Universitat Autònoma de Barcelona

From Quantum Source Compression to Quantum Thermodynamics

by

Zahra Baghali Khanian

under supervision of

Prof. Andreas Winter

and co-supervision of

Prof. Maciej Lewenstein

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in

Unitat de Física Teòrica: Informació i Fenòmens Quàntics
Departament de Física
Facultat de Ciències

Phd examination committee:
Prof. John Calsamiglia
Prof. Patrick Hayden
Prof. Michał Horodecki

Bellaterra, October, 2020

Grup d'Informació Quàntica

Abstract

This thesis addresses problems in the field of quantum information theory, specifically, quantum Shannon theory. The first part of the thesis is opened with concrete definitions of general quantum source models and their compression, and each subsequent chapter addresses the compression of a specific source model as a special case of the initially defined general models. First, we find the optimal compression rate of a general mixed state source which includes as special cases all the previously studied models such as Schumacher's pure and ensemble sources and other mixed state ensemble models. For an interpolation between the visible and blind Schumacher's ensemble model, we find the optimal compression rate region for the entanglement and quantum rates. Later, we comprehensively study the classical-quantum variation of the celebrated Slepian-Wolf problem and find the optimal rates considering per-copy fidelity; with block fidelity we find single letter achievable and converse bounds which match up to continuity of a function appearing in the bounds. The first part of the thesis is closed with a chapter on the ensemble model of quantum state redistribution for which we find the optimal compression rate considering per-copy fidelity and single-letter achievable and converse bounds matching up to continuity of a function which appears in the bounds.

The second part of the thesis revolves around information theoretical perspective of quantum thermodynamics. We start with a resource theory point of view of a quantum system with multiple non-commuting charges where the objects and allowed operations are thermodynamically meaningful; using tools from quantum Shannon theory we classify the objects and find explicit quantum operations which map the objects of the same class to one another. Subsequently, we apply this resource theory framework to study a traditional thermodynamics setup with multiple non-commuting conserved quantities consisting of a main system, a thermal bath and batteries to store various conserved quantities of the system. We state the laws of the thermodynamics for this system, and show that a purely quantum effect happens in some transformations of the system, that is, some transformations are feasible only if there are quantum correlations between the final state of the system and the thermal bath.

Resum

Aquesta tesi aborda problemes en el camp de la teoria de la informació quàntica, específicament, la teoria quàntica de Shannon. La primera part de la tesi comença amb definicions concretes de models de fonts quàntiques generals i la seva compressió, i cada capítol següent aborda la compressió d'un model de font específic com a casos especials dels models generals definits inicialment. Primer, trobem la taxa de compressió òptima d'una font d'estats barreja general que inclou com a casos especials tots els models prèviament estudiats, com les fonts pures i de col·lectivitats de Schumacher, i altres models de col·lectiuvitats d'estats barreja. Per a una interpolació entre els models de col·lectivitats visible i cec de Schumacher, trobem la regió de compressió òptima per les taxes d'entrellaçament i les taxes quàntiques. A continuació, estudiem exhaustivament la variació clàssic-quàntica del famós problema de Slepian-Wolf i trobem les taxes òptimes considerant la fidelitat per còpia; per la fidelitat de bloc trobem expressions tancades per les fites assolibles i inverses que coincideixen, sota la condició de que una funció que apareix a les dues fites sigui continua. La primera part de la tesi tanca amb un capítol sobre el model de col·lectivitats per la redistribució d'estats quàntics per al qual trobem la taxa de compressió òptima considerant la fidelitat per còpia i les fites assolibles i inverses, que de nou que coincideixen sota la condició de continuïtat d'una certa funció.

La segona part de la tesis gira al voltant de la termódinamica quàntica sota de la perspectiva de la teoria de la informació. Comencem amb un punt de vista de la teoria de recursos d'un sistema quàntic amb múltiples càrregues que no commuten i amb objectes i operacions permeses que son termodinàmicament significatives; utilitzant eines de la teoria quàntica de Shannon classifiquem els objectes i trobem operacions quàntiques explícites que relacionen els objectes de la mateixa classe entre sí. Posteriorment, apliquem aquest marc de la teoria de recursos per estudiar una configuració termodinàmica tradicional amb múltiples quantitats conservades que no commuten que consta d'un sistema principal, un reservori calòric i bateries per emmagatzemar diverses quantitats conservades del sistema. Enunciem les lleis de la termodinàmica per a aquest sistema, i mostrem que un efecte purament quàntic té lloc en algunes transformacions del sistema, és a dir, algunes transformacions només són factibles si hi ha correlacions quàntiques entre l'estat final del sistema i del reservori calòric.

