The Coherent Relative Entropy and the Work Cost of Quantum Processes

THESIS

submitted in partial fulfillment of the requirements for the degree of

BACHELOR OF SCIENCE
in
PHYSICS AND MATHEMATICS

Author :	Daan Otto
Student ID :	1817639
Supervisor MI :	Prof.dr. S. Fehr
Supervisor LION :	Dr. W. Löffler

Leiden, The Netherlands, June 19, 2020

The Coherent Relative Entropy and the Work Cost of Quantum Processes

Daan Otto

Instituut-Lorentz, Leiden University
P.O. Box 9500, 2300 RA Leiden, The Netherlands

June 19, 2020

Abstract

Information theory finds more and more applications within physics. Here, we look at the coherent relative entropy from a mathematical and physical perspective. We exploit the relation between this recently introduced entropy measure and the Rényi divergence of order infinity to show that properties of the coherent relative entropy follow almost directly from properties of the Rényi divergence. Besides, we discuss that properties of a new generalized coherent relative entropy also follow from these properties of the Rényi divergence. On the physics side we present a short survey on the relation between information theory and physics, where we discuss the Szilard engine and Landauer's principle, and we then explain the physical meaning of the coherent relative entropy in this context.

Contents

Chapter 1

Introduction

In 1961 Rolf Landauer pointed out in his paper [1] that information behaves like a physical concept. This result gave a solution to the infamous Maxwell's demon [2] and resulted in new studies where information theory and thermodynamics are considered to go hand in hand. Nowadays, researchers study this relation and its implications in the quantum regime by uniting quantum information theory and quantum thermodynamics [3–6]. This thesis will focus on the recently-introduced so-called coherent relative entropy and its interpretation as the work cost of quantum thermodynamic processes.

Background

In Refs. 6, 7 a new quantum theoretical measure named coherent relative entropy was introduced. Mathematically speaking, it is an entropy measure of a so-called CPTNI map $\mathcal{E}_{A\to B}$ applied to a state σ given as a density matrix. The map goes from input system A to output system B which both are characterised by a positive semi-definite operator. In Refs. 6, 7 they prove several properties of the coherent relative entropy. These are properties we expect an entropy measure to hold, for example the data processing inequality. In Refs. 6, 7 they prove these properties from scratch using non-trivial techniques, e.g., using semi-definite programming. This approach results in technical and lengthy proofs.

Next to introducing this new entropy measure and proving various properties, the main result of Refs. 6, 7 is to discuss and formally show the physical meaning of this new measure. The CPTNI map $\mathcal{E}_{A\to B}$ is applied to the input state σ with quantum thermodynamic systems as input and output systems. They showed that the coherent relative entropy captures the maximum amount of work extracted or the minimum amount of work needed when physically performing the map $\mathcal{E}_{A\to B}$ on an input state σ.

Our contributions

Here, we present a mathematical and physical contribution to the research field. Regarding our mathematical contribution we provide new quantum information theoretic insights into the coherent relative entropy. Our starting point is a connection between the coherent relative entropy to the Rényi Divergence of order infinity. This connection was also noted in Refs. 6, 7, yet they did not focus or elaborate much on this. In this thesis, we exploit this connection to provide the following two contributions. First, we extend this connection to the general Rényi divergence of order α, where the coherent relative entropy considered in Refs. 6, 7 corresponds to the case $\alpha = \infty$. This lets us introduce the general coherent relative entropy of order α. Secondly, this connection lets us re-prove some of the properties of the coherent relative entropy, as considered and proven in Refs. 6, 7 but now: (1) by means of simpler proofs that exploit corresponding properties of the Rényi divergence and (2) for our generalized version of the coherent relative entropy.

On the physics side, our contributions are as follows. We first give a self-contained introduction survey explaining that information behaves like a physical entity. We do this by discussing the Szilard engine [8], a thought experiment that in principle explains that information can be transformed into work. Also we present Landauer's principle [1] which states that erasing information costs work. We extend these notions from classical to quantum physics. Moreover, we present the recently introduced idea of treating a quantum register as a work storage system by identifying the quantum information it contains with the amount of work this information can be transformed into by identifying the quantum information it contains with the amount of work this information can be transformed into. We do this to eventually present the physical meaning of the coherent relative entropy in this context. The coherent relative entropy tells us the work cost or the amount of work extracted when a CPTNI map is physically performed. This work can then respectively be extracted from or stored in the work storage system. Finally, we give an example of the physical meaning of the coherent relative entropy in this context.

Structure

The structure of this thesis is as follows. Chapters 2 and 3 discuss some preliminary concepts of information theory which will be used throughout this thesis. In Chapter 4 we present the coherent relative entropy and our main results. We use the connection between the coherent relative entropy and the Rényi divergence to introduce the brand-new generalized coherent relative entropy and deduce its properties from those of the Rényi divergence. In Chapters 5 and 6, we move on to the part on physics in this thesis. We give the necessary background knowledge to ultimately discuss the physical relevance of the coherent relative entropy.

Chapter 2

Preliminary Quantum Information Theory

This chapter will explain some relevant concepts of quantum information theory. These concepts are necessary for the understanding of further claims and results discussed in this thesis. It is assumed that the reader has sufficient knowledge of linear algebra, yet the core concepts will be revisited within the context of quantum information theory. Most of these contents can be found in *Quantum Computation and Quantum Information* by Michael A. Nielsen and Isaac L. Chuang [9] and *Quantum Information Processing with Finite Resources* by Marco Tomamichel [10]. These references also form a good basis for those looking for more in-depth information.

2.1 Linear Algebra

Let $\mathcal{H}$ be a finite dimensional Hilbert space over the field of complex numbers $\mathbb{C}$. Elements of $\mathcal{H}$ are vectors and are notated as *ket-vectors*, $|\phi\rangle$. Given such a Hilbert space $\mathcal{H}$ there is the dual vector space $\mathcal{H}^* = \{f : \mathcal{H} \to \mathbb{C} \mid f \text{ linear}\}$. Elements of $\mathcal{H}^*$ are denoted as *bra-vectors*, $\langle\phi|$. If an orthonormal basis of $\mathcal{H}$ is chosen it is natural to think of ket-vectors and bra-vectors as columns vectors and row vectors respectively. Then, bra- and ket-vectors are related via the conjugate transpose, $|\phi\rangle = \langle\phi|^\dagger$. From this we have an intuition of matrix multiplication and the *inner product* naturally emerges $(|\phi\rangle, |\psi\rangle) = |\phi\rangle^\dagger|\psi\rangle = \langle\phi|\psi\rangle \in \mathbb{C}$. To be more concrete, we give an example where we let $\dim(\mathcal{H}) = d$. We can write ket-vectors as

$$|\phi\rangle = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{pmatrix} \in \mathbb{C}^d \qquad \text{and} \qquad |\psi\rangle = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_d \end{pmatrix} \in \mathbb{C}^d \ .$$

The inner product is

$$(|\phi\rangle, |\psi\rangle) = \begin{pmatrix} \overline{a_1} & \overline{a_2} & \dots & \overline{a_d} \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_d \end{pmatrix} = \sum_{i=1}^{d} \overline{a_i} \cdot b_i \ .$$

In a similar way, we define the *outer product* of $|\phi\rangle \in \mathcal{H}'$ and $|\psi\rangle \in \mathcal{H}$ as $|\phi\rangle\langle\psi| \in \mathcal{L}(\mathcal{H},\mathcal{H}')$. Here $\mathcal{L}(\mathcal{H},\mathcal{H}')$ is the set of all linear maps from $\mathcal{H}$ to $\mathcal{H}'$. Elements of this set are referred to as *operators*. When $\mathcal{H} = \mathcal{H}'$, we write $\mathcal{L}(\mathcal{H},\mathcal{H}) = \mathcal{L}(\mathcal{H})$ for the set of all linear maps from $\mathcal{H}$ to itself and its identity element will be denoted by $\mathbb{I}$. In addition, one can consider the vector space of *superoperators*, $\mathcal{L}(\mathcal{L}(\mathcal{H}),\mathcal{L}(\mathcal{H}'))$. When $\mathcal{H} = \mathcal{H}'$ with denote the identity element of $\mathcal{L}(\mathcal{L}(\mathcal{H}),\mathcal{L}(\mathcal{H}'))$ as *id*. When an orthonormal basis is chosen, it is natural to think of an operator $R \in \mathcal{L}(\mathcal{H})$ as a matrix, $R \in \mathbb{C}^{\dim(\mathcal{H})\times\dim(\mathcal{H})}$.

Before we define an orthonormal basis within this context, we introduce the set of *state vectors*. The set of elements with norm one is the set of state vectors which is denoted by $\mathcal{S}(\mathcal{H}) := \{|\phi\rangle \mid \sqrt{|\phi\rangle\langle\phi|} = 1\}$. A collection of state vectors $\{|i\rangle\}_{i\in I}$ is an *orthonormal basis* of $\mathcal{H}$ if

$$\sum_{i\in I} |i\rangle\langle i| = \mathbb{I} \; .$$

In an orthonormal basis all elements thus have norm one and are orthogonal to each other. When we look at two-dimensional Hilbert space $\mathcal{H}$, which we may assume to be $\mathcal{H} = \mathbb{C}^2$, we often use the orthonormal basis consisting of

$$|0\rangle := \begin{pmatrix}1\\0\end{pmatrix} \qquad \text{and} \qquad |1\rangle := \begin{pmatrix}0\\1\end{pmatrix} \; .$$

Next, we briefly recall certain operator properties, and some relations among these properties. An operator $R \in \mathcal{L}(\mathcal{H})$ is *positive semi-definite* - notation $R \geq 0$ - if for all $|\phi\rangle \in \mathcal{H}$ we have $\langle\phi|R|\phi\rangle \geq 0$. We denote $\mathcal{P}(\mathcal{H})$ as the set of all positive semi-definite operators in $\mathcal{L}(\mathcal{H})$. Positive semi-definite operators are *Hermitian* as well. An operator R is Hermitian if $R^\dagger = R$. If $RR^\dagger = R^\dagger R$, then R is *normal*. Besides, for two $L, R \in \mathcal{L}(\mathcal{H})$ the *Loewner order* is defined as $L \geq R$ meaning $L - R \geq 0$. An operator $V \in \mathcal{L}(\mathcal{H},\mathcal{H}')$ with $\dim(\mathcal{H}) \leq \dim(\mathcal{H}')$ is called an *isometry* if $V^\dagger V = \mathbb{I}$. If we require $\mathcal{H}' = \mathcal{H}$, we say V is a *unitary* operator, in which case it also holds that $VV^\dagger = \mathbb{I}$.

When using the bra-ket notation, it is convenient to introduce the *trace* as a map that sends the outer product to the inner product.

Definition 1. *The trace map is the unique linear map* $\mathrm{tr} : \mathcal{L}(\mathcal{H}) \to \mathbb{C}$ *such that for all* $|\phi\rangle \in \mathcal{H}$ *and for all* $\overline{\langle\phi|} \in \mathcal{H}^*$ *the following equality holds:*

$$\mathrm{tr}\left(|\phi\rangle\langle\psi|\right) = \langle\psi|\phi\rangle \; .$$

It is not too hard to see that this definition is well-defined. Note that the trace is cyclic, i.e. for all operators $R, L \in \mathcal{L}(\mathcal{H})$ we have the equality $\mathrm{tr}(RL) = \mathrm{tr}(LR)$. Since the trace is cyclic it is invariant under unitary similarity transformations. For U a unitary operator we have

$$\mathrm{tr}\left(URU^\dagger\right) = \mathrm{tr}\left(U^\dagger UR\right) = \mathrm{tr}\left(R\right) \; .$$

This definition of the trace map coincides with the common definition of the trace for operators. Let $R \in \mathcal{L}(\mathcal{H})$ and $\{|i\rangle\}_{i\in I}$ an orthonormal basis of $\mathcal{H}$, then the following familiar notion of the trace map is recovered:

$$\mathrm{tr}(R) = \mathrm{tr}(R\cdot\mathbb{I}) = \mathrm{tr}\left(\sum_{i\in I} R|i\rangle\langle i|\right) = \sum_{i\in I}\langle i|R|i\rangle \; .$$

Given a operator $R \in \mathcal{L}(\mathcal{H})$ there are a few more concepts important to mention:

- The *kernel* of R, $\ker(R) := \{|\phi\rangle \in \mathcal{H} \mid R|\phi\rangle = 0\}$;
- The *support* of R, $\mathrm{supp}(R) := \{|\phi\rangle \in \mathcal{H} \mid \langle\phi|\psi\rangle = 0, \forall|\psi\rangle \in \ker(R)\}$;
- The *rank* of R, $\mathrm{rk}(R) := \dim(\mathrm{supp}(R))$.

2.1.1 Functions on Operators

Whenever $R \in \mathcal{L}(\mathcal{H})$ is normal there exists a *spectral decomposition* of R:

$$R = \sum_{i=1}^{\dim(\mathcal{H})} \lambda_i |i\rangle\langle i| \ ,$$

Where $\lambda_1, \dots \lambda_{\dim(\mathcal{H})} \in \mathbb{C}$ are the eigenvalues of R, and $\{|i\rangle\}_{i \in I}$ is an orthonormal basis consisting of the corresponding normalized eigenvectors. Note that R is normal if and only if it is diagonalizable, so the existence of a spectral decomposition of R is equivalent to R being diagonalizable. We observe that in spectral decomposition R is expressed as the sum of its eigenvalues which in principle can be simplified by taking only the sum over the non-zero eigenvalues, because this does not change the outcome of that sum.

Now, given an function $f : \mathbb{C} \supseteq D \to \mathbb{C}$, we use spectral decomposition of an operator R to specify $f(R)$.

Definition 2. *Let $f : \mathbb{C} \supseteq D \to \mathbb{C}$ be a function and R a normal operator with spectral decomposition $R = \sum_{i=1}^{\dim(\mathcal{H})} \lambda_i |i\rangle\langle i|$ and eigenvalues λ_i in D. We define the the function f on operator R as*

$$f(R) := \sum_{i=1}^{\dim(\mathcal{H})} f(\lambda_i) |i\rangle\langle i| \ .$$

An example of such function is $f : \mathbb{C} \setminus \{0\} \to \mathbb{C}$ given by $f(R) = R^s$ for a $s \in \mathbb{Z}$. This function is consistent of our natural understanding of raising a normal matrix to the power of an integer when considering the non-zero eigenvalues of R. For $R \geq 0$ we generalize this function by letting $s \in \mathbb{R}$, while still only looking at the non-zero eigenvalues.

2.1.2 The Tensor Product

Up to this point we only considered one Hilbert space, we extend this by looking at multiple Hilbert spaces at once. Therefore we need to introduce the *tensor product* of vector spaces. Let $\mathcal{H}_A, \mathcal{H}_B$ be Hilbert spaces. Treating these as vector spaces, a new vector space emerges: the tensor product denoted as $\mathcal{H}_A \otimes \mathcal{H}_B$. Elements of the tensor product are finite linear combinations of vectors $|\phi_A\rangle \otimes |\phi_B\rangle$ with $|\phi_A\rangle \in \mathcal{H}_A$ and $|\phi_B\rangle \in \mathcal{H}_B$. From now on often refer to $\mathcal{H}_A$ as (quantum) system A, $\mathcal{H}_B$ as (quantum) system B and $\mathcal{H}_A \otimes \mathcal{H}_B$ as (bipartite quantum) system AB.

There are several properties of the tensor product which are important to note beforehand. For all $|\phi_A\rangle, |\psi_A\rangle \in \mathcal{H}_A, |\phi_B\rangle, |\psi_B\rangle \in \mathcal{H}_B$ and $\lambda \in \mathbb{C}$:

- $(|\phi_A\rangle + |\psi_A\rangle) \otimes |\phi_B\rangle = |\phi_A\rangle \otimes |\phi_B\rangle + |\psi_A\rangle \otimes |\phi_B\rangle$;
- $|\phi_A\rangle \otimes (|\phi_B\rangle + |\psi_B\rangle) = |\phi_A\rangle \otimes |\phi_B\rangle + |\phi_A\rangle \otimes |\psi_B\rangle$;

- $(\lambda|\phi_A\rangle) \otimes |\phi_B\rangle = \lambda\left(|\phi_A\rangle \otimes |\phi_B\rangle\right) = |\phi_A\rangle \otimes (\lambda|\phi_B\rangle)$.

