}$$

where the probability is computed over the choice of the initial state of the environment $\rho_E(0)$ *from any unitarily invariant measure on* $\mathcal{S}_=(\mathcal{H}_E)$.

Consider the case where $\{|i\rangle_S\}_{i=1,\ldots,d_S}$ is the eigenbasis of H_S and $\{|j\rangle_E\}_{j=1,\ldots,d_E}$ is the eigenbasis of H_E. We view H_{int} as a small perturbation of the unperturbed Hamiltonian $H_S \otimes \mathbb{1}_E + \mathbb{1}_S \otimes H_E$. Then, the unperturbed value of f_k is 1 and the first order correction is 0.

A weakness of both Theorem 9.6 and Theorem 9.8 is that an average or a probability is computed over the choice of an initial state of the environment from the Haar measure on all of $\mathcal{H}_E$. This allows for the possibility that there is a reasonably large subspace $\mathcal{H}_{\Omega_E} \subseteq \mathcal{H}_E$ which is such that initial states taken from it do not lead to memory effects in system. One might imagine, for example, that all initial states of E with a sufficiently large temperature do not lead to such memory effects, but that still most states from the Haar measure on all of $\mathcal{H}_E$ do.

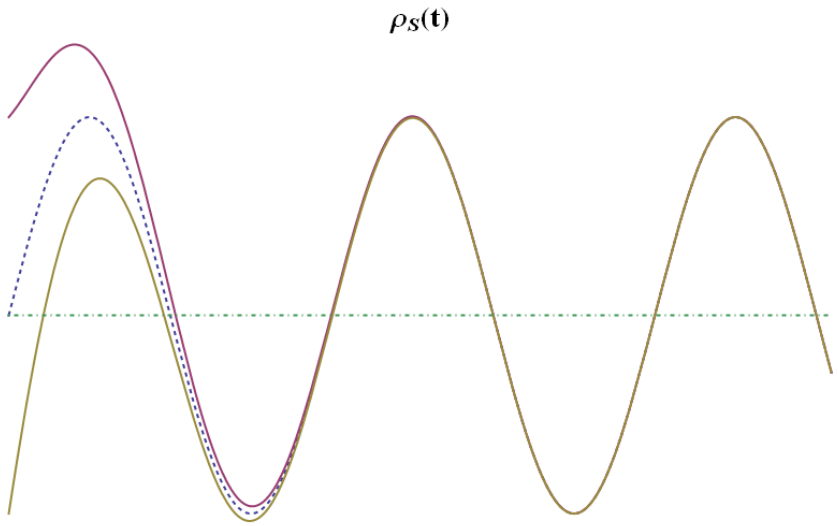

Figure 9.6: Different initial states (bold lines) approach a common evolving state (dashed line) which does not equilibrate towards its time-average (dashed-dotted line).

In Chapter 6 we mentioned initial state independence and equilibration as necessary conditions for thermalization. Here we showed that even when evolving under a non-integrable Hamiltonian it is possible that a state of the system does not become independent of its initial state for most environment states. From Chapter 6 we know, however, that it will equilibrate for most environment states if the environment is large enough. Hence the two conditions are independent of each other. An interesting and to the knowledge of the author open problem is, whether the independence also holds in the other direction. That is, is it possible that most initial states of the system become independent of their initial state (or at least that their time-average is independent) but that they do not equilibrate? The idea is drafted in Figure 9.6.

We will now proof Lemma 9.7. The above theorem then follows straighforwardly from Corollary 8.2 and a single application of the triangle inequality.

Proof. In order to shorten our notation we introduce the shorthands $\xi(k)_S \equiv |\xi(k)\rangle\langle\xi(k)|_S$ and $\hat{\xi}(k)_E \equiv |\hat{\xi}(k)\rangle\langle\hat{\xi}(k)|_E$. Sums with summation index k or l go from 1 to $d_S d_E$ and sums with summation index r go from 1 to d_E. By use of the assumed injectivity

(and hence also bijectivity) of the mapping, we have

$$\delta_{\xi(k),\xi(l)}\delta_{\hat{\xi}(k),\hat{\xi}(l)} = \delta_{kl} \ . \tag{9.80}$$

This implies that $\sum_k e^{-\mathbb{i}\,E_k t}\xi(k)_S \otimes \hat{\xi}(k)_E$ is a unitary, since

$$\begin{aligned}
&\left(\sum_k e^{-\mathbb{i}\,E_k t}\xi(k)_S \otimes \hat{\xi}(k)_E\right)\left(\sum_l e^{-\mathbb{i}\,E_l t}\xi(l)_S \otimes \hat{\xi}(l)_E\right)^\dagger \\
&= \left(\sum_k e^{-\mathbb{i}\,E_k t}\xi(k)_S \otimes \hat{\xi}(k)_E\right)\left(\sum_l e^{+\mathbb{i}\,E_l t}\xi(l)_S \otimes \hat{\xi}(l)_E\right) \\
&= \sum_{kl} e^{-\mathbb{i}(E_k-E_l)t}\delta_{\xi(k),\xi(l)}\delta_{\hat{\xi}(k),\hat{\xi}(l)}|\xi(k)\rangle\langle\xi(l)|_S \otimes |\hat{\xi}(k)\rangle\langle\hat{\xi}(l)|_E \\
&= \sum_k e^{-\mathbb{i}(E_k-E_k)t}|\xi(k)\rangle\langle\xi(k)|_S \otimes |\hat{\xi}(k)\rangle\langle\hat{\xi}(k)|_E \\
&= \mathbb{1}_{SE} \ .
\end{aligned} \tag{9.81}$$