Resumen

Esta tesis aborda problemas en el campo de la teoría de la información cuántica, específicamente, la teoría cuántica de Shannon. La primera parte de la tesis comienza con definiciones concretas de modelos de fuentes cuánticas generales y su compresión, y cada capítulo subsiguiente aborda la compresión de un modelo de fuente específico como casos especiales de los modelos generales definidos inicialmente. Primero, encontramos la tasa de compresión óptima de una fuente de estado mixto general que incluye como casos especiales todos los modelos previamente estudiados, como las fuentes pura y colectiva de Schumacher, y otros modelos colectivos de estado mixto. Para una interpolación entre el modelo colectivo visible y ciego de Schumacher, encontramos la región de tasa de compresión óptima para el entrelazamiento y las tasas cuánticas. A continuación, estudiamos exhaustivamente la variación clásico-cuántica del célebre problema de Slepian-Wolf y encontramos las tasas óptimas considerando la fidelidad por copia; con la fidelidad de bloque encontramos límites alcanzables e inversos que coinciden con la continuidad de una función que aparece en los límites. La primera parte de la tesis cierra con un capítulo sobre el modelo colectivo de redistribución de estado cuántico para el cual encontramos la tasa de compresión óptima considerando la fidelidad por copia y los límites alcanzables e inversos que coinciden con la continuidad de una función que aparece en los límites.

La segunda parte de la tesis gira en torno a la perspectiva teórica de la información de la termodinámica cuántica. Comenzamos con un punto de vista de la teoría de recursos de un sistema cuántico con múltiples cargas no conmutables con objetos y operaciones permitidas que son termodinámicamente significativas; usando herramientas de la teoría cuántica de Shannon clasificamos los objetos y encontramos operaciones cuánticas explícitas que mapean los objetos de la misma clase entre sí. Posteriormente, aplicamos este marco de la teoría de recursos para estudiar una configuración termodinámica tradicional con múltiples cantidades no conmutables compuesta por un sistema principal, un reservorio calórico y baterías para almacenar varias cantidades conservadas del sistema. Enunciamos las leyes de la termodinámica para este sistema, y mostramos que ocurre un efecto puramente cuántico en algunas transformaciones del sistema, es decir, algunas transformaciones solo son factibles si existen correlaciones cuánticas entre el estado final del sistema y del reservorio calórico.

Acknowledgements

I express my sincere gratitude to my supervisors Andreas Winter and Maciej Lewenstein for their continuous support and care in any aspect that I could possibly ask for. I started with studying various problems in quantum Shannon theory, and I enjoyed and learned from immense knowledge of Andreas Winter who gave me unlimited freedom and time to submerge myself in problems and wrap my head around them; I have been fascinated to see his perspective, scientific discipline and how he approaches science in general, and I feel privileged to have him as my mentor and role model both in academic and personal life. He later introduced me to Maciej Lewenstein in ICFO where I have learned from his profound knowledge in physics and how to make sense of complicated mathematical notions through physical interpretations without obsessing about equations. I cannot thank Maciej enough for his kindness, support and also the freedom and time that he gave me.

I am honored and delighted to defend my thesis in front of the experts of the field John Calsamiglia, Patrick Hayden and Michał Horodecki, whose scientific works have been a source of inspiration and guidance to me.

I have learned a lot and enjoyed discussing problems during my academic visits that I have had, specifically, I would like to thank Paul Skrzypczyk and Tony Short in the University of Bristol, Nilanjana Datta in Cambridge and Masahito Hayashi in Peng Cheng Lab.

Apart from scientific perspective, I got to enjoy my time in ICFO as a Phd student which is a great institute for anyone seeking professional academic training thanks to the organization, generosity and supportive environment of the institute. I am indebted to both academic and administrative staff. Moreover, thanks to the social environment there, I have made great friendships and enjoyed the fun specifically during various annual events.