Furthermore, the tensor product of operators naturally acts component wise on the tensor product of ket-vectors, i.e.

$$(R \otimes L)(|\phi_A\rangle \otimes |\phi_B\rangle) = R|\phi_A\rangle \otimes L|\phi_B\rangle \ .$$

This relation induces an isomorphism $\mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B) \cong \mathcal{L}(\mathcal{H}_A) \otimes \mathcal{L}(\mathcal{H}_B)$. Similarly, the tensor product of bra-vectors also acts component wise on the tensor product of operators and on the tensor product of ket-vectors. This first relation induces an isomorphism as well, namely $(\mathcal{H}_A \otimes \mathcal{H}_B)^* \cong \mathcal{H}_A^* \otimes \mathcal{H}_B^*$. We use these relations to deduce the following equalities:

$$(\langle\psi_A| \otimes \langle\psi_B|)(R \otimes L) = \langle\psi_A|R \otimes \langle\psi_B|L \ ,$$
$$(\langle\psi_A| \otimes \langle\psi_B|)(|\phi_A\rangle \otimes |\phi_B\rangle) = \langle\psi_A|\phi_A\rangle \otimes \langle\psi_B|\phi_B\rangle = \langle\psi_A|\phi_A\rangle \cdot \langle\psi_B|\phi_B\rangle \ ,$$

where in the second relation we use the trivial isomorphism, $\mathbb{C} \otimes V \cong V$ given by $\alpha \otimes v \mapsto \alpha \cdot v$ for V a vector space. Hence, there is also the following relation:

$$(R \otimes \langle\psi_B|)(|\phi_A\rangle \otimes |\phi_B\rangle) = R|\phi_A\rangle \otimes \langle\psi_B|\phi_B\rangle = R|\phi_A\rangle \cdot \langle\psi_B|\phi_B\rangle \ .$$

Recall that the trace sends the inner product to the outer product, so this last isomorphism can be used to deduce that

$$\operatorname{tr}(|\phi_A\rangle\langle\psi_A| \otimes |\phi_B\rangle\langle\psi_B|) = \operatorname{tr}(|\phi_A\rangle\langle\psi_A|) \cdot \operatorname{tr}(|\phi_B\rangle\langle\psi_B|) \ .$$

From linearity it follows that $\operatorname{tr}(R \otimes L) = \operatorname{tr}(R) \cdot \operatorname{tr}(L)$.

Later, we will use a property of the tensor product of operators, namely that exponentiation works component wise.

Lemma 1. *Let $R_A \in \mathcal{P}(\mathcal{H}_A), R_B \in \mathcal{P}(\mathcal{H}_B)$ and $p \in \mathbb{R}$. Then, we have the equality $(R_A \otimes R_B)^p = R_A^p \otimes R_B^p$.*

Proof. Consider the following spectral decompositions $R_A = \sum_i \lambda_i |e_i\rangle\langle e_i|$ and $R_B = \sum_j \mu_j |f_j\rangle\langle f_j|$ where all $\lambda_i, \mu_j > 0$. Spectral decomposition of the tensor product is then given by

$$R_A \otimes R_B = \sum_{i,j} \lambda_i \mu_j \left(|e_i\rangle \otimes |f_j\rangle\right)\left(\langle e_i| \otimes \langle f_j|\right) \ .$$

Raising this to the power of p can be seen as an function as discussed in Definition 2.

$$\begin{aligned}(R_A \otimes R_B)^p &= \sum_{i,j} \lambda_i^p \mu_j^p \left(|e_i\rangle \otimes |f_j\rangle\right)\left(\langle e_i| \otimes \langle f_j|\right) \\ &= \sum_i \lambda_i^p |e_i\rangle\langle e_i| \otimes \sum_j \mu_j^p |f_j\rangle\langle f_j| = R_A^p \otimes R_B^p \ .\end{aligned}$$

□

2.2 Density operators

Definition 3. *An operator $\rho \in \mathcal{L}(\mathcal{H})$ is called a density operator or density matrix if it is positive semi-definite operator, $\rho \geq 0$ and $\operatorname{tr}(\rho) = 1$. The set of density operators is denoted by $\mathcal{D}(\mathcal{H})$.*

When looking at a spectral decomposition $\rho = \sum_i \lambda_i |i\rangle\langle i|$ this definition translates into ρ being a density operator if $\sum_i \lambda_i = 1$ and $\lambda_i = \langle i|\rho|i\rangle \geq 0$ for all i. A density operator ρ is called *pure* if there is a state vector $|\phi\rangle \in \mathcal{S}(\mathcal{H})$ such that $\rho = |\phi\rangle\langle\phi|$, i.e. the density operator can be represented by one state vector. In physics, a pure density operator is the mathematical object to describe a deterministic quantum state. Besides, in spectral decomposition of ρ all but one of the λ_i are equal to zero. Not all density operators are pure, such density operators are called *mixed*. When a density operator is equal to $\mathbb{I}/\dim(\mathcal{H})$ we say it is *maximally mixed*.

We give an example of a pure density operator that is given by the following state vector

$$|\phi\rangle = \frac{2}{\sqrt{5}}|0\rangle + \frac{1}{\sqrt{5}}|1\rangle = \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} , \qquad |\phi\rangle\langle\phi| = \begin{pmatrix} 4/5 & 2/5 \\ 2/5 & 1/5 \end{pmatrix} .$$

A density operator is a *normalized operator*, it has trace one. Later on we will come across *sub-normalized* states as well. The set of sub-normalized states is denoted as

$$\mathcal{D}_{\leq}(\mathcal{H}) := \{\rho \in \mathcal{P}(\mathcal{H}) \mid 0 < \mathrm{tr}(\rho) \leq 1\} .$$

Now, we consider a density operator in the tensor product of two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$. Imagine you want to look only at system B. This can be done by *tracing out* system A and looking at the *reduced* density operator. Formally, when considering a density operator $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ as a description of a state in the bipartite quantum system AB it is possible to describe only the subsystem B.

Definition 4. *Let $\mathcal{H}_A$ and $\mathcal{H}_B$ Hilbert spaces. The partial trace tr_A is a superoperator defined as:*

$$\mathrm{tr}_A := \mathrm{tr} \otimes id_B : \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathbb{C} \otimes \mathcal{L}(\mathcal{H}_B) \cong \mathcal{L}(\mathcal{H}_B) ,$$

where the last isomorphism is naturally given by $\alpha \otimes R_B = \alpha \cdot R_B$. In a similar way the partial trace tr_B is defined as

$$\mathrm{tr}_B := id_A \otimes \mathrm{tr} : \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_A) \otimes \mathbb{C} \cong \mathcal{L}(\mathcal{H}_A) .$$

Take for example the pure density operator $\rho_{AB} = |\phi\rangle\langle\phi|$ given by the ket-vector $|\phi\rangle = \sum_i \alpha_i |e_i\rangle \otimes |f_i\rangle = \sum_i \alpha_i |e_i\rangle|f_i\rangle$, where the last equality is notation. In this context, $\{|e_i\rangle\}$ is an orthonormal basis of $\mathcal{L}(\mathcal{H}_A)$. Tracing out system A yields

$$\begin{aligned} \mathrm{tr}_A(\rho_{AB}) &= \mathrm{tr}_A \left(\sum_{i,j} \alpha_i \overline{\alpha_j} |e_i\rangle|f_i\rangle\langle e_j|\langle f_j| \right) = \sum_{i,j} \alpha_i \overline{\alpha_j}\, \mathrm{tr}(|e_i\rangle\langle e_j|) \otimes |f_i\rangle\langle f_j| \\ &= \sum_{i,j} \alpha_i \overline{\alpha_j} \langle e_j|e_i\rangle \cdot |f_i\rangle\langle f_j| = \sum_{i,j} \alpha_i \overline{\alpha_i} \langle e_i|e_i\rangle \cdot |f_i\rangle\langle f_i| = \sum_{i,j} \alpha_i^2 |f_i\rangle\langle f_i| , \end{aligned}$$

where in the last two equalities we used that the $|e_i\rangle$ form an orthonormal basis. Notice that it is quite easy to show that $\rho_B := \mathrm{tr}_A(\rho_{AB})$ is a density operator as well, referred to as the reduced density operator.

2.2.1 Completely Positive Trace Preserving (CPTP) Maps

In order for a superoperator $\mathcal{E} \in \mathcal{L}(\mathcal{L}(\mathcal{H}), \mathcal{L}(\mathcal{H}'))$ to describe a valid quantum evolution, it has to map density operators into density operators. This gives notion to the following slightly stronger condition which is necessary and sufficient.

Definition 5. *A superoperator $\mathcal{E} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_{A'}))$ is called a CPTP map if both of the following properties hold:*

1. *$\mathcal{E}$ is Completely Positive, i.e. for all Hilbert spaces $\mathcal{H}_B$ and operators $R_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ we have*
$$R_{AB} \geq 0 \Rightarrow \mathcal{E}(R_{AB}) = \mathcal{E} \otimes id_B(R_{AB}) \geq 0\ ;$$
2. *$\mathcal{E}$ is Trace Preserving, i.e. for all $R_A \in \mathcal{L}(\mathcal{H}_A)$ we have that*
$$\mathrm{tr} \circ \mathcal{E}(R_A) = \mathrm{tr}(R_A)\ .$$

Examples of a CPTP map are the partial trace or the following map given by an isometry. Let V be an isometry, the map M_V given by $M_V(\chi) := V\chi V^\dagger$ then is a CPTP map.

When describing quantum processes it is sometimes useful to consider *Completely Positive Trace Non-increasing* maps - referred to as CPTNI maps - instead of CPTP maps. The only difference with Definition 5 is the second requirement which changes into: for all $R_A \geq 0$ we have that $\mathrm{tr}(\mathcal{E}(R_A)) \leq \mathrm{tr}(R_A)$.

Also, it is possible to give a different equivalent definition of CPTP maps. For this we need to point out how a superoperator can be represented as an operator via the *Choi-Jamiołkowski isomorphism*.

Definition 6. *Let $\mathcal{E}_{A\to A'} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_{A'}))$ a superoperator map. Define the Choi matrix as*
$$J(\mathcal{E}) := (\mathcal{E}_{A\to A'} \otimes id_A)(|\Phi\rangle\langle\Phi|)\ .$$
With $|\Phi\rangle = \sum_i |i\rangle_A |i\rangle_A$, such that $\{|i\rangle_A\}$ is an orthonormal basis of $\mathcal{H}_A$.

If the superoperator $\mathcal{E}$ is completely positive, then $J(\mathcal{E})$ is positive semi-definite. Moreover, $\mathcal{E}$ is trace preserving if $\mathrm{tr}_{A'}(J(\mathcal{E})) = \mathbb{I}$ and trace non-increasing if $\mathrm{tr}_{A'}(J(\mathcal{E})) \leq \mathbb{I}$.

2.2.2 Purification

Looking back at our example on page 7, we began with a pure density operator and acquired a possibly mixed density operator by applying the partial trace. The following - referred to as *purification* - ensures that every mixed density operator can be obtained in such way.

Theorem 1. *Let $\rho_B \in \mathcal{D}(\mathcal{H}_B)$ be a density operator. Then there exists a state vector $|\phi\rangle \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ with $\mathcal{H}_A = \mathcal{H}_B$ such that $tr_A(|\phi\rangle\langle\phi|) = \rho_B$.*

Proof. Let $\rho_B \in \mathcal{L}(\mathcal{H}_B)$ with spectral decomposition $\rho_B = \sum_i^d \lambda_i |e_i\rangle\langle e_i|$ where $d = \dim(\mathcal{H}_B)$ and let $\{|i\rangle\}_{i\in I}$ be an orthonormal basis of $\mathcal{H}_A = \mathcal{H}_B$. Next, consider
$$|\phi\rangle = \sum_i^d \sqrt{\lambda_i}|i\rangle|e_i\rangle \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)\ .$$
When we trace out system A we obtain ρ_B:
$$\begin{aligned}\mathrm{tr}_A(|\phi\rangle\langle\phi|) &= \mathrm{tr}_A\left(\sum_{i,j}\sqrt{\lambda_i}\sqrt{\lambda_j}|i\rangle\langle j| \otimes |e_i\rangle\langle e_j|\right) = \sum_{i,j}\sqrt{\lambda_i\lambda_j}\langle j|i\rangle \otimes |e_i\rangle\langle e_j| \\ &= \sum_i \lambda_i |e_i\rangle\langle e_i| = \rho_B\ .\end{aligned}$$
□

Looking at Theorem 1 we say that $|\phi\rangle\langle\phi|$ is a purification of ρ_B. A question that follows from this proof is whether purification is unique. The answer is clearly no, because another orthonormal basis of $\mathcal{H}_A$ would not have changed the outcome. It turns out that purification is unique up to choice of orthonormal basis of $\mathcal{H}_A$.

2.3 Norm, Distance Measure and Metric

In this section we will introduce a *norm* and a *distance measure* for operators. Eventually, this distance measure lets us define a *metric* on the set of sub-normalized operators. This metric will tell us how similar two operators are. Later in this thesis, we will come across this norm and this metric again. We present some of its properties that become useful for proofs later on when discussing the coherent relative entropy in Chapter 4.

Definition 7. *For $p \in \mathbb{R} \setminus \{0\}$ the Schatten-p-norm for an operator $R \in \mathcal{L}(\mathcal{H})$ is defined as*

$$\|R\|_p := \mathrm{tr}(|R|^p)^{\frac{1}{p}} .$$

Additionally, the Schatten-∞-norm is defined as

$$\|R\|_\infty := \lambda_{\max}(|R|) .$$

Here $\lambda_{\max}(|R|)$ is the maximum eigenvalue of $|R| = \sqrt{R^\dagger R}$.

The Schatten-p-norm is a well-defined norm for $p \in [1, \infty)$. We will also encounter the Schatten-p-norm for $p \in (0, 1)$, yet contrary to its suggestive name it is not a norm. The Schatten-p-norm has the useful property that the norm of a tensor product is equal to the product of the norms.

Lemma 2. *Let $R_A \in \mathcal{L}(\mathcal{H}_A), R_B \in \mathcal{L}(\mathcal{H}_B)$ and $p \in \mathbb{R} \setminus \{0\}$. Then, we have the following equality:*

$$\|R_A \otimes R_B\|_p = \|R_A\|_p \otimes \|R_B\|_p .$$

Proof. We start by noting that

$$|R_A \otimes R_B| = \sqrt{(R_A \otimes R_B)^\dagger (R_A \otimes R_B)} = \sqrt{R_A^\dagger R_A \otimes R_B^\dagger R_B} .$$

From Lemma 1 it follows that

$$\sqrt{R_A^\dagger R_A \otimes R_B^\dagger R_B} = \sqrt{R_A^\dagger R_A} \otimes \sqrt{R_B^\dagger R_B} = |R_A| \otimes |R_B| .$$

Using Lemma 1 again, plus the fact that the the trace of a tensor product is the product of the traces of the components yields

$$\begin{aligned}\|R_A \otimes R_B\|_p^p &= \mathrm{tr}\left(|R_A \otimes R_B|^p\right) = \mathrm{tr}\left(\left(|R_A| \otimes |R_B|\right)^p\right) \\ &= \mathrm{tr}\left(|R_A|^p \otimes |R_B|^p\right) = \mathrm{tr}\left(|R_A|^p\right) \cdot \left(|R_B|^p\right) = \|R_A\|_p^p \cdot \|R_B\|_p^p .\end{aligned}$$

When taking the p-th square root the proof is completed. □

In the next part we discuss the *distance* between two density operators. One may think of such distance as a measure of the closeness of two operators. When two density operators are 'close' together, it is hard to tell them apart. On the other hand, if the measure yields that the density operators are 'further away' from each other, then it is easier to tell them apart. The *fidelity* is such a distance measure.

Definition 8. *Let $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ density operators. The fidelity is the function $F : \mathcal{D}(\mathcal{H}) \times \mathcal{D}(\mathcal{H}) \to [0,1]$ given by*

$$F(\rho,\sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1 = \operatorname{tr}|\sqrt{\rho}\sqrt{\sigma}| \ .$$

The fidelity increases when density operators are closer to each other, i.e. that it is a measure of 'similarity'. The fidelity equals 1 when the density operators are identical. For pure density operators $\rho = |\phi\rangle\langle\phi|$ and $\sigma = |\psi\rangle\langle\psi|$ the fidelity simplifies to $F(\rho,\sigma) = |\langle\phi|\psi\rangle|$. Next, we define the *generalized fidelity* [11] for sub-normalized states $\rho, \sigma \in \mathcal{D}_{\leq}(\mathcal{H})$:

$$F(\rho,\sigma) := \operatorname{tr}(|\sqrt{\rho}\sqrt{\sigma}|) - \sqrt{(1-\operatorname{tr}\rho)(1-\operatorname{tr}\sigma)} \in [0,1] \ .$$

Evidently, for density operators the generalized fidelity and fidelity from Definition 8 coincide. The general fidelity has some interesting properties which we briefly discuss below. Then, we present a metric and show that its properties follow from the properties we discuss now. One of these properties is the data processing inequality as proven in Ref. 10.