We first show that $\tau_S(t)$ has high fidelity with the state

$$\mathrm{Tr}_E\left[\left(\sum_k e^{-\mathbb{i}\,E_k t}\xi(k)_S \otimes \hat{\xi}(k)_E\right)(\phi_S \otimes \pi_E)\left(\sum_l e^{+\mathbb{i}\,E_l t}\xi(l)_S \otimes \hat{\xi}(l)_E\right)\right]$$

and then show that this state is identical with ϕ_S. According to (3.29) the fidelity can only increase under partial traces, that is, it can only decrease if we calculate it for purifications of the actual states, so

$$\begin{aligned}
&F^2\left\{\tau_S(t), \mathrm{Tr}_E\left[\left(\sum_k e^{-\mathbb{i}\,E_k t}\xi(k)_S \otimes \hat{\xi}(k)_E\right)(\phi_S \otimes \pi_E)\left(\sum_l e^{+\mathbb{i}\,E_l t}\xi(l)_S \otimes \hat{\xi}(l)_E\right)\right]\right\} \\
&= F^2\left\{\mathrm{Tr}_E\left[\left(\sum_k e^{-\mathbb{i}\,E_k t}|E_k\rangle\langle E_k|\right)(\phi_S \otimes \pi_E)\left(\sum_l e^{+\mathbb{i}\,E_l t}|E_l\rangle\langle E_l|\right)\right],\right. \\
&\qquad \left.\mathrm{Tr}_E\left[\left(\sum_k e^{-\mathbb{i}\,E_k t}\xi(k)_S \otimes \hat{\xi}(k)_E\right)(\phi_S \otimes \pi_E)\left(\sum_l e^{+\mathbb{i}\,E_l t}\xi(l)_S \otimes \hat{\xi}(l)_E\right)\right]\right\} \\
&\geq F^2\left\{\left(\sum_k e^{-\mathbb{i}\,E_k t}|E_k\rangle\langle E_k|\right)(\phi_S \otimes \Psi_{EE'})\left(\sum_l e^{+\mathbb{i}\,E_l t}|E_l\rangle\langle E_l|\right),\right. \\
&\qquad \left.\left(\sum_k e^{-\mathbb{i}\,E_k t}\xi(k)_S \otimes \hat{\xi}(k)_E\right)(\phi_S \otimes \Psi_{EE'})\left(\sum_l e^{+\mathbb{i}\,E_l t}\xi(l)_S \otimes \hat{\xi}(l)_E\right)\right\} \ .
\end{aligned} \tag{9.82}$$

Both these states are pure, so using eqn. (3.28) and that $|\Psi\rangle_{EE'} = \frac{1}{\sqrt{d_E}} \sum_r |r\rangle_E |r\rangle_{E'}$ we find

$$
\begin{aligned}
& F^2 \left\{ \tau_S(t), \mathrm{Tr}_E \left[\left(\sum_k e^{-\mathbb{i}\, E_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) (\phi_S \otimes \pi_E) \left(\sum_l e^{+\mathbb{i}\, E_l t} \xi(l)_S \otimes \hat{\xi}(l)_E \right) \right] \right\} \\
& \quad \geq \left| \langle \phi|_S \langle \Psi|_{EE'} \left(\sum_l e^{+\mathbb{i}\, E_l t} |E_l\rangle\langle E_l|_{SE} \right) \left(\sum_k e^{-\mathbb{i}\, E_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) |\phi\rangle_S |\Psi\rangle_{EE'} \right|^2 \\
& \quad = \left| \frac{1}{d_E} \sum_r \langle \phi|_S \langle r|_E \left(\sum_l e^{+\mathbb{i}\, E_l t} |E_l\rangle\langle E_l|_{SE} \right) \left(\sum_k e^{-\mathbb{i}\, E_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) |\phi\rangle_S |r\rangle_E \right|^2 \\
& \quad = \left| \frac{1}{d_E} \sum_r \langle \phi|_S \langle r|_E \left(\sum_l e^{+\mathbb{i}\, E_l t} |E_l\rangle\langle E_l|_{SE} \right) \left(\sum_k e^{-\mathbb{i}\, E_k t} |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right) \delta_{\phi,\xi(k)} \delta_{r,\hat{\xi}(k)} \right|^2 \\
& \quad = \left| \frac{1}{d_E} \sum_{kl} \delta_{\phi,\xi(k)} e^{-\mathbb{i}(E_k - E_l)t} \langle \xi(k)|_S \langle \hat{\xi}(k)|_E |E_l\rangle\langle E_l|_{SE} |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right|^2 \\
& \quad = \left| \frac{1}{d_E} \sum_{kl} \delta_{\phi,\xi(k)} e^{-\mathbb{i}(E_k - E_l)t} F^2 \left\{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right\} \right|^2 \qquad (9.83)
\end{aligned}
$$

By definition (9.77) the requirement $\delta(\phi) > \frac{1}{\sqrt{2}}$ requires that

$$
F^2 \left\{ |E_k\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right\} > \frac{1}{2} \qquad (9.84)
$$

if $\delta_{\phi,\xi(k)} = 1$. Since $\sum_l F^2 \left\{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right\} = 1$, this also implies that[2]

$$
\sum_{l:l\neq k} F^2 \left\{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right\} < \frac{1}{2} \qquad (9.85)
$$

[2] With $\sum_{l:l\neq k}$ we denote a sum over all values of l which are not equal to k. With $\sum_{k\neq l}$ we denote a sum over all pairs of possible values of k and l which are not equal.

if $\delta_{\phi,\xi(k)} = 1$. We conclude that

$$\begin{aligned}
&\left| \sum_{k \neq l} \delta_{\phi,\xi(k)} e^{-\mathbb{i}(E_k - E_l)t} F^2 \left\{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right\} \right| \\
&\quad \leq \sum_{k \neq l} \delta_{\phi,\xi(k)} F^2 \left\{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right\} \\
&\quad \leq \sum_{k = l} \delta_{\phi,\xi(k)} F^2 \left\{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right\} \ . \qquad (9.86)
\end{aligned}$$