I had great time and experience in UAB in our own quantum information group (Giq), where it is my academic home, thanks to the supportive and encouraging atmosphere that has been fostered here. I am indebted to Anna, Emili, John, Ramon and other Giq members for all their support and care; I have enjoyed the seminars, the time we have spent together during lunch and our annual Calçotadas.

I could never accomplish what I have accomplished so far without my background and the training that I have had in great schools in Iran, I specifically thank my teachers in Sharif university of technology where I was exposed to information theory and I got fascinated for the first time about quantum information. I am specifically grateful to Salman Beigi and Amin

Contents

Chapter 1

Introduction

1.1 Background and motivation

Information theory studies the transmission, processing, extraction, and utilization of information. The notion of classical information was first introduced by Shannon [1], who defined it operationally, as the minimum number of bits needed to communicate the message produced by a statistical source. This gave meaning to the Shannon entropy $H(X)$ of a source producing a random variable X. The amount of information that two random variables X and Y have in common was given a meaning through the mutual information $I(X:Y)$. Operationally it is the rate of communication possible through a noisy channel taking X to Y.

Quantum Shannon theory is a more general field which studies information on physical systems governed by the rules of quantum mechanics, therefore encompasses classical information as sub-field, and was mathematically founded by Holevo in 1973 [2] to study the transmission of information over quantum channels following the earliest understanding of the connection between quantum physics and information theory [3–6].

Surprisingly, von Neumann entropy, which is a generalization of Shannon entropy, was formulated before Shannon entropy in the context of thermodynamics and statistical mechanics, and it was not contemplated to convey informational interpretation. Despite this fact and Holevo's study of classical information on quantum systems [2,7], the concept of quantum information was obscure till 1995, when Schumacher showed that the von Neumann entropy has the operational interpretation of the number of *qubits* needed to transmit quantum states emitted by a statistical source [8].

After Schumacher's quantitative notion of quantum information, i.e. qubit, and understanding its complementary nature to classical information, quan-

tum Shannon theory has been further established in the last three decades by fundamental discoveries from source and channel coding to quantum cryptography, quantum error-correcting, measures of entanglement and so on [9–22].

In particular, the notion of a quantum source as a quantum state together with correlations with a reference system and its compression led to the discovery of operational meaning for quantum quantities such as quantum conditional entropy, which as opposed to its classical counterpart can obtain negative values. In this source compression task with side information, which is called state merging, the negative values of conditional entropy imply that the entanglement is generated after the compression is accomplished, and it can be used as a resource for future communications [23, 24].

Other quantum source compression problems such as quantum state redistribution and visible compression of mixed states gave operational meaning to quantum conditional mutual information and regularized entanglement of purification [25–27], respectively, and they have been used successfully as sub-protocols to accomplish tasks other than data compression [28]. Various source models and their compression have been considered throughout these years and each source appeared to be a distinct case with a unique compression behavior [23, 25–27, 29–31], and the compression of many other source models has been left open [30, 32]. These open questions and the lack of a source model, which can unify all these seemingly distinct models, is the underlying motivation for the first part of this thesis which focuses on the compression of quantum sources. We specifically solve the Schumacher's compression problem when the overall state together with the reference is a general mixed state. When there are side information systems, a general reference system appears to be hard to tackle, therefore we attack compression problems with classical references or so called ensemble sources.

Understanding compression and capacity problems apart from finding fundamental limits on the amount of communication and storage rates, has developed tools and quantitative notions, e.g typical subspaces and entropic quantities, which has been successfully used to deal with and interpret other quantum effects such as quantum thermodynamics and quantum coherence [33–35]. In particular, the innate relationship between information theory and thermodynamics has proved that integrated ideas from both fields are fruitful [36–39]. This has been the motivation for the second part of this thesis which focuses on quantum thermodynamics, where we consider a general framework with multiple conserved quantities and apply information theoretic tools to construct charge conserving operations. These explicit operations are extremely helpful to study traditional thermodynamics settings and laws.