Lemma 3. *Let $\rho, \sigma \in \mathcal{D}_{\leq}(\mathcal{H})$ sub-normalized states and $\mathcal{T} \in \mathcal{L}(\mathcal{L}(\mathcal{H}), \mathcal{L}(\mathcal{H}'))$ a CPTNI map. Then,*

$$F((\mathcal{T}(\rho), (\mathcal{T}(\sigma)) \geq F(\rho,\sigma) \ .$$

In other words, applying a CPTNI map to sub-normalized operators makes it harder to tell these operators apart.

Let us prove another property of the generalized fidelity, namely that it is invariant under isometries.

Lemma 4. *Let $\rho, \sigma \in \mathcal{D}_{\leq}(\mathcal{H})$ sub-normalized states and $V \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$ an isometry. Then, we have*

$$F(V\rho V^\dagger, V\sigma V^\dagger) = F(\rho,\sigma) \ .$$

Proof. Let $\rho, \sigma \in \mathcal{D}_{\leq}(\mathcal{H})$ sub-normalized states and $V \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$ an isometry. Start by noting that $V\rho V^\dagger$ is again a sub-normalized state, because $V\rho V^\dagger \geq 0$ and we have $0 < \operatorname{tr}(V\rho V^\dagger) = \operatorname{tr}(\rho) \leq 1$. Moreover, each semi positive-definite operator has a unique positive semi-definite square root:

$$\left(V\sqrt{\rho}V^\dagger\right)^2 = V\sqrt{\rho}V^\dagger V\sqrt{\rho}V^\dagger = V\rho V^\dagger = \left(\sqrt{V\rho V^\dagger}\right)^2 \ .$$

Hence, $V\sqrt{\rho}V^\dagger = \sqrt{V\rho V^\dagger}$ for all $\rho \in \mathcal{D}_{\leq}(\mathcal{H})$. Next, let's have a look at the terms of the fidelity. First we notice that

$$\sqrt{(1-\operatorname{tr}(V\rho V^\dagger))(1-\operatorname{tr}(V\sigma V^\dagger))} = \sqrt{(1-\operatorname{tr}(\rho))(1-\operatorname{tr}(\sigma))} \ .$$

Secondly, again using the property of isometries that $V^\dagger V = \mathbb{I}$ we obtain the following:

$$\begin{aligned}\operatorname{tr}\left(\left|\sqrt{V\rho V^\dagger}\sqrt{V\sigma V^\dagger}\right|\right) &= \operatorname{tr}\left(\left|V\sqrt{\rho}V^\dagger V\sqrt{\sigma}V^\dagger\right|\right) = \operatorname{tr}\left(\left|V\sqrt{\rho}\sqrt{\sigma}V^\dagger\right|\right) \\ &= \operatorname{tr}\left(\sqrt{V\sqrt{\sigma}\sqrt{\rho}V^\dagger V\sqrt{\rho}\sqrt{\sigma}V^\dagger}\right) = \operatorname{tr}\left(V\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}V^\dagger\right) = \operatorname{tr}\left(|\sqrt{\sigma}\sqrt{\rho}|\right) \ .\end{aligned}$$

We conclude that $F(V\rho V^\dagger, V\sigma V^\dagger) = F(\rho,\sigma)$. ∎

The next property we address is that the generalized fidelity is *supermultiplicative* as shown in Ref. 12. Formally stated in the lemma below.

Lemma 5. *Let $\rho_A \otimes \rho_B, \sigma_A \otimes \sigma_B \in \mathcal{D}_\leq(\mathcal{H}_A \otimes \mathcal{H}_B)$ sub-normalized states. Then, we have the following inequality:*

$$F(\rho_A \otimes \rho_B, \sigma_A \otimes \sigma_B) \geq F(\rho_A, \sigma_A) \cdot F(\rho_B, \sigma_B) .$$

These are the essential properties which we use later on. We continue by using the general fidelity to define a metric for sub-normalized operators.

Definition 9. *Let $\rho, \sigma \in \mathcal{D}_\leq(\mathcal{H})$ sub-normalized states. The purified distance is the map $P : \mathcal{D}_\leq(\mathcal{H}) \times \mathcal{D}_\leq(\mathcal{H}) \to [0,1]$ given by $P(\rho,\sigma) = \sqrt{1 - F(\rho,\sigma)^2}$.*

The purified distance is a metric on $\mathcal{D}_\leq(\mathcal{H})$ [10]. From this it follows that for instance the triangle inequality holds for the purified distance. Again it gives an intuition how similar two sub-normalized operators are. We say two sub-normalized states $\rho, \sigma \in \mathcal{D}_\leq(\mathcal{H})$ are ε-close to each other if $P(\rho,\sigma) \leq \varepsilon$. This will be denoted by $\rho \approx_\varepsilon \sigma$ and can be used to relax the requirement that two operators are equal into the two operators being ε-close to each other. This is exactly what we will use in Chapter 4. From the properties shown above for the generalized fidelity it follows sometimes trivially that purified distance has similar properties. We will use these properties in Chapter 4 as well to prove properties of the in Chapter 4 presented coherent relative entropy. The data processing inequality and invariance under isometries follow directly from Lemmas 4 and 5.

Corollary 1. *Let $\rho, \sigma \in \mathcal{D}_\leq(\mathcal{H})$ sub-normalized states and $\mathcal{T} \in \mathcal{L}(\mathcal{L}(\mathcal{H}), \mathcal{L}(\mathcal{H}')$ a CPTNI map. Then, we have*

$$P((\mathcal{T}(\rho), (\mathcal{T}(\sigma)) \leq P(\rho,\sigma) .$$

Corollary 2. *Let $\rho, \sigma \in \mathcal{D}_\leq(\mathcal{H})$ sub-normalized states and $V \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$ an isometry operator. Then, the following holds:*

$$P(V\rho V^\dagger, V\sigma V^\dagger) = P(\rho,\sigma) .$$

The next property does not follow trivially, therefore we will give its proof.

Lemma 6. *Let $\rho_A, \sigma_A \in \mathcal{D}_\leq(\mathcal{H}_A)$ and $\rho_B, \sigma_B \in \mathcal{D}_\leq(\mathcal{H}_B)$ sub-normalized states. There is the following inequality:*

$$P(\rho_A \otimes \rho_B, \sigma_A \otimes \sigma_B) \leq \sqrt{P(\rho_A,\sigma_A)^2 + P(\rho_B,\sigma_B)^2} .$$

Proof. First, we apply Lemma 5 to get

$$\begin{aligned} P(\rho_A \otimes \rho_B, \sigma_A \otimes \sigma_B) &= \sqrt{1 - F(\rho_A \otimes \rho_B, \sigma_A \otimes \sigma_B)^2} \\ &\leq \sqrt{1 - F(\rho_A,\sigma_A)^2 \cdot F(\rho_B,\sigma_B)^2} . \end{aligned}$$

Note that $F(\rho_A,\sigma_A)^2, F(\rho_B,\sigma_B)^2 \in [0,1]$. Therefore we have the following inequalities:

$$\begin{aligned} 0 &\leq \left(1 - F(\rho_A,\sigma_A)^2\right)\left(1 - F(\rho_B,\sigma_B)^2\right) \\ &= 1 - F(\rho_A,\sigma_A)^2 - F(\rho_B,\sigma_B)^2 + F(\rho_A,\sigma_A)^2 F(\rho_B,\sigma_B)^2 , \end{aligned}$$

$$2 - F(\rho_A,\sigma_A)^2 - F(\rho_B,\sigma_B)^2 \geq 1 - F(\rho_A,\sigma_A)^2 F(\rho_B,\sigma_B)^2 ,$$

where the second inequality can easily be deduced from the first one. Moreover, we can write this second inequality in a suggestive way in order to deduce the following,

$$\begin{aligned} P(\rho_A \otimes \rho_B, \sigma_A \otimes \sigma_B) &\leq \sqrt{1 - F(\rho_A, \sigma_A)^2 + 1 - F(\rho_B, \sigma_B)^2} \\ &= \sqrt{P(\rho_A, \sigma_A)^2 + P(\rho_B, \sigma_B)^2}\ . \end{aligned}$$

□

Chapter 3

Information Entropy

In the previous chapter, we introduced some important concepts of quantum information theory for this thesis. We continue by introducing entropy which is an established concept in information theory. Entropy is a measure of uncertainty. In this chapter, we first present the Shannon and the Rényi entropy - two kinds of well-known entropies - both for classical and quantum information theory. Next, we define the Rényi divergence which is the main object of study in this chapter. We discuss its properties which we will use in Chapter 4 to prove properties of the coherent relative entropy which we introduce in that chapter as well. Note that from now on we write $\log(x)$ for the binary logarithm of x, i.e. $\log_2(x)$. For more details, properties and proofs of information-theoretical entropies we refer the reader to Refs. 9, 13, 14.

3.1 Classical Entropy

As mentioned an entropy measures the amount of uncertainty in a system. When looking at a random variable X, entropy is the measure of uncertainty before its value is known to us. A commonly used entropy is the *Shannon entropy* which is a function of the probability distribution of such random variable. Here, we consider only finite probability spaces.

Definition 10. *Let X be a random variable with finite range $\mathcal{X}$. The Shannon entropy of X is defined as*

$$H(X) := -\sum_{x\in\mathcal{X}} P(X=x)\cdot\log\left(P(X=x)\right) \ .$$

This entropy can be generalized to the *Rényi entropy*.

Definition 11. *Let X be a random variable with finite range $\mathcal{X}$ and $\alpha \in [0,1)\cup(1,\infty)$. The Rényi entropy of order α, is defined as*

$$H_\alpha(X) := -\frac{1}{1-\alpha}\log\left(\sum_{x\in\mathcal{X}} P(X=x)^\alpha\right) \ .$$

We retrieve the Shannon entropy by taking the limit of H_α as α approaches 1:

$$\lim_{\alpha\to 1} H_\alpha(X) := H(X) \ .$$

Also, for the limit of H_α as α approaches infinity we get

$$\lim_{\alpha\to\infty} H_\alpha(X) := -\log\left(\max_{x\in\mathcal{X}}\{P(X=x)\}\right) \ .$$

3.2 Quantum Entropy

Now, we discuss similar entropies as mentioned in the previous section, yet in a quantum information-theoretical setting. We start with a quantum version of the Shannon entropy, *the Von Neumann entropy*.

Definition 12. *Let $\rho \in \mathcal{P}(\mathcal{H})$. We define the Von Neumann entropy as follows,*

$$H(\rho) := -\mathrm{tr}(\rho \log \rho) \ .$$

In addition, we extend the classical notion of the Rényi entropy to the *quantum Rényi entropy*.

Definition 13. *Let $\alpha \in [0,1) \cup (1,\infty)$ and $\rho \in \mathcal{D}(\mathcal{H})$ a density operator. Then, the Rényi entropy of order α is defined as*

$$H_\alpha(\rho) := \frac{1}{1-\alpha}\mathrm{tr}(\rho^\alpha) = \frac{1}{1-\alpha}\|\rho\|_\alpha^\alpha \ .$$

Similarly to the classical case, we retrieve the Von Neumann entropy for taking the limit of H_α as α approaches 1. The Rényi entropy of order 0 and infinity are also defined by taking the limit of H_α:

- $\lim_{\alpha\to 0} H_\alpha(\rho) = \log \mathrm{rk}(\rho)$;
- $\lim_{\alpha\to 1} H_\alpha(\rho) = H(\rho)$;
- $\lim_{\alpha\to\infty} H_\alpha(\rho) = -\log \|\rho\|_\infty$ *(max entropy)* .

Now, we define the following distance measure for positive semi-definite operators which will be the main object of study for the rest of this chapter.

Definition 14. *Let $\alpha \in (0,1) \cup (1,\infty)$ and $\rho, \sigma \in \mathcal{P}(\mathcal{H})$. The Rényi divergence of order α is defined as*

$$D_\alpha(\rho\|\sigma) := \frac{\alpha}{\alpha-1}\log\left\|\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right\|_\alpha$$

for $\mathrm{supp}(\sigma) \subseteq \mathrm{supp}(\rho)$*, or* $\alpha < 1$ *and* $\mathrm{supp}(\rho) \not\perp \mathrm{supp}(\sigma)$*. Otherwise* $D_\alpha(\rho\|\sigma) = \infty$*.*

Taking limit of D_α as α approaches infinity we obtain the Rényi divergence of order infinity:

$$D_\infty(\rho\|\sigma) = \log\left\|\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\right\|_\infty \ .$$

By taking limit of D_α as α approaches 1 we obtain the Rényi divergence of order 1, the *quantum relative entropy*:

$$D(\rho\|\sigma) := \begin{cases} \mathrm{tr}\left(\rho\log(\rho) - \log(\sigma)\right) & \mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma) \\ \infty & \text{otherwise} \end{cases} \ .$$

Here, we remark that in the literature e.g. Ref. 10 the Rényi divergence is defined similarly for normalized ρ. For ρ not normalized, ρ is normalized by dividing it by its trace. For our purpose Definition 14 works better.

3.3 Properties of the Rényi Divergence

In this part we will give some properties of the quantum Rényi entropy. These properties will be used later on this thesis in order to prove similar properties of the coherent relative entropy. Therefore not all proofs will be spelled out. Some properties we prove, for other proofs we refer to the Refs. 10, 15, 16.

Scaling

When scaling the operators we obtain a different measure for the quantum Rényi divergence as shown in the lemma beneath.

Lemma 7. *Let $a, b \in \mathbb{R}_{>0}$ be scalars, $\rho, \sigma \in \mathcal{P}(\mathcal{H})$. Then the following holds:*

$$D_\alpha(a\rho\|b\sigma) = D_\alpha(\rho\|\sigma) + \log\left(\frac{a^{\frac{\alpha}{\alpha-1}}}{b}\right) .$$

Proof. We assume $D_\alpha(\rho\|\sigma) \neq 0$. Then, we have

$$\begin{aligned} D_\alpha(a\rho\|b\sigma) &= \frac{\alpha}{\alpha-1}\log\left\|b^{\frac{1-\alpha}{2\alpha}}\sigma^{\frac{1-\alpha}{2\alpha}}a\rho b^{\frac{1-\alpha}{2\alpha}}\sigma^{\frac{1-\alpha}{2\alpha}}\right\|_\alpha \\ &= \frac{\alpha}{\alpha-1}\log b^{\frac{1-\alpha}{\alpha}}a\left\|\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right\|_\alpha = \log\left(\frac{a^{\frac{\alpha}{\alpha-1}}}{b}\right) + D_\alpha(\rho\|\sigma) . \end{aligned}$$

□

Note that for $\alpha = \infty$ Lemma 7 translates into $D_\infty(a\rho\|b\sigma) = D_\infty(\rho\|\sigma) + \log\left(\frac{a}{b}\right)$ since $\lim_{\alpha\to\infty}\frac{\alpha}{\alpha-1} = 1$.

Data Processing Inequality

Another property is the data processing inequality. This a similar property that also holds for the purified distance as shown in Corollary 1. In words it means that acting on a system will make two operators more indistinguishable, resulting in a lower Rényi divergence as proven in Ref. 15.

Lemma 8. *Let $\mathcal{T} \in \mathcal{L}(\mathcal{L}(\mathcal{H}), \mathcal{L}(\mathcal{H}'))$ be a CPTNI map and $\rho, \sigma \in \mathcal{P}(\mathcal{H})$. Then, for all $\alpha \geq 1/2$ the following holds:*

$$D_\alpha(\mathcal{T}(\rho)\|\mathcal{T}(\sigma)) \leq D_\alpha(\rho\|\sigma) .$$

Superadditivity

When looking at two systems the Rényi divergence of a factorizable bipartite system is the sum of the Rényi divergence the separate systems. We call this property superadditivity.