We split up the sum $\sum_{kl} = \sum_{k=l} + \sum_{k \neq l}$ and use that for $a, b \in \mathbb{C}$ with $|a| \geq |b|$ we have $|a + b| \geq |a| - |b|$ to obtain from (9.83)

$$\begin{aligned}
&F^2 \left\{ \tau_S(t), \mathrm{Tr}_E \left[\left(\sum_k e^{-\mathbb{i} E_k t} \xi(k)_S \otimes \hat{\xi}(k)_E \right) (\phi_S \otimes \pi_E) \left(\sum_l e^{+\mathbb{i} E_l t} \xi(l)_S \otimes \hat{\xi}(l)_E \right) \right] \right\} \\
&\geq \left(\frac{1}{d_E} \left| \sum_{k=l} \delta_{\phi,\xi(k)} e^{-\mathbb{i}(E_k - E_l)t} F^2 \left\{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right\} \right| \right. \\
&\qquad \left. - \frac{1}{d_E} \left| \sum_{k \neq l} \delta_{\phi,\xi(k)} e^{-\mathbb{i}(E_k - E_l)t} F^2 \left\{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right\} \right| \right)^2 \\
&\geq \left(\frac{1}{d_E} \sum_{k=l} \delta_{\phi,\xi(k)} F^2 \left\{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right\} \right. \\
&\qquad \left. - \frac{1}{d_E} \sum_{k \neq l} \delta_{\phi,\xi(k)} F^2 \left\{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right\} \right)^2 . \qquad (9.87)
\end{aligned}$$

Using that

$$\sum_{l: l \neq k} F^2 \left\{ |E_l\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right\} = 1 - F^2 \left\{ |E_k\rangle_{SE}, |\xi(k)\rangle_S |\hat{\xi}(k)\rangle_E \right\} = 1 - f_k^2 \qquad (9.88)$$

this simplifies to

$$
\begin{aligned}
&F^2\left\{\tau_S(t), \mathrm{Tr}_E\left[\left(\sum_k e^{-\mathbb{i}\,E_k t}\xi(k)_S\otimes\hat{\xi}(k)_E\right)(\phi_S\otimes\pi_E)\left(\sum_l e^{+\mathbb{i}\,E_l t}\xi(l)_S\otimes\hat{\xi}(l)_E\right)\right]\right\}\\
&\quad\geq\left(\frac{1}{d_E}\sum_k\delta_{\phi,\xi(k)}f_k^2-\frac{1}{d_E}\sum_k\delta_{\phi,\xi(k)}\left(1-f_k^2\right)\right)^2\\
&\quad=\left(\frac{1}{d_E}\sum_k\delta_{\phi,\xi(k)}\left(2f_k^2-1\right)\right)^2 \qquad (9.89)
\end{aligned}
$$

Applying the definition of $\delta(\phi)$ and the bijectivity of the mapping we finally obtain

$$
\begin{aligned}
&F^2\left\{\tau_S(t), \mathrm{Tr}_E\left[\left(\sum_k e^{-\mathbb{i}\,E_k t}\xi(k)_S\otimes\hat{\xi}(k)_E\right)(\phi_S\otimes\pi_E)\left(\sum_l e^{+\mathbb{i}\,E_l t}\xi(l)_S\otimes\hat{\xi}(l)_E\right)\right]\right\}\\
&\quad\geq\left(\frac{1}{d_E}\left(2\delta(\phi)^2-1\right)\sum_k\delta_{\phi,\xi(k)}\right)^2\\
&\quad=\left(2\delta(\phi)^2-1\right)^2\ . \qquad (9.90)
\end{aligned}
$$

As for the second part of the proof,

$$
\begin{aligned}
&\mathrm{Tr}_E\left[\left(\sum_k e^{-\mathbb{i}\,E_k t}\xi(k)_S\otimes\hat{\xi}(k)_E\right)(\phi_S\otimes\pi_E)\left(\sum_l e^{+\mathbb{i}\,E_l t}\xi(l)_S\otimes\hat{\xi}(l)_E\right)\right]\\
&\quad=\sum_{kl}\frac{1}{d_E}e^{-\mathbb{i}(E_k-E_l)t}\delta_{\phi,\xi(k)}\delta_{\phi,\xi(l)}\delta_{\hat{\xi}(k),\hat{\xi}(l)}|\xi(k)\rangle\langle\xi(l)|_S\\
&\quad=\sum_{kl}\frac{1}{d_E}e^{-\mathbb{i}(E_k-E_l)t}\delta_{\phi,\xi(k)}\delta_{\xi(k),\xi(l)}\delta_{\hat{\xi}(k),\hat{\xi}(l)}\phi_S \qquad (9.91)
\end{aligned}
$$

Applying (9.80) for the first equality and the bijectivity for the second this simplifies to

$$
\begin{aligned}
&\mathrm{Tr}_E\left[\left(\sum_k e^{-\mathbb{i}\,E_k t}\xi(k)_S\otimes\hat{\xi}(k)_E\right)(\phi_S\otimes\pi_E)\left(\sum_l e^{+\mathbb{i}\,E_l t}\xi(l)_S\otimes\hat{\xi}(l)_E\right)\right]\\
&\quad=\sum_k\frac{1}{d_E}\delta_{\phi,\xi(k)}\phi_S\\
&\quad=\phi_S\ . \qquad (9.92)
\end{aligned}
$$