Perhaps the most up-to-date and comprehensive review of the fast-growing

field of quantum thermodynamics is contained in the collection of essays in [40]. Still, some very fundamental questions concerning quantum thermodynamics have been answered in this thesis. For a non-specialist, These questions can be formulated as follows. Normally in both classical and quantum thermodynamics one deals with large system interacting with an even larger bath. In addition to energy, the system maybe characterized by many macroscopic conserved (in average) quantities, called here charges like total electric charge, total dipole moment, angular momentum, magnetization, total spin components etc. In quantum case, these quantities may correspond to non-commuting operators. How come that with repeated measurements on the system prepared in the same state, we obtain well defined average values of this charges? The repeated measurements of equally prepared systems can be mathematically treated by considering tensor product states of many copies of the systems. This mathematical construction is used in the thesis to define thermodynamically allowed transformation, which have to conserve all average values of the charges. To any quantum state, we associate a vector with entries of the expected charge values and entropy of that state. The set of all these vectors forms the phase diagram of the system, and show that it characterizes the equivalence classes of states under thermodynamically allowed transformations, which are proven rigorously to correspond to asymptotic unitary transformations that approximately conserve the charges.

Our theory provides a general theoretical framework, but leads also prediction of very concrete effects. In particular, we estimate how large an asymptotically large bath is necessary to attain the second law of thermodynamics, and permit a specified work transformation of a given system. In some situations, the necessary bath extension is relatively small, and then quantum setting requires an extended phase diagram exhibiting negative entropies. This corresponds to the purely quantum effect that at the end of the process, system and bath are entangled. Obviously, such processes are impossible classically! For large thermal bath, thermodynamically allowed transformation leave the system and the bath uncorrelated. In such case, the heat capacity of the bath becomes a function of how tightly the second law is attained.

1.2   The structure of the thesis

The reminder of this chapter is dedicated to introduce some notation and preliminary material, which are prerequisite for the subsequent chapters. In summary as mentioned above, the thesis is based on two main themes: part I and part II revolving around quantum source compression and quantum

thermodynamics, respectively.

As for the source compression part, we start with chapter 2 where we first expand on the notion of a quantum source and continue with a rigorous definition of asymptotic source compression task which encompasses as special cases all the reviewed compression problems in the context of asymptotic quantum source compression and unifies them under a common base. Later in the chapter, we define side information and distributed settings for compressing the information. As for the resources available for communication, we consider noiseless qubit channel and shared entanglement between the parties.

In chapter 3, we consider the most general source where the overall state with the reference system is a general mixed state. This model covers all the previously studied models such as Schumacher's ensemble and pure sources [8] and the ensemble of mixed state source [31]. We find the optimal trade-off between the entanglement and quantum communication rates. The optimal rates are in terms of a decomposition of the source introduced in [41] which is a generalization of the well-known decomposition discovered by Koashi and Imoto in [42]. When there are side information systems or the compression task is distributed, the general models defined in chapter 2 appears to be very complicated, and even much simpler models have been left open since the early exploration of the compression problems [29, 30, 32]; therefore in the subsequent chapters we consider special cases where the source states are from an ensemble, that is the reference system is partly classical.

In chapter 4, we consider an interpolation between visible and blind Schumacher compression, that is the encoder has access to a side information system which can reduce to a classical system with the information about the identity of the states and a trivial system in the visible and blind scenarios, respectively. We find the optimal trade-off between the entanglement and quantum rates which depending on whether the ensemble is reducible or not, the entanglement consumption reduces the quantum rate or does not help it at all.

Chapter 5 is about the distributed compression of a hybrid classical-quantum source which is an extension of the celebrated Slepian-Wolf problem [43]. Two important sub-problems of this distributed compression problem are classical data compression with quantum side information (at the decoder), which is addressed in [29,30], and quantum data compression with classical side information (at the decoder), which is the main focus of this chapter. For a class of generic sources we show that the compression rate can be strictly larger than the conditional entropy contrary to the fully classical problem of Slepian-Wolf where the rate of the side information case is always governed by the conditional entropy. However, in general the quantum com-

pression rate reduces by a factor of half of the mutual information between the classical variable and the environment system of the encoder.