Lemma 9. *Let $\rho_A, \sigma_A \in \mathcal{P}(\mathcal{H}_A)$ and $\rho_B, \sigma_B \in \mathcal{P}(\mathcal{H}_B)$. Then, we have*

$$D_\alpha(\rho_A\|\sigma_A) + D_\alpha(\rho_B\|\sigma_B) = D_\alpha(\rho_A \otimes \rho_B\|\sigma_A \otimes \sigma_B) .$$

Proof. We work out the right hand side of the equality to obtain the left hand side

$$D_\alpha(\rho_A \otimes \rho_B \| \sigma_A \otimes \sigma_B) = \frac{\alpha}{\alpha - 1} \log \left\| (\sigma_A \otimes \sigma_B)^{\frac{1-\alpha}{2\alpha}} (\rho_A \otimes \rho_B) (\sigma_A \otimes \sigma_B)^{\frac{1-\alpha}{2\alpha}} \right\|_\alpha .$$

Next, we apply Lemma 1 to obtain

$$D_\alpha(\rho_A \otimes \rho_B \| \sigma_A \otimes \sigma_B) = \frac{\alpha}{\alpha - 1} \log \left\| \left(\sigma_A^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}} \right) (\rho_A \otimes \rho_B) \left(\sigma_A^{\frac{1-\alpha}{2\alpha}} \otimes \sigma_B^{\frac{1-\alpha}{2\alpha}} \right) \right\|_\alpha .$$

Lastly, we use Lemma 2 to show

$$\begin{aligned} D_\alpha(\rho_A \otimes \rho_B \| \sigma_A \otimes \sigma_B) &= \frac{\alpha}{\alpha - 1} \log \left[\left\| \sigma_A^{\frac{1-\alpha}{2\alpha}} \rho_A \sigma_A^{\frac{1-\alpha}{2\alpha}} \right\|_\alpha \cdot \left\| \sigma_B^{\frac{1-\alpha}{2\alpha}} \rho_B \sigma_B^{\frac{1-\alpha}{2\alpha}} \right\|_\alpha \right] \\ &= \frac{\alpha}{\alpha - 1} \log \left\| \sigma_A^{\frac{1-\alpha}{2\alpha}} \rho_A \sigma_A^{\frac{1-\alpha}{2\alpha}} \right\|_\alpha + \frac{\alpha}{\alpha - 1} \log \left\| \sigma_B^{\frac{1-\alpha}{2\alpha}} \rho_B \sigma_B^{\frac{1-\alpha}{2\alpha}} \right\|_\alpha = D_\alpha(\rho_A \| \sigma_A) + D_\alpha(\rho_B \| \sigma_B) . \end{aligned}$$

□

Isometry Invariance

Just like the generalized fidelity - Lemma 4 - and the purified distance - Corollary 2 - the Rényi divergence is invariant isometry transformations, which is shown in Ref. 10.

Lemma 10. *Let $V \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$ be an isometry and $\rho, \sigma \in \mathcal{P}(\mathcal{H})$. Then, for all $\alpha \geq 0$ we have*

$$D_\alpha \left(V \rho V^\dagger \| V \sigma V^\dagger \right) = D_\alpha(\rho \| \sigma) .$$

Triangle-like Inequality

In general, the Rényi divergence does not satisfy the triangle inequality. We consider the case that it does hold a property that looks like the triangle inequality as explained in Ref. 16.

Lemma 11. *Let $\rho, \sigma, \chi \in \mathcal{P}(\mathcal{H})$. Then, for all $\alpha \in [1/2, \infty)$ the following holds:*

$$D_\alpha(\rho \| \sigma) \leq D_\alpha(\rho \| \chi) + D_\infty(\chi \| \sigma) .$$

Thus, for $\alpha = \infty$ this translates into the triangle inequality as one is familiar with.

Chapter 4

Coherent Relative Entropy

In this chapter, we discuss and contribute to the so-called coherent relative entropy. This new entropy notion was recently introduced in Refs. 6, 7. In these papers they discuss the physical interpretation of this measure and argue why it is called an entropy. They show various properties of the coherent relative entropy, e.g. the data processing inequality. In their proofs of these properties they often use non-trivial techniques, for example from semi-definite programming. This results in lengthy and technical proofs.

Here, we provide new quantum information theoretic insight into this novel entropy measure. The starting point of our contribution is a relation between the coherent relative entropy and the Rényi divergence of order infinity; this connection was already mentioned in Refs. 6, 7, but little attention was given to it. In this thesis we exploit this connection in the following two ways. On the one hand, by extending this connection to the Rényi divergence of general order α, we obtain a natural generalization of the coherent relative entropy to a general order α, where the original definition corresponds to $\alpha = \infty$. On the other hand, it allows us to re-prove some of the properties of the coherent relative entropy, as considered and proven in Refs. 6, 7, but now: (1) by means of simpler proofs that exploit corresponding properties of the Rényi divergence, (2) for our generalized version of the coherent relative entropy.

4.1 Process Matrix

In this section the *process matrix* is presented which is component of the coherent relative entropy. Before we give its definition it is useful that we first discuss the framework we use. For the coherent relative entropy we have to consider two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$. Recall that we refer to these spaces as system A and system B respectively. Also, we consider the *reference system* R_A which is isomorphic to system A. Take a density operator with spectral decomposition $\sigma_A = \sum_i \lambda_i |i\rangle\langle i|_A$ as well as a CPTNI map $\mathcal{E}_{A\to B}$ which preserves the trace of σ_A. We call σ_A the *input state*. The process matrix holds information about the input state and the CPTNI map. Next, let us examine a purification of $|\sigma\rangle_A$, namely $|\sigma\rangle\langle\sigma|_{AR_A}$ with $|\sigma\rangle_{AR_A} = \sum_i \sqrt{\lambda_i}|i\rangle|i\rangle$. Now, all is set to define the process matrix.

Definition 15. *Let $\sigma_A \in \mathcal{D}(\mathcal{H}_A)$ a density operator and let $\mathcal{E}_{A\to B} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B))$ be a CPTNI map such that $\mathrm{tr}(\mathcal{E}_{A\to B}(\sigma_A)) = \mathrm{tr}(\sigma_A)$. The process matrix of σ_A and $\mathcal{E}_{A\to B}$ is* *defined as*

$$\rho_{BR_A} := (\mathcal{E}_{A\to B} \otimes id_{R_A})(|\sigma\rangle\langle\sigma|_{AR_A}) = \sum_{i,j} \sqrt{\lambda_i \lambda_j}\, \mathcal{E}_{A\to B}(|i\rangle\langle j|) \otimes |i\rangle\langle j| \ .$$

We note that the process matrix differs from the Choi-Jamiołkowski isomorphism presented in Definition 6, but looks quite similar. The process matrix is an operator that encodes information of the CPTNI map $\mathcal{E}_{A\to B}$ and the input state σ_A. It also gives an intuitive notion that the system R_A remembers what the input state was. Observe that ρ_{BR_A} uniquely determines σ_A as well as the CPTNI $\mathcal{E}_{A\to B}$ on the support of σ_A.

Lemma 12. *Let ρ_{BR_A} be a process matrix given by input state σ_A and CPTNI map $\mathcal{E}_{A\to B}$. Then, ρ_{BR_A} uniquely determines this input state σ_A and CPTNI map $\mathcal{E}_{A\to B}$ on the support of σ_A.*

Proof. We retrieve σ_A by applying the partial trace to the process matrix:

$$\sigma_A = \sum_{i,j} \sqrt{\lambda_i \lambda_j}\, \text{tr}\left[\mathcal{E}\left(|i\rangle\langle j|\right)\right] \cdot |i\rangle\langle j| = \sum_{i,j} \sqrt{\lambda_i \lambda_j} \langle j|i\rangle \cdot |i\rangle\langle j| = \sum_i \lambda_i |i\rangle\langle i| \ .$$

This uniquely determines the input state σ_A. We retrieve the CPTNI map with the following observation,

$$(\mathbb{I} \otimes \langle i|)\rho_{BR_A}(\mathbb{I} \otimes |j\rangle) = \sqrt{\lambda_i \lambda_j}\, \mathcal{E}(|i\rangle\langle j|) \ .$$

This determines $\mathcal{E}(|i\rangle\langle j|)$ for $|i\rangle, |j\rangle$ in the support of σ_A. □

We continue by examining how the process matrix behaves under CPTNI maps. Let us start with a process matrix ρ_{BR_A} which encodes an input state σ_A and a CPTNI map $\mathcal{E}_{A\to B}$. Next, let $\mathcal{F}_{B\to C}$ be a CPTNI map such that $\text{tr}(\mathcal{F}_{B\to C}(\mathcal{E}_{A\to B}(\sigma_A))) = 1$. Applying this map to the process matrix yields the following process matrix:

$$\rho_{CR_A} := \mathcal{F}_{B\to C} \otimes \text{id}_{R_A}(\rho_{BR_A}) \ .$$

Which encodes input state σ_A and CPTNI map $\mathcal{F}_{B\to C} \circ \mathcal{E}_{A\to B}$. An example of CPTNI maps are M_{V_B} and M_{V_A} given by isometries $V_A \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_{A'})$ and $V_B \in \mathcal{L}(\mathcal{H}_B, \mathcal{H}_{B'})$. Given these isometries such that σ_A and $\mathcal{E}_{A\to B}(\sigma_A)$ are in the support of V_A and V_B respectively, we apply these maps on the process matrix we achieve the following:

$$\begin{aligned}\rho'_{BR_A} = M_{V_B} \otimes M_{V_A}(\rho_{BR_A}) &= \sum_{i,j} \sqrt{\lambda_i \lambda_j}\, M_{V_B}\mathcal{E}(|i\rangle\langle j|) \otimes M_{V_A}|i\rangle\langle j| \\ &= \sum_{i,j} \sqrt{\lambda_i \lambda_j}\, M_{V_B}\mathcal{E}M_{V_{A^\dagger}}(M_{V_A}|i\rangle\langle j|) \otimes M_{V_A}|i\rangle\langle j| \ .\end{aligned}$$

Recall that $M_V(\cdot) = V \cdot V^\dagger$. Besides, we used the property of isometries that $V_A^\dagger V_A = \mathbb{I}$. Notice that ρ'_{BR_A} encodes CPTNI map $M_{V_B}\mathcal{E}M_{V_{A^\dagger}}$ which is trace preserving on the encoded input state $\sigma'_A = M_{V_A}\sigma_A$.

4.2 The Coherent Relative Entropy and its Relation to the Rényi Divergence

In this section, we will introduce the coherent relative entropy. We will adopt the definition as used in Ref. 6. Additionally, we present its connection to the Rényi divergence of order infinity which forms the starting point of our mathematical contributions. Here, we use this relation to define a new generalized coherent relative entropy.

Definition 16. *Consider two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ with $\Gamma_A \in \mathcal{P}(\mathcal{H}_A)$ and $\Gamma_B \in \mathcal{P}(\mathcal{H}_B)$. Then, for any process matrix ρ_{BR_A} representing an input state $\sigma_A \in \mathcal{D}(\mathcal{H}_A)$ and a CPTNI map $\mathcal{E}_{A\to B} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B))$ the coherent relative entropy is defined as*

$$\hat{D}_{A\to B}(\rho_{BR_A}\|\Gamma_A, \Gamma_B) := \max_{\mathcal{T}} \max \left\{ \lambda \mid \mathcal{T}(\Gamma_A) \leq 2^{-\lambda}\Gamma_B \right\} ,$$

where the optimization is over all CPTNI maps $\mathcal{T}_{A\to B} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B))$ such that $\mathcal{T}(\sigma_{AR_B}) = \rho_{BR_A}$.

Recall that ρ_{BR_A} determines σ_A and a CPTNI map $\mathcal{E}$ on the support of σ_A. So within the optimization over the CPTNI maps $\mathcal{T}$, there is a degree of freedom outside the support of σ_A. When σ_A has full support, i.e. its kernel is trivial, this degree of freedom is lost and it is necessary that $\mathcal{T} = \mathcal{E}$.

We may want to relax this definition by not requiring $\mathcal{T}(\sigma_{AR_B})$ and ρ_{BR_A} to be exactly equal. It can be more useful to consider cases when $\mathcal{T}(\sigma_{AR_B})$ is close enough to ρ_{BR_A}. We can quantify this by saying $\mathcal{T}(\sigma_{AR_B})$ is ε-close to ρ_{BR_A} in terms of the purified distance as introduced in Definition 9. For this we need the *smooth coherent relative entropy*.

Definition 17. *Consider two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ with respective operators $\Gamma_A \geq 0$ and $\Gamma_B \geq 0$. Then, for any process matrix ρ_{BR_A} representing an input state σ_A and a CPTNI map $\mathcal{E}_{A\to B}$ and for all $\varepsilon \geq 0$, the smooth coherent relative entropy is defined as*

$$\hat{D}^{\varepsilon}_{A\to B}(\rho_{BR_A}\|\Gamma_A, \Gamma_B) := \max_{\mathcal{T}} \max \left\{ \lambda \mid \mathcal{T}(\Gamma_A) \leq 2^{-\lambda}\Gamma_B \right\} ,$$

where the optimization is over all CPTNI maps $\mathcal{T}_{A\to B} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B))$ such that $\mathcal{T}(\sigma_{AR_A}) \approx_{\varepsilon} \rho_{BR_A}$.

We note that that for $\varepsilon = 0$ the smooth coherent relative entropy coincides with the coherent relative entropy as stated in Definition 16. When looking more closely at the constraint $\mathcal{T}(\Gamma_A) \leq 2^{-\lambda}\Gamma_B$, it is possible to derive an elegant relation between the coherent relative entropy and the Rényi divergence of order infinity. This relation also is given in proposition 12 of Ref. 6 and will be the starting point of our approach when introducing a generalized coherent relative entropy and proving properties of the coherent relative entropy.

Theorem 2. *Consider Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ with respective operators $\Gamma_A \geq 0$ and $\Gamma_B \geq 0$. Then, for any process matrix ρ_{BR_A} representing an input state σ_A and a CPTNI map $\mathcal{E}_{A\to B}$ and for all $\varepsilon \geq 0$ there is the following relation:*

$$\hat{D}^{\varepsilon}_{A\to B}(\rho_{BR_A}\|\Gamma_A, \Gamma_B) = \max_{\mathcal{T}_{A\to B}} \{ -D_{\infty}(\mathcal{T}(\Gamma_A)\|\Gamma_B) \mid \mathcal{T}(\sigma_{AR_A}) \approx_{\varepsilon} \rho_{BR_A} \} ,$$

where the optimization is over all CPTNI maps $\mathcal{T}_{A\to B} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_A), \mathcal{L}(\mathcal{H}_B))$.

Proof. Let us have a look at one of the requirements in the optimization process of the coherent relative entropy, $\mathcal{T}(\Gamma_A) \leq 2^{-\lambda}\Gamma_B$. We note that this expression is equivalent to

$$\Gamma_B^{-1/2}\mathcal{T}(\Gamma_A)\Gamma_B^{-1/2} \leq 2^{-\lambda}\Gamma_B^0 .$$

Here, Γ_B^0 is the identity on the support of Γ_B. We note that the support of $\Gamma_B^{-1/2}\mathcal{T}(\Gamma_A)\Gamma_B^{-1/2}$ is contained in the support of $2^{-\lambda}\Gamma_B^0$. This means that the eigenbasis of $\Gamma_B^{-1/2}\mathcal{T}(\Gamma_A)\Gamma_B^{-1/2}$

is an eigenbasis of $2^{-\lambda}\Gamma_B^0$ too. The operator $2^{-\lambda}\Gamma_B^0$ has eigenvalues $2^{-\lambda}$ and 0. Therefore this inequality is equivalent to the eigenvalues of $\Gamma_B^{-1/2}\mathcal{T}(\Gamma_A)\Gamma_B^{-1/2}$ being at most $2^{-\lambda}$. We recall that $\|\Gamma_B^{-1/2}\mathcal{T}(\Gamma_A)\Gamma_B^{-1/2}\|_\infty$ is the Schatten-∞-norm which is the largest eigenvalue of the operator $\Gamma_B^{-1/2}\mathcal{T}(\Gamma_A)\Gamma_B^{-1/2}$. This means we can rewrite our inequality into

$$\|\Gamma_B^{-1/2}\mathcal{T}(\Gamma_A)\Gamma_B^{-1/2}\|_\infty \leq 2^{-\lambda} \ .$$

By applying the logarithm of base two - which is monotonic - on both sides the Rényi divergence of order infinity emerges:

$$D_\infty(\mathcal{T}(\Gamma_A)\|\Gamma_B) = \log \left\|\Gamma_B^{-1/2}\mathcal{T}(\Gamma_A)\Gamma_B^{-1/2}\right\|_\infty \leq -\lambda \ .$$

For the coherent relative entropy we optimize in such way to obtain the maximum of λ, hence we want to make $|\lambda|$ as large as possible. We notice that $|\lambda|$ takes its maximum value when $-\lambda$ is equal to the minimal value of $D_\infty(\mathcal{T}\,(\Gamma_A)\|\Gamma_B)$. Note that $D_\infty(\mathcal{T}\,(\Gamma_A)\|\Gamma_B)$ is a function of $\mathcal{T}$, hence $-\lambda = \min_{\mathcal{T}}\{D_\infty(\mathcal{T}\,(\Gamma_A)\|\Gamma_B)\}$.