We find

$$F\left\{\tau_S(t), \phi_S\right\} \geq 2\delta(\phi)^2 - 1 \tag{9.93}$$

and by use of (3.31)

$$\begin{aligned} \left\|\tau_S(t) - \phi_S\right\|_1 &\leq 2\sqrt{1 - F(\tau_S(t), \phi_S)^2} \\ &\leq 2\sqrt{1 - \left(2\delta(\phi)^2 - 1\right)^2} \\ &= 4\delta(\phi)\sqrt{1 - \delta(\phi)^2} \end{aligned} \tag{9.94}$$

which is lower than 2 if $\delta(\phi)$ is larger than $\frac{1}{\sqrt{2}}$. □

Appendix A

Various technical lemmas

Lemma A.1. *Let $\rho_A \in \mathcal{S}_{\leq}(\mathcal{H}_A)$ and $0 \leq \Pi_A \leq \mathbb{1}_A$. Then,*

$$H_{\min}(A)_{\Pi_A \rho_A \Pi_A} \geq H_{\min}(A)_{\rho_A} \ . \tag{A.1}$$

Proof. This follows directly from [BCR09, Lemma B.24.]. We make the identification $\mathcal{H}_B \cong \mathrm{span}_{\mathbb{C}} \{|0\rangle_B\}$ and realize that in this case

$$\rho_{AB} = \rho_A \otimes |0\rangle\langle 0|_B \tag{A.2}$$

for every $\rho_{AB} \in \mathcal{S}_{\leq}(\mathcal{H}_{AB})$ and that

$$\sigma_B, \omega_B = |0\rangle\langle 0|_B \tag{A.3}$$

for all $\sigma_B, \omega_B \in \mathcal{S}_{=}(\mathcal{H}_B)$. It follows that with the identifications made the condition

$$\mathbb{1}_A \otimes \omega_B - \Pi_{AB}(\mathbb{1}_A \otimes \sigma_B)\Pi_{AB} \geq 0 \tag{A.4}$$

is automatically fulfilled. The assertion then follows by direct application of the lemma cited. □

Lemma A.2. *[BCR09, Lemma B.25.] Let $\varepsilon > 0$ and $\rho_A \in \mathcal{S}_{\leq}(\mathcal{H}_A)$. Then there exists $0 \leq \Pi_A \leq \mathbb{1}_A$ such that $\rho_A \in \mathcal{B}^{\varepsilon}(\Pi_A \rho_A \Pi_A)$ and*

$$H_{\max}^{\frac{\varepsilon^2}{6}}(A)_{\rho} \geq H_R(A)_{\Pi\rho\Pi} + 2 \cdot \log \frac{\varepsilon^2}{6} \ . \tag{A.5}$$

Lemma A.3. *[AGZ09, Corollary 4.4.28] For a function $f : \mathcal{A} \to \mathcal{B}$ from a set $\mathcal{A}$ to a set $\mathcal{B}$ endowed with distance measures $d_{\mathcal{A}}$ and $d_{\mathcal{B}}$ the* Lipschitz constant *is defined as*

$$L(f) := \sup_{a_1, a_2 \in \mathcal{A}} \frac{d_{\mathcal{B}}\left(f(a_1), f(a_2)\right)}{d_{\mathcal{A}}\left(a_1, a_2\right)} \ . \tag{A.6}$$

For a function $f : \mathbb{U}(\mathbb{C}^d) \to \mathbb{C}$ let $\langle f \rangle_U$ be the Haar measure average of f. Then,

$$\Pr_U \{|f(U) - \langle f \rangle_U| \geq \delta\} \leq 2e^{-d\delta^2/4L(f)^2} \tag{A.7}$$

where the probability is computed for the choice of U from the Haar measure and where the relevant distance measure on $\mathbb{U}(\mathbb{C}^d)$ is $\|\ldots\|_2$.

Lemma A.4. *Let $\rho_{AR} \in \mathcal{S}_=(\mathcal{H}_A \otimes \mathcal{H}_R)$. Then*

$$\int_{\mathbb{U}(A)} U\rho_{AR}U^\dagger \mathrm{d}U = \pi_A \otimes \rho_R \ . \tag{A.8}$$

Proof. Since we are working with finite dimensional Hilbert spaces, we have $\rho_{AR} \in \mathrm{Herm}(\mathcal{H}_A \otimes \mathcal{H}_R) \cong \mathrm{Herm}(\mathcal{H}_A) \otimes \mathrm{Herm}(\mathcal{H}_R)$. Let $\left\{\tau_A^{(i)}\right\}_i$ be a basis of the real vector space $\mathrm{Herm}(\mathcal{H}_A)$ and $\left\{\eta_R^j\right\}_j$ a basis of $\mathrm{Herm}(\mathcal{H}_R)$. We assume that both bases are normalized, i.e. $\mathrm{Tr}_A\, \tau_A^{(i)} = \mathrm{Tr}_R\, \eta_R^{(j)} = 1$. Then ρ_{AR} is uniquely decomposable in the form

$$\rho_{AR} = \sum_{ij} c_{ij} \tau_A^{(i)} \otimes \eta_R^{(j)} \tag{A.9}$$

with $\sum_{ij} c_{ij} = 1$. Every term of the form $\int_{\mathbb{U}(A)} U\tau_A^{(i)}U^\dagger \mathrm{d}U$ commutes with every $V \in \mathbb{U}(A)$ (by use of the invariance of the Haar measure) and hence by Schur's Lemma and the invariance of the trace under conjugation with a unitary

$$\int_{\mathbb{U}(A)} U\tau_A^{(i)}U^\dagger \mathrm{d}U = \pi_A \ . \tag{A.10}$$