We recall that $\mathcal{T}(\sigma_{AR_A})$ still needs to be ε-close to ρ_{BR_A}. When we add this constraint, we get the following relation:

$$-\hat{D}^\varepsilon_{A\to B}(\rho_{BR_A}\|\Gamma_A,\Gamma_B) = \min_{\mathcal{T}_{A\to B}}\left\{D_\infty(\mathcal{T}\,(\Gamma_A)\|\Gamma_B) \mid \mathcal{T}(\sigma_{AR_A}) \approx_\varepsilon \rho_{BR_A}\right\} \ .$$

Lastly, we use the property that $\max_x\{-x\} = -\min_x\{x\}$, to obtain

$$\begin{aligned}\hat{D}^\varepsilon_{A\to B}(\rho_{BR_A}\|\Gamma_A,\Gamma_B) &= -\min_{\mathcal{T}_{A\to B}}\left\{D_\infty(\mathcal{T}\,(\Gamma_A)\|\Gamma_B) \mid \mathcal{T}(\sigma_{AR_A}) \approx_\varepsilon \rho_{BR_A}\right\} \\ &= \max_{\mathcal{T}_{A\to B}}\left\{-D_\infty(\mathcal{T}\,(\Gamma_A)\|\Gamma_B) \mid \mathcal{T}(\sigma_{AR_A}) \approx_\varepsilon \rho_{BR_A}\right\} \ .\end{aligned}$$

□

So there is a relation between coherent relative entropy and Rényi divergence of order infinity. Instinctively the idea of a generalization of this relation follows. As we defined the Rényi divergence of order α we can define coherent relative entropy of order α as well.

Definition 18. *Consider two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ with respective operators $\Gamma_A,\Gamma_B \geq 0$. Let $\alpha \in [0,1)\cup(1,\infty)$. Then, for any process matrix ρ_{BR_A} representing an input state σ_A and a CPTNI map $\mathcal{E}_{A\to B}$ and for all $\varepsilon \geq 0$, we define the smooth coherent relative entropy of order α as*

$$\hat{D}^\varepsilon_{\alpha,A\to B}(\rho_{BR_A}\|\Gamma_A,\Gamma_B) := \max_{\mathcal{T}_{A\to B}}\left\{-D_\alpha(\mathcal{T}\,(\Gamma_A)\|\Gamma_B) \mid \mathcal{T}(\sigma_{AR_A}) \approx_\varepsilon \rho_{BR_A}\right\} \ ,$$

where the optimization is over all CPTNI maps $\mathcal{T}_{A\to B} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_A),\mathcal{L}(\mathcal{H}_B))$.

When we write the coherent relative entropy without a subscript α we refer to the coherent relative entropy (of order infinity) as presented in Definition 17.

4.3 Properties of the Coherent Relative Entropy

The coherent relative entropy holds some properties that we want in order for it to be called an entropy. In Refs. 6, 7 these properties are proven from scratch to a large extent using non-trivial techniques, e.g., from semi-definite programming. As a consequence, these proofs in Refs. 6, 7 are technical and lengthy. Here, instead, we exploit the observed connection between the coherent relative entropy to the Rényi divergence as discussed in Theorem 2 in order to show that many of these properties of the coherent relative entropy are inherited from similar properties of the Rényi divergence as discussed in subsection 3.3. This results in short proofs, and sheds new insight into these properties. Besides, we generalize these properties for the coherent relative entropy of order α.

Scaling

The coherent relative entropy behaves in a similar way as the Rényi divergence when we scale the positive semi-definite operators. This property is also discussed in proposition 8 of Ref. 6. Here, the proof follows almost directly from Lemma 7.

Lemma 13. *Let $a, b \in \mathbb{R}_{>0}$ be scalars, $\Gamma_A, \Gamma_B \geq 0$ positive semi-definite operators, ρ_{BR_A} a process matrix. Then, for all $\alpha, \varepsilon \geq 0$ we have*

$$\hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|a\Gamma_A, b\Gamma_B) = \hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|\Gamma_A, \Gamma_B) + \log\left(\frac{b}{a^{\frac{\alpha}{\alpha-1}}}\right) .$$

Proof. From linearity of CPTNI maps it follows that $\mathcal{T}(a\Gamma_A) = a\mathcal{T}(\Gamma_A)$, hence

$$D_\alpha(\mathcal{T}(a\Gamma_A)\|b\Gamma_B) = D_\alpha(a\mathcal{T}(\Gamma_A)\|b\Gamma_B) .$$

Using Lemma 7 we get

$$-D_\alpha(a\mathcal{T}(\Gamma_A)\|b\Gamma_B) = -D_\alpha(\mathcal{T}(\Gamma_A)\|\Gamma_B) + \log\left(\frac{b}{a^{\frac{\alpha}{\alpha-1}}}\right) .$$

Optimizing both sides we obtain the coherent relative entropy:

$$\hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|a\Gamma_A, b\Gamma_B) = \hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|\Gamma_A, \Gamma_B) + \log\left(\frac{b}{a^{\frac{\alpha}{\alpha-1}}}\right) .$$

□

Data Processing Inequality

In proposition 19 of Ref. 6 the data processing inequality for coherent relative entropy is shown by using optimization techniques. In this thesis, the proof of the data processing inequality is a direct consequence of the data processing inequality of the Rényi divergence as discussed in Lemma 8 and the purified distance as discussed Corollary 1.

Lemma 14. *Let $\mathcal{F}_{B\to C} \in \mathcal{L}(\mathcal{L}(\mathcal{H}_B), \mathcal{L}(\mathcal{H}_C))$ be a CPTP map, $\Gamma_A, \Gamma_B \geq 0$ and ρ_{BR_A} a process matrix. Then for all $\alpha \geq 1/2$ and $\varepsilon \geq 0$ we have*

$$\hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|\Gamma_A, \Gamma_B) \leq \hat{D}^{\varepsilon}_{\alpha,A\to C}(\mathcal{F}(\rho_{BR_A})\|\Gamma_A, \mathcal{F}(\Gamma_B)) .$$

Proof. Let $\overline{\mathcal{T}}_{A\to B}$ be the CPTNI map such that

$$\hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|\Gamma_A, \Gamma_B) = -D_\alpha(\overline{\mathcal{T}}(\Gamma_A)\|\Gamma_B) .$$

From Lemma 8 it follows that

$$-D_\alpha(\overline{\mathcal{T}}(\Gamma_A)\|\Gamma_B) \leq -D_\alpha(\mathcal{F}(\overline{\mathcal{T}}(\Gamma_A))\|\mathcal{F}(\Gamma_B)) .$$

Moreover, from Corollary 1 it follows that

$$P\left(\mathcal{F}(\overline{\mathcal{T}}(\sigma_{AR_A})), \mathcal{F}(\rho_{BR_A})\right) \leq P\left(\overline{\mathcal{T}}(\sigma_{AR_A}), \rho_{BR_A}\right) \leq \varepsilon\,.$$

Therefore, noting that $\mathcal{F}(\rho_{BR_A})$ is the process matrix of initial state σ_A and CPTP map $\mathcal{F} \circ \mathcal{E}$, we have

$$\begin{aligned}
&\hat{D}^{\varepsilon}_{\alpha,A\to C}(\mathcal{F}(\rho_{BR_A})\|\Gamma_A, \mathcal{F}(\Gamma_B)) \\
&\quad = \max_{\mathcal{T}_{A\to C}} \{-D_\alpha(\mathcal{T}(\Gamma_A)\|\mathcal{F}(\Gamma_B))\,|\, \mathcal{T}\circ\mathcal{F}(\rho_{BR_A}) \approx_\varepsilon \mathcal{F}(\sigma_{AR_A})\} \\
&\quad \geq -D_\alpha\left(\mathcal{F}\circ\overline{\mathcal{T}}(\Gamma_A)\|\mathcal{F}(\Gamma_B)\right) \geq -D_\alpha\left(\overline{\mathcal{T}}(\Gamma_A)\|\Gamma_B\right) = \hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|\Gamma_A, \Gamma_B)\,.
\end{aligned}$$

□

Isometry Invariance

In Corollary 2 and in and Lemma 10 is shown that respectively the purified distance and the Rényi divergence are invariant under isometry. The coherent relative entropy holds this property as well. In proposition 7 of Ref. 6, this property is also mentioned. However, its proof is not totally spelled out. In this thesis, we will provide an insightful proof.

Lemma 15. *Let $V_A \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_{A'})$, $V_B \in \mathcal{L}(\mathcal{H}_B, \mathcal{H}_{B'})$ isometries, Γ_A, Γ_B positive semidefinite operators, ρ_{BR_A} a process matrix, such that Γ_A and ρ_{R_A} are in the support of V_A and that Γ_B and σ_B are in the support of V_B. Then, for all $\alpha, \varepsilon \geq 0$ we have*

$$\hat{D}^{\varepsilon}_{\alpha,A\to B}(M_{V_B}\otimes M_{V_A}(\rho_{BR_A})\|M_{V_A}(\Gamma_A), M_{V_B}(\Gamma_B)) = \hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|\Gamma_A, \Gamma_B)\,.$$

Proof. As discussed on page 18 we note that if process matrix ρ_{BR_A} encodes input state σ_A and CPTNI map $\mathcal{E}_{A\to B}$, then $\rho'_{BR_A} = M_{V_B}\otimes M_{V_A}(\rho_{BR_A})$ encodes input state $\sigma'_A = M_{V_A}(\sigma_A)$ and the CPTNI map $M_{V_B}\circ\mathcal{E}_{A\to B}\circ M_{V_{A^\dagger}}$. Also, notice that for the purification of σ'_A we similarly take $\sigma'_{AR_A} = M_{V_A}\otimes M_{V_A}(\sigma_{AR_A})$. Let $\mathcal{T}_{A\to B}$ be a CPTNI such that

$$\rho_{BR_A} \approx_\varepsilon \mathcal{T}(\sigma_{AR_A})\,.$$

From Corollary 2 it follows that this is equivalent to

$$\rho'_{BR_A} \approx_\varepsilon M_{V_B}\circ\mathcal{T}\otimes M_{V_A}(\sigma_{AR_A}) = M_{V_B}\circ\mathcal{T}\circ M_{V_{A^\dagger}}(\sigma'_{AR_A}) = \mathcal{T}'(\sigma'_{AR_A})\,.$$

In the last equality we used $V_A^\dagger V_A = \mathbb{I}$, a property of isometries. Thus ρ_{BR_A} is ε-close to $\mathcal{T}(\sigma_{AR_A})$ if and only if ρ'_{BR_A} is ε-close to $\mathcal{T}'(\sigma'_{AR_A})$. Besides, we observe that $\mathcal{T}' = M_{V_B}\circ\mathcal{T}\circ M_{V_{A^\dagger}}$ is again a CPTNI map. In general, we have the following relation:

$$\hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|\Gamma_A, \Gamma_B) \geq -D_\alpha(\mathcal{T}(\Gamma_A)\|\Gamma_B) = -D_\alpha(M_{V_B}\circ\mathcal{T}(\Gamma_A)\|M_{V_B}(\Gamma_B))\,,$$

where we applied Lemma 10 for the last equality. Again, using the property of isometries and the equivalence of the ε-closeness discussed above we get the following relation,

$$\begin{aligned}
&\hat{D}^{\varepsilon}_{\alpha,A\to B}(M_{V_B}\otimes M_{V_A}(\rho_{BR_A})\|M_{V_A}(\Gamma_A), M_{V_B}(\Gamma_B)) \\
&\quad \geq -D_\alpha(M_{V_B}\circ\mathcal{T}\circ M_{V_{A^\dagger}}(M_{V_A}(\Gamma_A))\|M_{V_B}(\Gamma_B)) \\
&\quad = -D_\alpha(\mathcal{T}(\Gamma_A)\|\Gamma_B) \leq \hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|\Gamma_A, \Gamma_B)\,.
\end{aligned}$$

For $\mathcal{T}$ such that last inequality becomes an equality we get

$$\hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|\Gamma_A,\Gamma_B) \geq \hat{D}^{\varepsilon}_{\alpha,A\to B}(M_{V_B}\otimes M_{V_A}(\rho_{BR_A})\|M_{V_A}(\Gamma_A),M_{V_B}(\Gamma_B)) .$$

Furthermore, let $\overline{\mathcal{T}}'$ be the CPTNI map such that

$$\begin{aligned}\hat{D}^{\varepsilon}_{\alpha,A\to B}(M_{V_B}\otimes M_{V_A}(\rho_{BR_A})\|M_{V_A}(\Gamma_A),M_{V_B}(\Gamma_B)) &= -D_\alpha(\overline{\mathcal{T}}'(M_{V_A}(\Gamma_A))\|M_{V_B}(\Gamma_B))\\ &\geq -D_\alpha(M_{V_B}\circ\mathcal{T}\circ M_{V_{A^\dagger}}(M_{V_A}(\Gamma_A))\|M_{V_B}(\Gamma_B)) ,\end{aligned}$$

where the last inequality follows from the fact that $\overline{\mathcal{T}}'$ maximizes the Rényi divergence and $M_{V_B}\circ\mathcal{T}\circ M_{V_{A^\dagger}}$ may not be the optimal solution. This means the following:

$$\hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|\Gamma_A,\Gamma_B) \leq \hat{D}^{\varepsilon}_{\alpha,A\to B}(M_{V_B}\otimes M_{V_A}(\rho_{BR_A})\|M_{V_A}(\Gamma_A),M_{V_B}(\Gamma_B)) .$$

We finally conclude that

$$\hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|\Gamma_A,\Gamma_B) = \hat{D}^{\varepsilon}_{\alpha,A\to B}(M_{V_B}\otimes M_{V_A}(\rho_{BR_A})\|M_{V_A}(\Gamma_A),M_{V_B}(\Gamma_B)) .$$

□

Chain Rule

The coherent relative entropy also holds some kind of chain rule. The coherent relative entropy of a concatenation of processes can never be lower than the sum of the individual processes. This property corresponds to proposition 20 in Ref. 6.

Lemma 16. *Let $\mathcal{H}_A,\mathcal{H}_B,\mathcal{H}_C$ be Hilbert spaces with respectively positive semi-definite operators Γ_A,Γ_B and Γ_C. In addition, let R_A and R_B be reference systems. Let σ_A be a density operator and $\mathcal{E}_{A\to B},\mathcal{F}_{B\to C}$ CPTNI maps, such that $\mathrm{tr}[\mathcal{F}\circ\mathcal{E}(\sigma_A)] = \mathrm{tr}[\mathcal{E}(\sigma_A)] = 1$. Then for all $\varepsilon,\varepsilon'\geq 0$ and $\alpha \geq 1/2$ we have*

$$\begin{aligned}\hat{D}^{\varepsilon}_{\alpha,A\to B}(\mathcal{E}(\sigma_{AR_A})\|\Gamma_A,\Gamma_B) + &\hat{D}^{\varepsilon'}_{B\to C}(\mathcal{F}(\rho_{BR_A})\|\Gamma_B,\Gamma_C)\\ &\leq \hat{D}^{\varepsilon+\varepsilon'}_{\alpha,A\to C}(\mathcal{F}\circ\mathcal{E}(\sigma_{AR_A})\|\Gamma_A,\Gamma_C) .\end{aligned}$$

Proof. Let $\overline{\mathcal{T}}_{A\to B}$ and $\overline{\mathcal{T}}'_{B\to C}$ be the CPTNI maps such that

$$\begin{aligned}\hat{D}^{\varepsilon}_{\alpha,A\to B}(\mathcal{E}(\sigma_{AR_A})\|\Gamma_A,\Gamma_B) &= -D_\alpha(\overline{\mathcal{T}}(\Gamma_A)\|\Gamma_B) \qquad \text{and}\\ \hat{D}^{\varepsilon'}_{B\to C}(\mathcal{F}(\rho_{BR_A})\|\Gamma_B,\Gamma_C) &= -D_\infty\left(\overline{\mathcal{T}}'(\Gamma_B)\|\Gamma_C\right) .\end{aligned}$$

We use data processing inequality of the Rényi divergence as discussed Lemma 8, to obtain the following inequality:

$$\begin{aligned}&-D_\alpha\left(\overline{\mathcal{T}}(\Gamma_A)\|\Gamma_B\right) - D_\infty\left(\overline{\mathcal{T}}'(\Gamma_B)\|\Gamma_C\right)\\ &\qquad\leq -D_\alpha\left(\overline{\mathcal{T}}'\circ\overline{\mathcal{T}}(\Gamma_A)\|\overline{\mathcal{T}}'(\Gamma_B)\right) - D_\infty\left(\overline{\mathcal{T}}'(\Gamma_B)\|\Gamma_C\right) .\end{aligned}$$

Next, we apply Lemma 11, which results in the following inequality:

$$-D_\alpha\left(\overline{\mathcal{T}}'\circ\overline{\mathcal{T}}(\Gamma_A)\|\overline{\mathcal{T}}'(\Gamma_B)\right) - D_\infty\left(\overline{\mathcal{T}}'(\Gamma_B)\|\Gamma_C\right) \leq -D_\alpha\left(\overline{\mathcal{T}}'\circ\overline{\mathcal{T}}(\Gamma_A)\|\Gamma_C\right) .$$

Next, we show that $\overline{\mathcal{T}}' \circ \overline{\mathcal{T}}$ is indeed an option in the optimization process for $\hat{D}^{\varepsilon+\varepsilon'}_{\alpha,A\to C}(\mathcal{F} \circ \mathcal{E}(\sigma_{AR_A})\|\Gamma_A, \Gamma_C)$. Remember that the purified distance is a metric, hence it obeys the triangle inequality:

$$P\left(\overline{\mathcal{T}}' \circ \overline{\mathcal{T}}(\sigma_{AR_A}), \mathcal{F} \circ \mathcal{E}(\sigma_{AR_A})\right) \leq P\left(\overline{\mathcal{T}}' \circ \overline{\mathcal{T}}(\sigma_{AR_A}), \overline{\mathcal{T}}' \circ \mathcal{E}(\sigma_{AR_A})\right) + P\left(\overline{\mathcal{T}}' \circ \mathcal{E}(\sigma_{AR_A}), \mathcal{F} \circ \mathcal{E}(\sigma_{AR_A})\right) .$$

When we apply Corollary 1, we obtain

$$P\left(\overline{\mathcal{T}}' \circ \overline{\mathcal{T}}(\sigma_{AR_A}), \overline{\mathcal{T}}' \circ \mathcal{E}(\sigma_{AR_A})\right) + P\left(\overline{\mathcal{T}}' \circ \mathcal{E}(\sigma_{AR_A}), \mathcal{F} \circ \mathcal{E}(\sigma_{AR_A})\right) \leq P\left(\overline{\mathcal{T}}(\sigma_{AR_A}), \mathcal{E}(\sigma_{AR_A})\right) + P\left(\overline{\mathcal{T}}'(\rho_{BR_B}), \mathcal{F}(\rho_{BR_B})\right) \leq \varepsilon + \varepsilon' .$$

Thus, $\overline{\mathcal{T}}' \circ \overline{\mathcal{T}}$ is indeed an option in the optimization process. We conclude that

$$\hat{D}^{\varepsilon}_{\alpha,A\to B}(\mathcal{E}(\sigma_{AR_A})\|\Gamma_A, \Gamma_B) + \hat{D}^{\varepsilon'}_{B\to C}(\mathcal{F}(\rho_{BR_B})\|\Gamma_B, \Gamma_C) \leq \hat{D}^{\varepsilon+\varepsilon'}_{\alpha,A\to C}(\mathcal{F} \circ \mathcal{E}(\sigma_{AR_A})\|\Gamma_A, \Gamma_C) .$$

□

Superadditivity

The notion of the superadditivity property of the Rényi divergence - Lemma 9 - gives rise to a similar property for the coherent relative entropy as well. This property is also discussed in proposition 9 of Ref. 6.

Lemma 17. *Let $\mathcal{H}_A, \mathcal{H}_{A'}, \mathcal{H}_B, \mathcal{H}_{B'}$ be Hilbert spaces with respectively positive semi-definite operators $\Gamma_A, \Gamma_{A'}, \Gamma_B$ and $\Gamma_{B'}$. In addition, let ρ_{BR_A} and $\zeta_{B'R_{A'}}$ process matrices, $\alpha \geq 1/2$ and $\varepsilon, \varepsilon' \geq 0$. Then,*

$$\hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|\Gamma_A, \Gamma_B) + \hat{D}^{\varepsilon'}_{\alpha,A'\to B'}(\zeta_{B'R_{A'}}\|\Gamma_{A'}, \Gamma_{B'}) \leq \hat{D}^{\varepsilon''}_{\alpha,A\otimes A'\to B\otimes B'}(\rho_{BR_A} \otimes \zeta_{B'R_{A'}}\|\Gamma_A \otimes \Gamma_{A'}, \Gamma_B \otimes \Gamma_{B'}) ,$$

where $\varepsilon'' = \sqrt{\varepsilon^2 + \varepsilon'^2}$.

Proof. Let $\overline{\mathcal{T}}_{A\to B}, \overline{\mathcal{T}}'_{A'\to B'}$ be the CPTNI maps such that

$$\hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|\Gamma_A, \Gamma_B) = -D_\alpha(\overline{\mathcal{T}}(\Gamma_A)\|\Gamma_B) \quad \text{and}$$
$$\hat{D}^{\varepsilon'}_{\alpha,A'\to B'}(\zeta_{B'R_{A'}}\|\Gamma_{A'}, \Gamma_{B'}) = -D_\alpha(\overline{\mathcal{T}}'(\Gamma_{A'})\|\Gamma_{B'}) .$$

When we add these two we get

$$\hat{D}^{\varepsilon}_{\alpha,A\to B}(\rho_{BR_A}\|\Gamma_A, \Gamma_B) + \hat{D}^{\varepsilon'}_{\alpha,A'\to B'}(\zeta_{B'R_{A'}}\|\Gamma_{A'}, \Gamma_{B'}) = -\left(D_\alpha(\overline{\mathcal{T}}(\Gamma_A)\|\Gamma_B) + D_\alpha(\overline{\mathcal{T}}'(\Gamma_{A'})\|\Gamma_{B'})\right)$$

Let us apply Lemma 9 over here to obtain

$$-\left(D_\alpha(\overline{\mathcal{T}}(\Gamma_A)\|\Gamma_B) + D_\alpha(\overline{\mathcal{T}}'(\Gamma_{A'})\|\Gamma_{B'})\right) = -D_\alpha\left(\overline{\mathcal{T}}(\Gamma_A) \otimes \overline{\mathcal{T}}'(\Gamma_{A'})\|\Gamma_B \otimes \Gamma_{B'}\right) = -D_\alpha\left(\overline{\mathcal{T}} \otimes \overline{\mathcal{T}}'(\Gamma_A \otimes \Gamma_{A'})\|\Gamma_B \otimes \Gamma_{B'}\right) .$$

Now, we will show that

$$\left(\overline{\mathcal{T}} \otimes \overline{\mathcal{T}}'\right)_{A\otimes A'\to B\otimes B'} \left(\sigma_{AR_A} \otimes \sigma_{A'R_{A'}}\right) \approx_{\varepsilon''} \rho_{BR_A} \otimes \zeta_{B'R_{A'}} \ .$$

We us the inequality of Lemma 6 to obtain

$$\begin{aligned}
& P\left(\left(\overline{\mathcal{T}} \otimes \overline{\mathcal{T}}'\right)_{A\otimes A'\to B\otimes B'} \left(\sigma_{AR_A} \otimes \sigma_{A'R_{A'}}\right), \rho_{BR_A} \otimes \zeta_{B'R_{A'}}\right) \\
&= P\left(\overline{\mathcal{T}}\left(\sigma_{AR_A}\right) \otimes \overline{\mathcal{T}}'\left(\sigma_{A'R_{A'}}\right), \rho_{BR_A} \otimes \zeta_{B'R_{A'}}\right) \\
&\leq \sqrt{P\left(\overline{\mathcal{T}}(\sigma_{AR_A}), \rho_{BR_A}\right)^2 + P\left(\overline{\mathcal{T}}'(\sigma_{A'R_{A'}}), \zeta_{B'R_{A'}}\right)^2} \leq \sqrt{\varepsilon^2 + \varepsilon'^2} = \varepsilon'' \ .
\end{aligned}$$

□

Chapter 5

Information is Physical

In the previous chapters, we considered information and entropy as mathematical concepts. This chapter will be devoted to information and entropy as physical concepts. We give the relevant background knowledge before we present the physical relevance of the coherent relative entropy in Chapter 6. First, we give a self-contained literature review on information as a physical entity. We will argue with a renowned thought experiment that information is indeed physical and can be traded for work. The approach of this experiment is classical, yet we will extend this to a quantum system. In the second part, we discuss the results of more recent studies. In particular, the so-called *information battery*. The idea of this information battery is to treat a quantum register as a work storage system.

5.1 Szilard Engine

Szilard came up with a thought experiment to investigate the infamous Maxwell Demon. The outcome of this experiment gives a link between information and work. For this thought experiment, we use a slightly different one than the original one [8] because it gives a more intuitive notion of what happens. However, the ideas and outcomes are still very much identical. For information on the ideas that are presented here, see Refs. 17–19.

The experiment starts with a cylinder with volume $\mathcal{V}$ containing a single particle, see fig. 5.1 for an illustration of the experiment. The particle is in thermal equilibrium with the cylinder walls. Both ends of the cylinder are blocked by a piston. The particle cannot apply enough force on the pistons to move them further outwards than in the initial situation, step (1) in fig. 5.1. Moreover, the whole set-up is connected to a heat bath making sure the temperature is constant and the particle does not lose thermal energy [2]. The first thing we do is inserting a partition halfway in the cylinder (2). We assume that the amount of work required to insert and remove such partition is negligible. Due to this insertion, there are two compartments, one is empty, the other contains the particle and both have volume $\mathcal{V}/2$. Both possible trajectories are represented in fig. 5.1. Next, we measure in which compartment the particle is without modifying the particle. For now, it is not important how this information is obtained, only that we gained one bit of information. This bit of information needs to be stored somewhere, indicated by the filled box in step (3) in fig. 5.1. It can be in one of three states, a blank state '?' when there is no knowledge, '0' when the particle is in the left compartment and '1' if it is in the right compartment. Note that we now have obtained one bit of information. One can push the piston on the side that does not contain the particle until it touches the partition and can remove the partition (4). Pushing the piston does not cost any work since there are no particles able to resist the movement. After the removal of the partition, the particle now can exert pressure on the just moved piston due to its collisions with the piston (5). Because the set-up is connected to a heat bath, the expansion happens isothermally, and heat is extracted from the bath. We now have lost the information in which compartment the particle is, we lost our bit of information (6). When the piston moves work can indeed be extracted, e.g. connect a pulley to the piston which lifts a weight when it moves. Finally, while the particle is in its original state, the memory system is not. Thus in the last step we reset the memory system to its blank state (7).

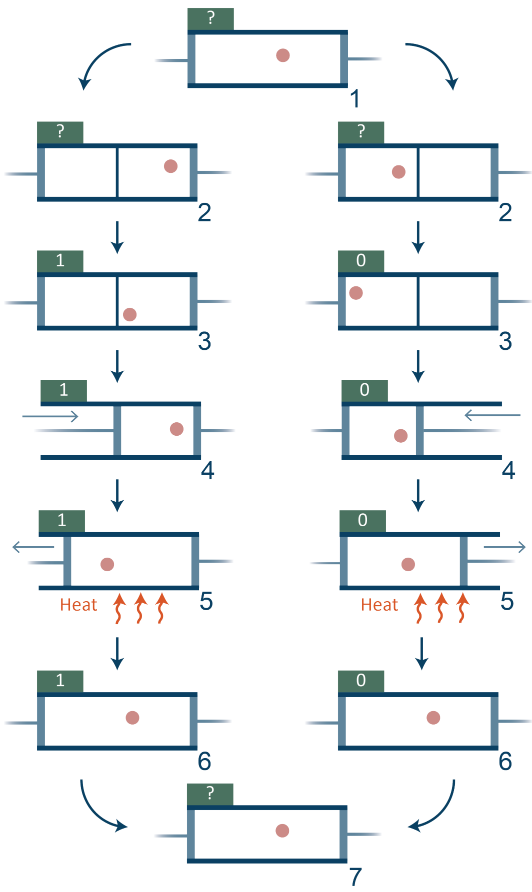

Figure 5.1: *An illustration how the Szilard engine works. Two possible trajectories, one where the particle is the left-hand side and one where the particle is in the right-hand side. Besides, the green box represents the memory system.*

A natural question that arises is how much work can be extracted. This can be calculated with the use of this experiment and the ideal gas law:

$$p = \frac{Nk_BT}{V},$$

where p is the *pressure*, N is the *amount of particles*, k_B is the *Boltzmann constant*, T is the *temperature* and V is the *volume*. In our experiment, we have a gas consisting of one particle, hence $N = 1$. The amount of work done by the system, when the gas changes from a state with volume $\mathcal{V}/2$ to one with volume $\mathcal{V}$, while T is constant all the way is given by

$$W = \int_{\frac{\mathcal{V}}{2}}^{\mathcal{V}} pdV = \int_{\frac{\mathcal{V}}{2}}^{\mathcal{V}} \frac{k_BT}{V} dV = k_BT \left(\ln(\mathcal{V}) - \ln\left(\frac{\mathcal{V}}{2}\right)\right) = k_BT\ln(2)\,.$$

In other words, it is possible to extract $k_BT\ln(2)$ work in this process. After the experiment the system is back at its initial state. The experiment can thus be carried out again and again. After n cycles we have converted $nk_BT\ln(2)$ heat into work.

Result 1. *During one cycle the Szilard engine converts $k_BT\ln(2)$ heat into work.*

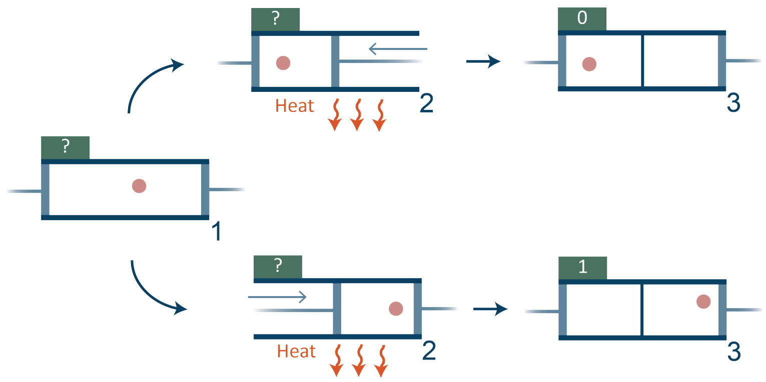

Figure 5.2: *The two possible trajectories for gaining one bit of information.*

We have seen how much work can be extracted while losing information about the position of the particle. Now, we briefly discuss how we can obtain information about the location of a particle and how much works this costs, shown in fig. 5.2. First, the location of the particle is unknown, step (1) in fig. 5.2. Next, by compressing the volume (2) and inserting a partition (3) we know where the particle is. A similar calculation as done for the Szilard engine shows that to obtain this bit we need to invest at least $k_B T \ln(2)$ work. This leads to the following generalized version of Result 1.

Result 2. *From one bit of information $k_B T \ln(2)$ work can be extracted. On the other hand, to gain one bit of information at least $k_B T \ln(2)$ work needs to be invested. Information behaves like a physical quantity.*

5.1.1 Landauer's Principle

There is a catch to the Szilard engine as discussed above, it violates the second law of thermodynamics. In Szilard's thought experiment heat is converted into work without loss. When considering both the system and the heat bath together, there is a decrease of heat and an energy gain due to the performed work. Heat energy is converted into mechanical energy while the system cools down, hence the - physical - entropy has decreased, this violates the second law of thermodynamics.

Indeed there is something we forgot to take into account. If we only look at the particle, then is back at its initial state in step (6) of fig. 5.1. However, in step (7) the memory system is reset. This is exactly where *Landauer's principle* comes into play: Resetting the memory device, erasing the recorded bit, results in a dissipation of heat of at least $k_B T \ln(2)$ [1].

Let us derive Landauer's principle, where we follow a similar approach as presented in Refs. 19, 20. We want to describe a process that resets one bit of information regardless of its initial value. The bit is successfully erased or reset when it is impossible to recover its initial state. We can do this by implementing the irreversible map that sends the value of the bit always to 0. Again, we take a cylinder with one particle inside to model a bit. We assume that we initially know the value of the bit, the particle is either in the left - value 0 the in memory system - or the right compartment - value 1 in the memory system - where a partition makes sure the particle cannot move from one to the other compartment. In fig. 5.3 is shown how the reset process works, starting in situation (1). The resetting process has two steps beginning with removing the partition (2). Next, a piston isothermally compresses the volume and pushing the particle to the left and generates heat (3). The partition is inserted back and the bit has value 0 (4). Note that this system models the irreversible map, hence the initial state of the bit cannot be recovered and the information we initially had has been erased. The value of the bit has been reset. This process requires at least $k_B T \ln(2)$ work to be carried out, so there is a dissipation of $k_B T \ln(2)$ heat which flows from the memory system - the cylinder - to the environment.

Returning to the Szilard engine, the full cycle should also take Landauer's principle into account. We measure where the particle is, this is stored into our memory system. The particle now extracts $k_B T \ln(2)$ work while heat is transferred from the heat bath to the cylinder. The particle is back at its original state. However, the memory system should erase its information for the total system to be in its original state. This is exactly what happens in step (7) in fig. 5.1. According to Landauer's principle, the erasure of this bit of information is associated with at least $k_B T \ln(2)$ heat which is transferred from the memory system to the heat bath. Therefore, a full cycle of the Szilard does not only convert $k_B T \ln(2)$ heat into mechanical work, it generates at least $k_B T \ln(2)$ heat as well. Ergo, the system heats up and there is no apparent violation of the second law of thermodynamics.