We find

$$\begin{aligned} \int_{\mathbb{U}(A)} U\rho_{AR}U^\dagger \mathrm{d}U &= \sum_{ij} c_{ij} \int_{\mathbb{U}(A)} U\tau_A^{(i)}U^\dagger \mathrm{d}U \otimes \eta_R^{(j)} \\ &= \pi_A \otimes \left(\sum_{ij} c_{ij} \eta_R^{(j)} \right) \\ &= \pi_A \otimes \rho_R \end{aligned} \tag{A.11}$$

where the last equality follows from the definition of the partial trace (2.10). □

Lemma A.5. *For $\rho_{ABC} \in \mathcal{S}_{\leq}(\mathcal{H}_{ABC})$ we have*

$$H_{\min}(AB|C)_\rho \leq H_{\min}(A|C)_\rho + \log d_B \ . \tag{A.12}$$

Proof. By definition,

$$H_{\min}(AB|C)_\rho := \sup_{\sigma_C \in \mathcal{S}_=(\mathcal{H}_C)} \sup \left\{ \lambda \in \mathbb{R} : 2^{-\lambda} \mathbb{1}_{AB} \otimes \sigma_C \geq \rho_{ABC} \right\} \ . \tag{A.13}$$

Let $\omega_C \in \mathcal{S}_=(\mathcal{H}_C)$ and $\mu \in \mathbb{R}$ be arguments which achieve the supremum (such arguments exist since we are dealing with finite dimensional Hilbert spaces), so $H_{\min}(AB|C)_\rho = \mu$ and $2^{-\mu} \mathbb{1}_{AB} \otimes \omega_C \geq \rho_{ABC}$. Since the partial trace of a positive semi-definite operator is positive semi-definite again, we have $2^{-\mu} d_B \mathbb{1}_A \otimes \omega_C \geq \rho_{AC}$ and thus

$$\begin{aligned} H_{\min}(A|C)_\rho &= \sup_{\sigma_C \in \mathcal{S}_=(\mathcal{H}_C)} \sup \left\{ \lambda \in \mathbb{R} : 2^{-\lambda} \mathbb{1}_A \otimes \sigma_C \geq \rho_{AC} \right\} \\ &\geq \mu - \log d_B \\ &= H_{\min}(AB|C)_\rho - \log d_B \ . \end{aligned} \tag{A.14}$$

□

Lemma A.6 (Swap trick)**.** *Let $\mathcal{H}_{A'}$ be isomorphic to $\mathcal{H}_A$ with basis vectors $|i\rangle_{A'}$ which are the images of the basisvectors $|i\rangle_A$ under said isomorphism. Define the "SWAP"-operator*

$$\mathbb{S}_{A \leftrightarrow A'} := \sum_i |i\rangle_A \langle i|_{A'} + |i\rangle_{A'} \langle i|_A \in \mathrm{Herm}(\mathcal{H}_A \otimes \mathcal{H}_{A'}) \ . \tag{A.15}$$

Then for all $M, N \in \operatorname{End} \mathcal{H}_A$

$$\mathrm{Tr}(MN) = \mathrm{Tr}_{AA'} \left[(M_A \otimes N_{A'}) \mathbb{S}_{A \leftrightarrow A'} \right] \ . \tag{A.16}$$

Proof.

$$\begin{aligned} \mathrm{Tr}(MN) &= \sum_{ij} \langle i|M|j\rangle_A \langle j|N|i\rangle_A \\ &= \sum_{ij} \langle i|M|j\rangle_A \langle j|N|i\rangle_{A'} \\ &= \mathrm{Tr}_{AA'} \left[(M_A \otimes N_{A'}) \mathbb{S}_{A \leftrightarrow A'} \right] \ . \end{aligned} \tag{A.17}$$

□

Lemma A.7. *Let $\sigma : \mathcal{S}_n \to \mathrm{End}\left[(\mathbb{C}^d)^{\otimes n}\right]$ be the natural action of the symmetric group $\mathcal{S}_n$ which for $\Pi \in \mathcal{S}_n$ defined by*

$$\sigma(\Pi)\left(|e_1\rangle \otimes \ldots \otimes |e_n\rangle\right) := |e_{\Pi^{-1}(i)}\rangle \otimes \ldots \otimes |e_{\Pi^{-1}(n)}\rangle \tag{A.18}$$

Let $M \in \mathrm{End}\left[(\mathbb{C}^d)^{\otimes n}\right]$. Let

$$\Gamma(M) := \int_{\mathbb{U}(\mathbb{C}^d)} U^{\otimes n} M (U^\dagger)^{\otimes n} \mathrm{d}U \tag{A.19}$$

where $\mathrm{d}U$ is the Haar measure on $\mathbb{U}(\mathbb{C}^d)$. Γ is the orthogonal (with respect to the Hilbert-Schmidt inner product $(A, B) = \mathrm{Tr}(A^\dagger B)$) projector onto $\mathrm{span}_{\mathbb{C}}\{\sigma(\Pi) : \Pi \in \mathcal{S}_n\}$. I.e.

$$\Gamma(M) := \sum_{\Pi \in \mathcal{S}_n} \alpha_\Pi \sigma(\Pi) \tag{A.20}$$

where the α_Π are such that $(\sigma(\Pi), M) = (\sigma(\Pi), \Gamma(M))$.

Proof. This is a standard result in Schur-Weyl duality an a rewriting of e.g. [CS06, Proposition 2.2.]. □

Lemma A.8. *For $\rho \otimes \rho \in \mathcal{S}_=(\mathcal{H}_\Omega \otimes \mathcal{H}_{\Omega'})$ we have*

$$\int_{\mathbb{U}(\Omega)} U^{\otimes 2}(\rho \otimes \rho)(U^\dagger)^{\otimes 2} \mathrm{d}U = \frac{d_\Omega - p(\rho)}{d_\Omega^3 - d_\Omega} \cdot \mathbb{1}_{\Omega\Omega'} + \frac{p(\rho) d_\Omega - 1}{d_\Omega^3 - d_\Omega} \cdot \mathbb{S}_{\Omega \leftrightarrow \Omega'} \tag{A.21}$$

where the integral is with respect to the Haar measure and where $\mathbb{S}_{\Omega \leftrightarrow \Omega'}$ denotes the "SWAP"-operator as introduced in Lemma A.6.