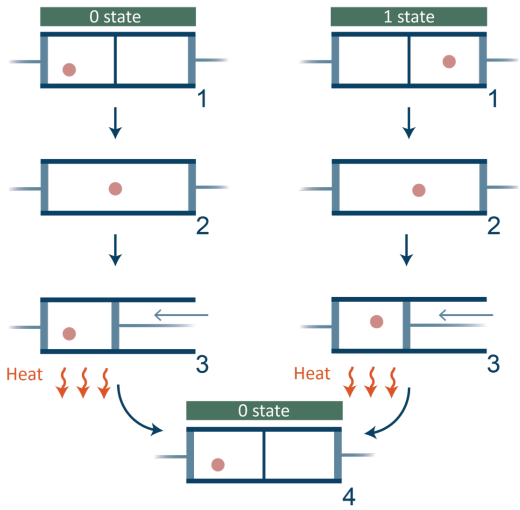

Figure 5.3: *Regardless the value of the bit the resetting process shown in this figure makes sure the bit reset to the 0 value. It is not possible to reverse this process.*

Result 3. *In the Szilard engine $k_B T \ln(2)$ work is extracted from one bit of information. This cools the system down with $k_B T \ln(2)$. However, due to Landauer's principle this cooling is compensated for. Resetting the memory system heats up the whole system with at least $k_B T \ln(2)$. Ultimately, energy conservation is not violated.*

5.2 Information in Quantum Systems

Up to this point, everything has been reviewed in the classical world. Now, we generalize this to quantum systems. We do this to lay the foundation for the physical interpretation of the coherent relative entropy which we discuss in Chapter 6. First, we describe the Szilard engine in a quantum system.

In the Szilard engine, we start with a particle moving freely. The probability of the particle being in the left compartment is $1/2$ and the same for it being in the right compartment. Let $|0\rangle$ be the state where the particle is in the left compartment and $|1\rangle$ for it being in the right one. The particle is initially in the state

$$\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \mathbb{I}/2\,.$$

This is a maximally mixed qubit. When we have information about the location of the particle and a partition is inserted the particle either is in state $|0\rangle\langle 0|$ or $|1\rangle\langle 1|$, hence its state is pure. After one cycle the particle is again in the maximally mixed state. Adopting our findings - Result 2 - from classical systems, we obtain the following result.

Result 4. *When a pure qubit is turned into a maximally mixed qubit we can extract $k_B T \ln(2)$ work.*

5.2.1 Extracting Work from Qubits

We have seen that from transforming a pure qubit into a maximally mixed qubit we can extract $k_B T \ln(2)$ work. The maximum amount of work that can be extracted from a pure qubit is $k_B T \ln(2)$ [21]. When a qubit ρ is in a mixed state its maximum amount of extractable work is given by the following generalization [22] :

$$W = k_B T \ln(2) - k_B T H(\rho) \ .$$

Recall that $H(\rho)$ is the Von Neumann entropy as defined in Definition 12. When we apply this equation to the situation where we start with a pure qubit and end with a maximally mixed qubit, we obtain as the difference the amount of extractable work:

$$\begin{aligned} \Delta W &= k_B T \ln(2) - k_B T S(\rho_{\text{pure}}) - (k_B T \ln(2) - k_B T S(\rho_{\text{mixed}})) \\ &= -k_B T S(\rho_{\text{pure}}) + k_B T S(\rho_{\text{mixed}}) = k_B T \ln(2) \ . \end{aligned}$$

This is what we have expected, one pure qubit allows the extraction of $k_B T \ln(2)$ of work. This leads to a generalization of Result 4.

Result 5. *When a pure qubit is turned into a maximally mixed qubit we can extract $k_B T \ln(2)$ work. When a maximally mixed qubit is turned into a pure qubit we need to perform $k_B T \ln(2)$ work.*

This last result can be generalized to the concept of a *quantum register* - which consists of several qubits - as a work storage system. A quantum register in this context may be called an *information battery* [6, 7, 23, 24]. We use this information battery when describing the set-up for the physical meaning of the coherent relative entropy in Chapter 6. When work can be extracted from a certain process this work can be stored in the information battery. When a certain process can only be carried by performing work, it is possible to use the work that is stored in this information battery. An example of an information battery is shown in fig. 5.4.

An information battery is a system consisting of n qubits. We connect it to a heat bath to make sure the temperature remains constant. This system starts in the state $2^n \mathbb{I}_{2^n}$, which means that all qubits are in maximally mixed state $\frac{1}{2}\mathbb{I}_2$. We say the information battery has a work storage of $n k_B T \ln(2)$, as it is possible to extract $n k_B T \ln(2)$ work from some process and use this work to change the state of all n qubits from maximally mixed to pure.

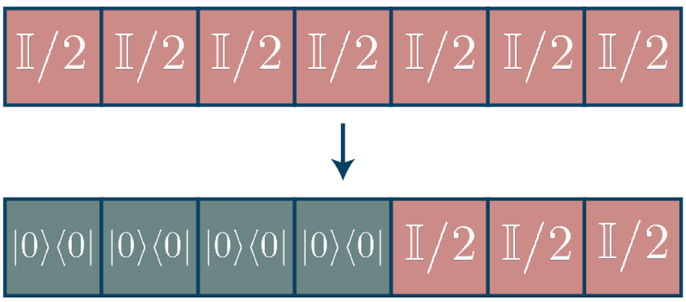

Figure 5.4: *An example of an information battery with 7 qubits. First, all qubits are maximally mixed and the system has a work storage of $7k_B T \ln(2)$. Then we store $4k_B T \ln(2)$ of work in this battery yielding the lower displayed state of the register with 4 pure and 3 maximally mixed qubits.*

For example, imagine a situation like the Szilard engine in which $k_B T \ln(2)$ of work is extracted. This energy can be stored in our information battery, which state changes into e.g.

$$2^{n-1}\mathbb{I}_{2^{n-1}} \otimes |0\rangle\langle 0| \ .$$

In this situation, $n-1$ qubits are in maximally mixed state one qubit is in a pure state $|0\rangle\langle 0|$. Note that $|0\rangle\langle 0|$ is chosen arbitrarily, because it could have been $|1\rangle\langle 1|$ as well. This process can be executed multiple times until our information battery is in the state

$$2^{\lambda}\mathbb{I}_{2^{\lambda}} \otimes |0\rangle\langle 0|^{\otimes(n-\lambda)} \ .$$

Here λ qubits are in maximally mixed state $n-\lambda$ qubits are in pure state. At this point, the information battery has a storage capacity of $\lambda k_B T \ln(2)$ and $(n-\lambda)k_B T \ln(2)$ can be extracted from the information battery.

Chapter 6

Work Cost of Quantum Processes

In this final chapter, we explain in a broader context the physical relevance of the coherent relative entropy as discussed in Chapter 4. For this, we use the relation between information theory and physics as discussed in Chapter 5. In particular, the information battery. Firstly we present the set-up of quantum thermodynamics we here use. Then, we use the coherent relative entropy to determine the work cost of a quantum process in our set-up. This result was shown in Ref. 6. Here, we give a self-contained explanation of that result in a within this context.

6.1 Quantum Thermodynamics

We will start by introducing our framework for quantum thermodynamics. For more in depth discussions on quantum thermodynamics we refer to Refs. 25, 26. Consider a quantum system described by Hamiltonian H_A which is connected to a heat bath. There is also an information battery where work can be extracted and stored. In this system thermalization will take place, the system will equilibrate to the *equilibrium state* given by

$$\gamma_A = \frac{e^{-\beta H_A}}{\mathrm{tr}\left(e^{-\beta H_A}\right)} .$$

Here $\beta = {}^{1}/_{k_B T}$ and this state is often referred to as the *Gibbs state*.

Besides, to each quantum system A we assign a positive semi-definite operator Γ_A. In our framework this will be the non-normalized Gibbs state $\Gamma_A := e^{-\beta H_A}$ with H_A the Hamiltonian of system A. We elaborate on our framework by defining the allowed operations. The allowed operations are the Γ-*sub-preserving maps* which are a CPTNI maps $\mathcal{E}_{A\to B}$ such that $\mathcal{E}_{A\to B}(\Gamma_A) \leq \Gamma_B$. When we carry out a Γ-*sub-preserving map* and extract work, this work can be stored in the information battery. When we want to implement an operation which is not an allowed operation we may extract work from the information battery. Recall that the coherent relative entropy optimizes over al CPTNI maps $\mathcal{T}_{A\to B}$ for which $\mathcal{T}_{A\to B}(\Gamma_A) \leq 2^{-\lambda}\Gamma_B$ holds.

The reason that these states characterise our system and that the Γ-sub-preserving map are the allowed maps is explained by quantum resource theory. Discussing quantum thermodynamics as a quantum resource theory is beyond the scope of this thesis. If the reader wants to dive into this, we recommend Refs. 27–31.

6.2 Work Cost or Gain of Processes

We arrive at one of the main questions, what is the information gain when carrying out an allowed operation? Adding to this, what is the work cost when implementing an operation which is not allowed? It turns out that this can be calculated via the coherent relative entropy. This result was also found and discussed thoroughly in Ref. 6. In fig. 6.1 there is a visual representation of this process. We need two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ with respective matching positive semi-definite operators Γ_A and Γ_B. Both system A and B are connected to a heat bath and there is an information battery X. Let us start in initial state σ_A and apply the CPTNI map $\mathcal{E}_{A\to B}$. Let ρ_{BR_A} be the process matrix of σ_A and $\mathcal{E}_{A\to B}$, then the optimal implementation of this map on this input state is given by the coherent relative entropy:

$$\hat{D}^{\epsilon}_{A\to B}(\rho_{BR_A}\|\Gamma_A,\Gamma_B) = \lambda\ .$$

When λ is positive this means λ pure qubits can be extracted from this process which can be stored in our information battery. Intuitively, when λ is negative $-\lambda$ qubits need to be invested to perform this map. These qubits should be extracted from our information battery. Recalling the relation between information and physics we conclude that the work we can optimally extract is given by

$$W_{\text{extracted}} = \lambda k_B T \ln(2) = \hat{D}^{\epsilon}_{A\to B}(\rho_{BR_A}\|\Gamma_A,\Gamma_B) k_B T \ln(2)\ .$$

Notice that the amount of work needed is also given by $W_{\text{extracted}}$, yet occurs when its value is negative.

We will prove this result in Theorem 3 by showing that an allowed operation in our framework which implements a CPTNI map $\mathcal{T}_{A\to B}$ and changes λ maximally mixed into pure qubits in the information battery, suffices the relation $\mathcal{T}_{A\to B}(\Gamma_A) \leq 2^{-\lambda}\Gamma_B$. This theorem is also presented in proposition 1 of Ref. 6. Moreover, we show that for any CPTNI map that suffices $\mathcal{T}_{A\to B}(\Gamma_A) \leq 2^{-\lambda}\Gamma_B$ there exists a free operation working on our system and the information battery that implements $\mathcal{T}_{A\to B}$ and 'charges' the information battery with λ.

Theorem 3. *Let $\mathcal{T}_{A\to B}$ be a CPTNI map and let $\lambda \in \mathbb{R}$. Then the following are equivalent:*

1. $$\mathcal{T}_{A\to B}(\Gamma_A) \leq 2^{-\lambda}\Gamma_B\ ;$$

2. *Let X be a large information battery system. For any integers $0 \leq \lambda_1, \lambda_2 \leq \dim(X)$ such that $\lambda_1 - \lambda_2 \leq \lambda$, there is a Γ-sub-preserving trace non-increasing map $\mathcal{F}_{AX\to BX}$ such that for all σ_A we have*

$$\mathcal{F}_{AX\to BX}\left(\sigma_A \otimes \left(2^{-\lambda_1}\mathbb{1}_{2^{\lambda_1}}\right)\right) = \mathcal{T}_{A\to B}(\sigma_A) \otimes \left(2^{-\lambda_2}\mathbb{1}_{2^{\lambda_2}}\right)\ .$$

Here, we denote $2^{-\mu}\mathbb{1}_{2^{\mu}}$ for the uniform mixed state of rank μ in the Hilbert space $\mathcal{H}_X$.

Proof. We start by proving 1. $\Rightarrow$ 2.
Let $\lambda_1, \lambda_2 \in \mathbb{Z}_{\geq 0}$ such that $\lambda_1 - \lambda_2 \leq \lambda$ and X an information battery system with $\Gamma_X = \mathbb{1}_X$ and of dimension at least $\max\{\lambda_1, \lambda_2\}$. We define $\mathcal{F}_{AX\to BX}$ as

$$\mathcal{F}_{AX\to BX}(\cdot) := \mathcal{T}_{A\to B}\left(\mathrm{tr}_X\left(\mathrm{id}_A \otimes \mathbb{1}_{2^{\lambda_1}}(\cdot)\right)\right) \otimes \frac{\mathbb{1}_{2^{\lambda_2}}}{2^{\lambda_2}}\ .$$

Then for all σ_A we have $\mathcal{F}_{AX\to BX}\left(\sigma_A \otimes (2^{-\lambda_1}\mathbb{1}_{2^{\lambda_1}})\right) = \mathcal{T}_{A\to B}(\sigma_A) \otimes (2^{-\lambda_2}\mathbb{1}_{2^{\lambda_2}})$. Notice that $\mathcal{F}_{AX\to BX}$ is a composition of trace non-increasing maps and hence is trace non-increasing as well. The only thing left to show is that $\mathcal{F}_{AX\to BX}$ is Γ-sub-preserving map. First, we plug $\Gamma_A \otimes \Gamma_X$ into $\mathcal{F}_{AX\to BX}$ to obtain

$$\begin{aligned}\mathcal{F}_{AX\to BX}(\Gamma_A \otimes \Gamma_X) &= \mathcal{T}_{A\to B}\left(\mathrm{tr}_X\left(\mathrm{id}_A \otimes \mathbb{1}_{2^{\lambda_1}}(\Gamma_A \otimes \mathbb{1}_X)\right)\right) \otimes \frac{\mathbb{1}_{2^{\lambda_2}}}{2^{\lambda_2}} \\ &= 2^{-\lambda_2}\mathcal{T}_{A\to B}\left(2^{\lambda_1}\Gamma_A\right) \otimes \mathbb{1}_{2^{\lambda_2}} = 2^{\lambda_1-\lambda_2}\mathcal{T}_{A\to B}\left(\Gamma_A\right) \otimes \mathbb{1}_{2^{\lambda_2}} ,\end{aligned}$$

where in the last equality we used the property that $\mathcal{T}_{A\to B}$ is a linear function. Next, we use our assumption $\mathcal{T}_{A\to B}(\Gamma_A) \le 2^{-\lambda}\Gamma_B$ and that $\mathbb{1}_{2^{\lambda_2}} \le \mathbb{1}_X = \Gamma_X$ to acquire

$$2^{\lambda_1-\lambda_2}\mathcal{T}_{A\to B}\left(\Gamma_A\right) \otimes \mathbb{1}_{2^{\lambda_2}} \le 2^{\lambda_1-\lambda_2+\lambda}\Gamma_B \otimes \Gamma_X \le \Gamma_B \otimes \Gamma_X .$$

In the last equality we used that $\lambda_1 - \lambda_2 \le \lambda$. This shows that $\mathcal{F}_{AX\to BX}$ is Γ-sub-preserving.

We continue by proving 2. $\Rightarrow$ 1.
Let $\lambda_1, \lambda_2 \in \mathbb{Z}_{\ge 0}$ such that $\lambda_1 - \lambda_2 \le \lambda$. Note that we can retrieve $\mathcal{T}_{A\to B}$ from $\mathcal{F}_{AX\to BX}$:

$$\begin{aligned}&\mathrm{tr}_X\left((\mathrm{id}_B \otimes \mathbb{1}_{2^{\lambda_2}})\,\mathcal{F}_{AX\to BX}\left(\sigma_A \otimes \left(2^{-\lambda_1}\mathbb{1}_{2^{\lambda_1}}\right)\right)\right) = \mathrm{tr}_X\left(\mathcal{T}_{A\to B}(\sigma_A) \otimes \left(2^{-\lambda_1}\mathbb{1}_{2^{\lambda_1}}\right)\right) \\ &= 2^{-\lambda_1+\lambda_1}\mathcal{T}_{A\to B}(\sigma_A) = \mathcal{T}_{A\to B}(\sigma_A) .\end{aligned}$$

By plugging Γ_A into $\mathcal{T}_{A\to B}$ and using that $\mathbb{1}_{2^{\lambda_1}} \le \mathbb{1}_X = \Gamma_X$ we obtain

$$\begin{aligned}\mathcal{T}_{A\to B}(\Gamma_A) &= \mathrm{tr}_X\left((\mathrm{id}_B \otimes \mathbb{1}_{2^{\lambda_2}})\,\mathcal{F}_{AX\to BX}\left(\Gamma_A \otimes \left(2^{-\lambda_1}\mathbb{1}_{2^{\lambda_1}}\right)\right)\right) \\ &\le 2^{-\lambda_1}\,\mathrm{tr}_X\left((\mathrm{id}_B \otimes \mathbb{1}_{2^{\lambda_2}})\,\mathcal{F}_{AX\to BX}\left(\Gamma_A \otimes \Gamma_X\right)\right) .\end{aligned}$$

Now, we use the Γ-sub-preserving property of $\mathcal{F}_{AX\to BX}$ to get

$$\begin{aligned}&2^{-\lambda_1}\,\mathrm{tr}_X\left((\mathrm{id}_B \otimes \mathbb{1}_{2^{\lambda_2}})\,\mathcal{F}_{AX\to BX}\left(\Gamma_A \otimes \Gamma_X\right)\right) \\ &\le 2^{-\lambda_1}\,\mathrm{tr}_X\left(\Gamma_B \otimes \mathbb{1}_{2^{\lambda_2}}\Gamma_X\right) = 2^{-(\lambda_1-\lambda_2)}\Gamma_B .\end{aligned}$$

We retrieve $\mathcal{T}_{A\to B} \le 2^{-\lambda}\Gamma_B$ when sequence (λ_1, λ_2) is chosen such that $\lambda_1 - \lambda_2$ approaches λ . □

We conclude that the coherent relative entropy gives the amount of work extracted or needed such that a CPTNI map on an input state can be carried out.

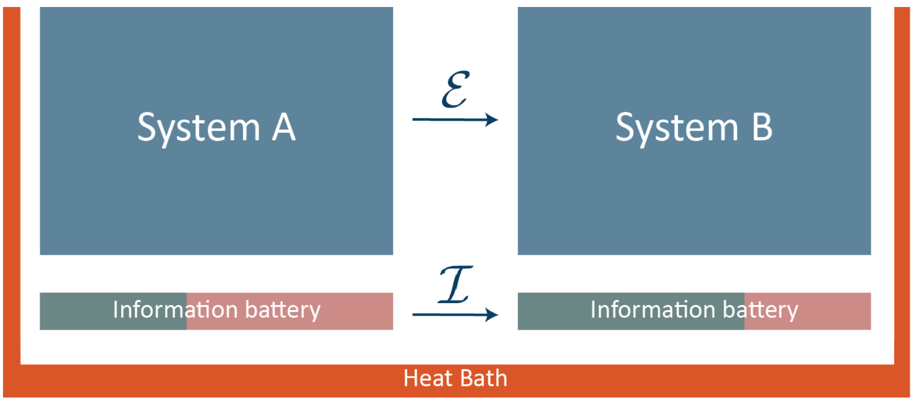

Figure 6.1: *When performing $\mathcal{E}_{A\to B}$ we extract work and store this in the information battery via $\mathcal{I}$. The map $\mathcal{E}_{A\to B} \otimes \mathcal{I}$ is then an allowed operation.*

6.2.1 Example

We illustrate the relation between the coherent relative entropy and work by an example. This example will be the Szilard engine as covered in section 5.1. In that section, we found that one bit can be transformed into $k_B T \ln(2)$ work. As we know the outcome of this experiment, it is a good example to demonstrate our findings.

In this setting we start in a system A where the particle can be in either the left or the right side. The Hamiltonian H_A is such that we have corresponding positive semi-definite operator $\Gamma_A = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \mathbb{I}/2$. Where $|0\rangle$ corresponds to the particle being in the left compartment and $|1\rangle$ to the particle being in the right compartment. We will review two processes. In the first process the particle is forced into the left compartment and is in system B which is essentially the same system. So, we have Hamiltonian $H_B = H_A$ and corresponding positive semi-definite operator $\Gamma_B = \Gamma_A$. The second process is when the particle pushes the piston back and we are back in system A. For convenience we set $\varepsilon = 0$ during both processes.

To start with the first process, we consider the input state $\sigma_A = \mathbb{I}/2$ as the particle is equally likely to be in the left or right compartment. In addition, we consider CPTNI map $\mathcal{E}_{A\to B}(\cdot) = |0\rangle\langle 0|$ as the particle is forced in the left compartment. The work cost or gain is given by the coherent relative entropy:

$$\hat{D}_{A\to B}\left(\rho_{BR_A} \| \mathbb{I}/2, \mathbb{I}/2\right) = \max_{\mathcal{T}_{A\to B}} \left\{ -D_\infty(\mathcal{T}(\mathbb{I}/2) \| \mathbb{I}/2) \mid \mathcal{T}(\sigma_{AR_A}) = \rho_{BR_A} \right\} \ .$$

Because the input state has rank 2 and $\varepsilon = 0$, the only candidate for optimization is $\mathcal{T} = \mathcal{E}$. This gives the following result:

$$\hat{D}_{A\to B}\left(\rho_{BR_A} \| \mathbb{I}/2, \mathbb{I}/2\right) = -D_\infty(\mathcal{T}(\mathbb{I}/2) \| \mathbb{I}/2) = -D_\infty(|0\rangle\langle 0| \| \mathbb{I}/2) = -1 \ .$$

We can interpret this result by saying that forcing a particle to be in the left compartment costs at least $k_B T \ln(2)$ work.

The second process starts in state $\sigma_B = |0\rangle\langle 0|$ as the particle is in the left compartment. Also, we implement CPTNI map $\mathcal{E}_{B\to A}(\cdot) = \mathbb{I}/2$. Again we look at the coherent relative entropy,

$$\hat{D}_{B\to A}\left(\rho_{AR_B} \| \mathbb{I}/2, \mathbb{I}/2\right) = \max_{\mathcal{T}_{B\to A}} \left\{ -D_\infty(\mathcal{T}(\mathbb{I}/2) \| \mathbb{I}/2) \mid \mathcal{T}(\sigma_{BR_B}) = \rho_{AR_B} \right\} \ .$$

The input state is $\sigma_B = |0\rangle\langle 0|$, its purification is $\sigma_{BR_B} = |0\rangle\langle 0| \otimes |0\rangle\langle 0|$, thus the process matrix is $\rho_{AR_B} = \mathbb{I}/2 \otimes |0\rangle\langle 0|$. We optimize over all CPTNI maps $\mathcal{T}$ such that $\mathcal{T}(\sigma_{BR_B}) = \rho_{AR_B}$. This comes down to $\mathcal{T}(|0\rangle\langle 0|) = \mathbb{I}/2$. We want to maximize

$$-D_\infty(\mathcal{T}(\mathbb{I}/2) \| \mathbb{I}/2) = -\log \left\| \mathbb{I}/2^{-\frac{1}{2}} \mathcal{T}(\mathbb{I}/2) \mathbb{I}/2^{-\frac{1}{2}} \right\|_\infty \ .$$

Because the logarithm is monotonic, this means that we want to minimize the largest eigenvalue of $\mathbb{I}/2^{-\frac{1}{2}} \mathcal{T}(\mathbb{I}/2) \mathbb{I}/2^{-\frac{1}{2}}$. Let us have a look at $\mathcal{T}(\mathbb{I}/2)$. When we use the fact that $\mathcal{T}$ is linear, we get

$$\mathcal{T}(\mathbb{I}/2) = \frac{1}{2}\mathcal{T}(|0\rangle\langle 0|) + \frac{1}{2}\mathcal{T}(|1\rangle\langle 1|) = \mathbb{I}/4 + \frac{1}{2}\mathcal{T}(|1\rangle\langle 1|) \ .$$

Within optimization, there is the degree of freedom for $\mathcal{T}(|1\rangle\langle 1|)$ as long as $\mathcal{T}$ is CPTNI. Therefore, when expressing its outcome in the basis $\{|0\rangle, |1\rangle\}$, we get

$$\mathcal{T}(|1\rangle\langle 1|) = \begin{pmatrix} a_0 & a_1 \\ \overline{a_1} & 1 - a_0 - a_2 \end{pmatrix} \qquad \text{and} \qquad \mathcal{T}(\mathbb{I}/2) = \frac{1}{2}\begin{pmatrix} a_0 + 1/2 & a_1 \\ \overline{a_1} & 3/2 - a_0 - a_2 \end{pmatrix},$$

where $a_0 \in [0,1], a_1 \in \mathbb{C}$ and $a_2 \in [0, 1-a_0]$ in order to make sure $\mathcal{T}$ is completely positive and trace non-increasing. Now, we examine $\mathbb{I}/2^{-\frac{1}{2}}\mathcal{T}(\mathbb{I}/2)\mathbb{I}/2^{-\frac{1}{2}}$ and its eigenvalues. We have

$$\mathbb{I}/2^{-\frac{1}{2}}\mathcal{T}(\mathbb{I}/2)\mathbb{I}/2^{-\frac{1}{2}} = \begin{pmatrix} a_0 + 1/2 & a_1 \\ \overline{a_1} & 3/2 - a_0 - a_2 \end{pmatrix} \quad .$$

One can calculate that the eigenvalues of this matrix are

$$\lambda_\pm = 1 - \frac{1}{2}a_2 \pm \frac{1}{2}\sqrt{(2-a_2)^2 + 4a_0^2 - 4a_0 + 4a_0a_2 + 2a_2 + 4|a_1|^2 - 3}\,.$$

Recall we want to minimize the maximum eigenvalue. We notice that the maximum eigenvalue increases if $|a_1|$ increases. Therefore, we set $a_1 = 0$. Our eigenvalues then simplify into

$$\lambda_1 = a_0 + 1/2 \qquad \text{and} \qquad \lambda_2 = 3/2 - a_0 - a_2\,.$$

We observe that $\lambda_1 \geq 1/2$ with equality when $a_0 = 0$. When we also set $a_2 = 1$, we get $\lambda_1 = \lambda_2 = 1/2$. This is the minimum value of the maximum eigenvalue and we found a solution to our optimization:

$$\begin{aligned}\hat{D}_{B\to A}\left(\rho_{AR_B}\|\mathbb{I}/2, \mathbb{I}/2\right) &= \max_{\mathcal{T}_{B\to A}}\{-D_\infty(\mathcal{T}(\mathbb{I}/2)\|\mathbb{I}/2) \mid \mathcal{T}(\rho_{AR_B}) = \sigma_{BR_B}\} \\ &= -\log\left\|\begin{matrix} 1/2 & 0 \\ 0 & 1/2\end{matrix}\right\|_\infty = 1\,.\end{aligned}$$

Just as we expected, it possible to maximally extract $k_B T \ln(2)$ work when this process is carried out. These two processes show that with the coherent relative entropy we can show what we already have seen, one bit of information can be traded for $k_B T \ln(2)$ work.

Discussion

In this thesis, we presented a self-contained research in quantum information theory and some of its applications in quantum thermodynamics. Our contributions strongly rely on the connection between the coherent relative entropy and the Rényi divergence. This relation allowed us to contribute in two ways. First, we generalized the coherent relative entropy to the coherent relative entropy of order α. Secondly, we exploited this relation to prove properties of the coherent relative entropy of order α by using properties of the Rényi divergence of order α. This led to proofs which were sometimes trivial and shorter than when proven from the ground up. In Chapter 5 we gave a research survey on information as a physical concept. The coherent relative entropy also finds its application in quantum thermodynamics as we discussed in Chapter 6. If we want to perform a process from one system to another, given an initial state, the coherent relative entropy tells us the optimal amount of work that can be extracted from this process. The extracted work can then be stored in an information battery where work is stored in the form of pure-state qubits. On the other hand, when a process initially could not be carried out for free, the coherent relative entropy gives the minimal amount of work needed to perform this operation. Again, this work could theoretically be extracted from an information battery.

We have identified a few remaining questions. For example, the coherent relative entropy tells us the work cost of a map from one system to another given an initial state, but this relies on the way the physical system is described. The Γ-sub-preserving maps are argued to be the allowed maps. However, in this thesis, we did not go in-depth into why these maps are indeed the maps that describe nature. Quantum resource theory might give the answer to this.

Lastly, we recommend further research. The results are theoretical, similar experimental findings could strengthen the results found in this thesis. Besides, the new-found coherent relative of order α could be an interesting research topic. More properties of the coherent relative entropy of order α can be deduced from properties of the Rényi divergence. The physical interpretation of the α order coherent relative entropy is also an open question at this moment.

Acknowledgements

I wish to pay my special regards to my supervisors, Serge Fehr and Wolfgang Löffler. Serge Fehr convincingly guided me along the way and always gave valuable feedback. His eager for knowledge goes beyond his expertise, which helped to realize the goal of this project. Wolfgang Löffler his support is characterized by infectious enthusiasm. His approach on the matter gave new useful insights. Altogether, thanks to both supervisors the goal of this thesis has been achieved.

Moreover, I wish to acknowledge the great support of my friends and family. They kept me going on and this work would not have been possible without their support.

References

[1] R. Landauer, *Irreversibility and heat generation in the computing process*, IBM journal of research and development **5**, 183 (1961).

[2] H. S. Leff and A. F. Rex, *Maxwell's Demon : Entropy, Information, Computing.*, Princeton Legacy Library, Princeton University Press, 1990.

[3] P. J. Coles, M. Berta, M. Tomamichel, and S. Wehner, *Entropic uncertainty relations and their applications*, Reviews of Modern Physics **89** (2017).

[4] J. Åberg, *Truly work-like work extraction via a single-shot analysis*, Nature Communications **4** (2013).

[5] M. Horodecki and J. Oppenheim, *Fundamental limitations for quantum and nanoscale thermodynamics*, Nature Communications **4** (2013).

[6] P. Faist and R. Renner, *Fundamental Work Cost of Quantum Processes*, Physical Review X **8** (2018).

[7] P. Faist, *Quantum Coarse-Graining: An Information-Theoretic Approach to Thermodynamics*, 2016.

[8] L. Szilard, *On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings*, Behavioral Science **9**, 301 (1964).

[9] I. L. C. Michael A. Nielsen, *Quantum Computation and Quantum Information*, Cambridge University Press, 10th anniversary edition, 2010.

[10] M. Tomamichel, *Quantum Information Processing with Finite Resources*, SpringerBriefs in Mathematical Physics (2016).

[11] M. Tomamichel, *A Framework for Non-Asymptotic Quantum Information Theory*, 2012.

[12] P. Mendonca, R. Napolitano, M. Marchiolli, C. Foster, and Y.-C. Liang, *Alternative fidelity measure between quantum states*, Physical Review A **78**, 052330 (2008).

[13] M. M. Wilde, *Quantum Information Theory*, (2009).

[14] M. M. Wilde, *From classical to quantum Shannon theory*, Quantum Information Theory , xi.

[15] S. Beigi, *Sandwiched Rényi divergence satisfies data processing inequality*, Journal of Mathematical Physics **54**, 122202 (2013).

[16] D. Sutter, *Approximate quantum Markov chains*, in *Approximate Quantum Markov Chains*, pages 75–100, Springer, 2018.

[17] C. H. Bennett, *The thermodynamics of computation-a review*, International Journal of Theoretical Physics **21**, 905 (1982).

[18] C. H. Bennett, *Demons, engines and the second law*, Scientific American **257**, 108 (1987).

[19] K. Maruyama, F. Nori, and V. Vedral, *Colloquium: The physics of Maxwell's demon and information*, Reviews of Modern Physics **81**, 1 (2009).

[20] M. B. Plenio and V. Vitelli, *The physics of forgetting: Landauer's erasure principle and information theory*, Contemporary Physics **42**, 25 (2001).

[21] M. F. Frenzel, D. Jennings, and T. Rudolph, *Reexamination of pure qubit work extraction*, Physical review E **90**, 052136 (2014).

[22] R. Alicki, M. Horodecki, P. Horodecki, and R. Horodecki, *Thermodynamics of quantum information systems-Hamiltonian description*, Open Systems & Information Dynamics **11**, 205 (2004).

[23] S.-Y. Wang, *Informational work storage in quantum thermodynamics*, 2017.

[24] P. Faist, F. Dupuis, J. Oppenheim, and R. Renner, *The minimal work cost of information processing*, Nature communications **6**, 1 (2015).

[25] S. Vinjanampathy and J. Anders, *Quantum thermodynamics*, Contemporary Physics **57**, 545 (2016).

[26] R. Alicki and R. Kosloff, *Introduction to Quantum Thermodynamics: History and Prospects*, arXiv: Quantum Physics , 1 (2018).

[27] E. Chitambar and G. Gour, *Quantum resource theories*, Reviews of Modern Physics **91** (2019).

[28] J. Goold, M. Huber, A. Riera, L. d. Rio, and P. Skrzypczyk, *The role of quantum information in thermodynamics-a topical review*, Journal of Physics A: Mathematical and Theoretical **49**, 143001 (2016).

[29] N. H. Y. Ng and M. P. Woods, *Resource Theory of Quantum Thermodynamics: Thermal Operations and Second Laws*, Thermodynamics in the Quantum Regime , 625 (2018).

[30] M. Lostaglio, *An introductory review of the resource theory approach to thermodynamics*, Reports on Progress in Physics **82**, 114001 (2019).

[31] P. Faist, J. Oppenheim, and R. Renner, *Gibbs-preserving maps outperform thermal operations in the quantum regime*, New Journal of Physics **17**, 043003 (2015).