Proof. This integral can be solved by use of Lemma A.7. We write

$$\Gamma(\rho \otimes \rho) = \int_{\mathbb{U}(\Omega)} U^{\otimes 2}(\rho \otimes \rho)(U^\dagger)^{\otimes 2} \mathrm{d}U = \alpha \mathbb{1}_{\Omega\Omega'} + \beta \mathbb{S}_{\Omega \leftrightarrow \Omega'} \ . \tag{A.22}$$

By use of the orthogonality of the projection (applied in the second equality below) we find

$$\begin{aligned} 1 &= (\rho \otimes \rho, \mathbb{1}_{\Omega\Omega'}) \\ &= (\Gamma(\rho \otimes \rho), \mathbb{1}_{\Omega\Omega'}) \\ &= \alpha\left(\mathbb{1}_{\Omega\Omega'}, \mathbb{1}_{\Omega\Omega'}\right) + \beta\left(\mathbb{S}_{\Omega \leftrightarrow \Omega'}, \mathbb{1}_{\Omega\Omega'}\right) \\ &= \alpha \sum_{kl=1}^{d_\Omega} \langle kl|kl\rangle_{\Omega\Omega'} + \beta \sum_{kl=1}^{d_\Omega} \langle lk|kl\rangle_{\Omega\Omega'} \\ &= \alpha d_\Omega^2 + \beta d_\Omega \end{aligned} \tag{A.23}$$

and similarly, applying Lemma A.6 in the first equality below,

$$\begin{aligned} p(\rho) &= \mathrm{Tr}\left[(\rho\otimes\rho)\,\mathbb{S}_{\Omega\leftrightarrow\Omega'}\right] \\ &= (\rho\otimes\rho, \mathbb{S}_{\Omega\leftrightarrow\Omega'}) \\ &= (\Gamma(\rho\otimes\rho), \mathbb{S}_{\Omega\leftrightarrow\Omega'}) \\ &= \alpha\,(\mathbb{1}_{\Omega\Omega'}, \mathbb{S}_{\Omega\leftrightarrow\Omega'}) + \beta\,(\mathbb{S}_{\Omega\leftrightarrow\Omega'}, \mathbb{S}_{\Omega\leftrightarrow\Omega'}) \\ &= \alpha\sum_{kl=1}^{d_\Omega}\langle kl|lk\rangle_{\Omega\Omega'} + \beta\sum_{kl=1}^{d_\Omega}\langle lk|lk\rangle_{\Omega\Omega'} \\ &= \alpha d_\Omega + \beta d_\Omega^2\ . \end{aligned} \tag{A.24}$$

Solving the system of equations (A.23) and (A.24) for α and β and inserting the obtained values into (A.22) yields the integral provided in the lemma. □

Lemma A.9. *If*

$$\frac{d}{dt}f(t) \le \alpha K\cdot f(t)^{\frac{\alpha-1}{\alpha}} \tag{A.25}$$

for an $\alpha > 0$ then

$$f(t) \le \left(f(0)^{\frac{1}{\alpha}} + Kt\right)^{\alpha} \tag{A.26}$$

for $t \ge 0$.

Proof. We introduce the auxiliary function

$$h(t) := f(t)^{\frac{1}{\alpha}} - Kt \tag{A.27}$$

which by use of the assumption is monotously decreasing since

$$\frac{\mathrm{d}h(t)}{\mathrm{d}t} = \frac{1}{\alpha}\cdot f(t)^{\frac{1-\alpha}{\alpha}}\cdot\frac{\mathrm{d}}{\mathrm{d}t}f(t) - K \le K - K\ . \tag{A.28}$$

Since $h(0) = f(0)^{\frac{1}{\alpha}}$ we conclude that $h(t) \le f(0)^{\frac{1}{\alpha}}$ for all $t \ge 0$ which is equivalent to the assertion. □

Lemma A.10. *Let $|\phi\rangle\langle\phi|_A \in \mathcal{S}_=(\mathcal{H}_A)$ be a pure state. Then,*

$$H_{\min}^{\varepsilon}(A)_\phi \le \log\frac{1}{1-\varepsilon^2} = \frac{\varepsilon^2}{\ln 2} + \mathcal{O}(\varepsilon^4)\ . \tag{A.29}$$

Proof. For the state $\mu|\phi\rangle\langle\phi|_A \in \mathcal{S}_{\leq}(\mathcal{H}_A)$ with $0 \leq \mu \leq 1$ we have $P(|\phi\rangle\langle\phi|_A, \mu|\phi\rangle\langle\phi|_A) = \sqrt{1-\mu}$ and $H_{\min}(A)_{\mu|\phi\rangle\langle\phi|} = -\log\mu$ which directly yields the assertion. □

Lemma A.11. *[TCR09, Theorem 9] For $\rho_{AB} \in \mathcal{S}_{=}(\mathcal{H}_{AB})$ and $n \in \mathbb{N}$ we have for $n \geq \frac{8}{5}\log\frac{2}{\varepsilon^2}$*

$$\frac{1}{n}H^{\varepsilon}_{\min}(A^n|B^n)_{\rho^{\otimes n}} \geq H(A|B)_\rho - \frac{1}{\sqrt{n}}\cdot 4\sqrt{\log\left(\frac{2}{\varepsilon^2}\right)}\cdot\log\left(2^{-\frac{1}{2}H_{\min}(A|B)_\rho} + 2^{\frac{1}{2}H_{\max}(A|B)_\rho} + 1\right) . \tag{A.30}$$

Remark: The quantity $\Upsilon(A|B)_{\rho|\rho}$ appearing in the original version of [TCR09, Theorem 9] has been replaced by use of the inequality between Definition 4 and Lemma 8 on p. 5 of [TCR09].

Lemma A.12. *For $\rho_A \in \mathcal{S}_{=}(\mathcal{H}_A)$ we have for $n \geq \frac{8}{5}\log\frac{2}{\varepsilon^2}$*

$$\frac{1}{n}H^{\varepsilon}_{\max}(A^n)_{\rho^{\otimes n}} \leq H(A)_\rho + \frac{1}{\sqrt{n}}\cdot 4\sqrt{\log\frac{2}{\varepsilon^2}}\cdot\log\left(\sqrt{d_A}+2\right) . \tag{A.31}$$

Proof. Let ϕ_{AP} be a purification of ρ_A. Then by use of (3.42) and Lemma A.11 we have

$$\begin{aligned}\frac{1}{n}H^{\varepsilon}_{\max}(A^n)_{\rho^{\otimes n}} &= -\frac{1}{n}H^{\varepsilon}_{\min}(A^n|P^n)_{\phi^{\otimes n}} \\ &\leq -H(A|P)_\phi + \frac{1}{\sqrt{n}}\cdot 4\sqrt{\log\left(\frac{2}{\varepsilon^2}\right)}\cdot\log\left(2^{-\frac{1}{2}H_{\min}(A|P)_\phi} + 2^{\frac{1}{2}H_{\max}(A|P)_\phi} + 1\right) .\end{aligned} \tag{A.32}$$

The first summand is

$$-H(A|P)_\phi = -H(AP)_\phi + H(P)_\phi = H(A)_\phi . \tag{A.33}$$

By use of (3.51) we have

$$H_{\min}(A|P)_\phi \geq -\log d_A \tag{A.34}$$

since we can always find a purifying system P with $d_P = d_A$. Using that

$$\langle\phi|\mathbb{1}_A \otimes \sigma_P|\phi\rangle_{AP} \leq \langle\phi|\mathbb{1}_A \otimes \mathbb{1}_P|\phi\rangle_{AP} = 1 \tag{A.35}$$

for every $\sigma_P \in \mathcal{S}_{=}(\mathcal{H}_P)$ we find from (3.11) that

$$H_{\max}(A|P)_\phi \leq 0 \tag{A.36}$$

which yields the assertion. □

Bibliography

[ADHW06] A. Abeyesinghe, I. Devetak, P. Hayden, A. Winter, *The mother of all protocols: Restructuring quantum informations family tree*, Proceedings of the Royal Society A, **465**:25372563 (2009).

[AGZ09] G. W. Anderson, A. Guionnet, O. Zeitouni, *An Introduction to Random Matrices*, Cambridge University Press 2009

[Ber09] M. Berta, *Single-shot Quantum State Merging*, arXiv:0912.4495v1 (2009).

[BCR09] M. Berta, M. Christandl, R. Renner, *The Quantum Reverse Shannon Theorem based on One-Shot Information Theory*, arXiv:0912.3805v2 (2009).

[BHLS03] C.H. Bennett, A.W. Harrow, D.W. Leung, J.A. Smolin, *On the capacities of bipartite Hamiltonians and unitary gates*, IEEE Transactions on Information Theory **49**, 1895 (2003).

[BCCRR10] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, R. Renner, *The Uncertainty Principle in the Presence of Quantum Memory*, Nat. Phys. **6**, 659 (2010).

[Bha97] R. Bhatia, *Matrix Analysis* Springer-Verlag New York, Inc. 1997.

[Bra07] S. Bravyi, *Upper bounds on entangling rates of bipartite Hamiltonians*, Phys. Rev. Lett. **76**, 052319 (2007).

[CLV03] A. M. Childs, D. W. Leung, G. Vidal, *Reversible simulation of bipartite product Hamiltonians*, IEEE Transactions on Information Theory, **50**, 1189 (2003).

[Cho75] M. D. Choi, *Completely positive linear maps on complex matrices*, Linear algebra and its applications, 10(3):285 (1975).

[CS06] B. Collins, P. Śniady, *Integration with respect to the Haar measure on unitary, orthogonal and symplectic group*, Communications in Mathematical Physics, **264**:773795 (2006).

[dRARDV11] L. del Rio, J. Aberg, R. Renner, O. Dahlsten, V. Vedral, *The thermodynamic meaning of negative entropy*, Nature **474**, 6163 2011

[Dup09] F. Dupuis, *The decoupling approach to quantum information theory*, PhD Thesis, arXiv:1004.1641 (2009).

[DBWR10] F. Dupuis, M. Berta, J. Wullschleger, R. Renner, *The Decoupling Theorem*, arXiv:1012.6044v1 (2010).

[FvdG99] C. Fuchs, J. van de Graaf, *Cryptographic distinguishability measures for quantum-mechanical states*, IEEE Transactions on Information Theory, 45(4):1216-1227, (1999).

[GM02] J. Gemmer, G. Mahler, *Distribution of local entropy in the Hilbert space of bi-partite quantum systems: Origin of Jaynes' principle*, arXiv:quant-ph/0201136v1 (2002).

[GMM09] J. Gemmer, M. Michel, G. Mahler, *Quantum Thermodynamics*, vol. 784. Springer, Berlin / Heidelberg (2000).

[Gog10a] C. Gogolin, *Einselection without pointer states*, arXiv:0908.2921v4 (2010).

[Gog10b] C. Gogolin, *Pure State Quantum Statistical Mechanics*, arXiv:1003.5058 (2010).

[GME11] C. Gogolin, M. P. Müller, J. Eisert, *Absence of Thermalization in Nonintegrable Systems*, Phys. Rev. Lett. **106**, 040401 (2011).

[GLTZ06] S. Goldstein, J. L. Lebowitz, R. Tumulka, N. Zanghi, *Canonical Typicality*, Phys. Ref. Lett. **96**, 050403 2006.

[Has10] M. B. Hastings, *Locality in Quantum Systems*, arXiv:1008.5137v1 (2010).

[HP07] P. Hayden, J. Preskill, *Black holes as mirrors: quantum information in random subsystems*, arXiv:0708.4025v2 (2007).

[HL09] A.W. Harrow, R.A. Low, *Random Quantum Circuits are Approximate 2-designs*, Communications in Mathematical Physics **291**, 257302 (2009).

[Hel69] C. W. Helstrom, *Quantum detection and estimation theory*, Journal of Statistical Physics, 1(2):231252, June 1969 (1969).

[HGJ11] H. Hinrichsen, C. Gogolin, P. Janotta, *Non-equilibrium Dynamics, Thermalization and Entropy Production*, arXiv:1102.0103v1 (2011).

[HJ05] A. R. Horn, C. R. Johnson, *Matrix analysis*, Cambridge University Press (2005).

[HOW05] M. Horodecki, J. Oppenheim, A. Winter, *Partial quantum information*, Nature **436**, 673-676 (2005).

[Jam72] A. Jamiołkowski, *Linear transformations which preserve trace and positive semidefiniteness of operators*, Reports on Mathematical Physics, 3(4):275278 (1972).

[Jay57] E.T. Jaynes, *Information Theory and Statistical Mechanics*, Phys. Ref. **106**(4), Reports on Mathematical Physics, 3(4), 620-630 (1957).

[GHH07] G. Kimura, H. Ohno, H. Hayashi, *How to detect a possible correlation from the information of a sub-system in quantum mechanical systems*, Phys. Rev. A **76** 042123 (2007).

[Kit58] C. Kittel, *Elementary Statistical Physics*, John Wiley & Sons, Inc. (1985).

[KRS09] R. König, R. Renner, C. Schaffner, *The operational meaning of min- and max-entropy*, IEEE Transactions on Information Theory, 55(9), (2009).

[LPSW09] N. Linden, S. Popescu, A. Short, A. Winter, *Quantum mechanical evolution towards thermal equilibrium*, Phys. Rev. E **79**, 061103 (2009).

[LPSW10] N. Linden, S. Popescu, A. Short, A. Winter, *On the speed of fluctuations around thermodynamic equilibrium*, New J. Phys. **12**, 055021 (2010).

[ML97] N. Margolus, L. B. Levitin, *The maximum speed of dynamical evolution*, arXiv:quant-ph/9710043v2 (1997).

[MS01] V. D. Milman, G. Schechtman, *Asymptotic Theory of Finite Dimensional Normed Spaces*, Springer Verlag, LNM 1200, Berlin (2001).

[Nie97] M. A. Nielsen, *Computable functions, quantum measurements, and quantum dynamics*, Phys. Rev. Lett. **79**, 29152918 (1997).

[NC00] M. A. Nielsen, I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, New York (2000).

[PSW06a] S. Popescu, A. Short, A. Winter, *The foundations of statistical mechanics from entanglement: Individual states vs. averages*, arXiv:quant-ph/0511225v3 (2006).

[PSW06b] S. Popescu, A. Short, A. Winter, *Entanglement and the foundations of statistical mechanics*, Nature Physics **2**, 754-758 (2006).

[Ren05] R. Renner, *Security of quantum key distribution*, PhD Thesis, arXiv:quant-ph/0512258 (2005).

[RKIA11] C. A. Rodriguez-Rosario, G. Kimura, H. Imai, A. Aspuru-Guzik, *Lazy states: sufficient and necessary condition for zero quantum entropy rates under any coupling to the environment*, Phys. Rev. Lett. **106**, 050403 (2011).

[RW04] R. Renner, S. Wolf, *Smooth Rényi entropy and applications*, Proceedings of the 2004 International Symposium on Information Theory (2004).

[RGE11] A. Riera, C. Gogolin, J. Eisert, *Thermalization in nature and on a quantum computer*, arXiv:1102.2389v1 (2011).

[Sha48] C. E. Shannon, *A Mathematical Theory of Communication*, The Bell System Technical Journal **27** (1948).

[Sho11] A. Short, *Equilibration of quantum systems and subsystems*, New J. Phys. **13** 053009 (2011).

[Sti55] W. F. Stinespring, *Positive functions on C*-algebras*, Proceedings American Mathematical Society, 6:211 (1955).

[Sze11] O. Szehr, *Decoupling Theorems*, Master's Thesis, ETH Zürich (2011).

[Tas98] H. Tasaki, *From Quantum Dynamics to the Canonical Distribution: General Picture and a Rigorous Example*, Phys. Ref. Lett. **80**, 1373-1376 (1998).

[TCR09] M. Tomamichel, R. Colbeck, R. Renner, *A fully quantum asymptotic equipartition property*, IEEE Trans. Inf. Theory **55**, 5840 (2009).

[TCR10] M. Tomamichel, R. Colbeck, R. Renner, *Duality Between Smooth Min- and Max-Entropies*, IEEE Trans. Inf. Theory **56**, 4674 (2010).

[Uhl76] A. Uhlmann, *The transition probability in the state space of a *-algebra*, Reports on Mathematical Physics, 9:273 (1976).

[vDH02] W. van Dam, P. Hayden, *Renyi-entropic bounds on quantum communication*, arXiv:quant-ph/0204093v1 (2002).