Chapter 6 closes the first part of the thesis where we consider the most general ensemble model of pure states with side information available both at the encoder and decoder side. When the overall state of the parties and the reference system is pure, the problem is known as quantum state redistribution [25, 26, 44]. We find the optimal quantum compression rate and confirm that preserving correlations with a a hybrid classical-quantum reference, which is less stringent than preserving the correlations with the purified reference, can lead to strictly smaller quantum rates. Indeed, this model includes as special cases the sources considered in chapter 4 and chaprer 5, however, in the former chapter the figure of merit is block fidelity whereas in the last two chapters the optimal rates are obtained by considering per-copy fidelity; considering block fidelity in the last two chapters, we find upper and lower bounds which would match if the corresponding function defining the bounds is continuous.

The second part of the thesis consists of two chapters. In chapter 7, we develop a general resource theory with allowed operations which are thermodynamically meaningful. The objects of this resource theory are quantum states and the allowed operations are those asymptotically commuting with a general set of charges associated with the quantum system. In order to explicitly construct these operations we use tools and notions such as quantum typicality and approximate microcanonical subspace. Later in chapter 8, we use the developed operations to study a traditional thermodynamics setting with multiple conserved quantities consisting of a work system, a thermal bath and many batteries to store each charge. We extend the notion of charge-entropy diagram to a diagram with conditional entropy to find out which transformations are feasible and show that some transformations are feasible only if the final states of the work system and the thermal bath are entangled, i.e. a purely quantum effect enlarges the set of feasible transformations for the work system.

Finally, the last six chapters are essentially based on the following publications and preprints:

- **Chapter 3:**

  [45] Z. B. Khanian and A. Winter, "General mixed state quantum data compression with and without entanglement assistance," *pre-print (2019)*, arXiv: 1912.08506.

  [46] Z. B. Khanian and A. Winter, "Entanglement-assisted quantum data compression," in: *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*,

Los Angeles, CA, USA, pp. 1852-1857, June 2020.

- **Chapter 4:**

  [47] Z. B. Khanian and A. Winter, "Entanglement-assisted quantum data compression," *preprint (2019)*, arXiv: 1901.06346.

  [48] Z. B. Khanian and A. Winter, "Entanglement-assisted quantum data compression," in: *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, pp. 1147–1151, July 2019.

- **Chapter 5:**

  [49] Z. B. Khanian and A. Winter, "Distributed compression of correlated classical-quantum sources or: the price of ignorance," *IEEE Trans. Inf. Theory*, vol. 66, no. 9, pp. 5620-5633, Sep 2020. arXiv: 1811.09177.

  [50] Z. B. Khanian and A. Winter, "Distributed compression of correlated classical-quantum sources," in: *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, pp. 1152-1156, July 2019.

- **Chapter 6:**

  [51] Z. B. Khanian and A. Winter, "Rate distortion perspective of quantum state redistribution," *in preparation*.

  [52] Z. B. Khanian and A. Winter, "Quantum state redistribution for ensemble sources," in: *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Los Angeles, CA, USA, pp. 1858-1863, June 2020.

- **Chapter 7 and Chapter 8:**

  [53] Z. B. Khanian, M. N. Bera, A. Riera, M. Lewenstein and A. Winter, "Resource theory of heat and work and everything else," *in preparation*.

During my Phd, I have also worked on the following thermodynamics project which is not included in this thesis:

- [54] M. N. Bera, A. Riera, M. Lewenstein, Z. B. Khanian, and A. Winter, "Thermodynamics as a Consequence of Information Conservation," *Quantum*, vol. 3, 2018. arXiv[quant-ph]:1707.01750.

1.3 Notation and preliminaries

In this section, we introduce some conventions, notation and facts that we use throughout this thesis.

Quantum systems are associated with (finite dimensional) Hilbert spaces A, B, etc., whose dimensions are denoted $|A|$, $|B|$, respectively. The state of such quantum system is entirely characterized by a density operator, say ρ, acting on the associated Hilbert space which is a positive semidefinite operator with trace 1. Also, we use the notation $\phi = |\phi\rangle\langle\phi|$ to denote the density operator of the pure state vector $|\phi\rangle$. Moreover, a system is called classical if all the states of the system are diagonal in a fixed orthonormal basis.

The evolution of a quantum system is characterized by a quantum channel or a so-called completely positive and trace preserving (CPTP) map which is a linear map taking operators on a Hilbert space to operators on the same or a different Hilbert space [55], however, since there is no risk of confusion, we denote a CPTP map by the input and output Hilbert spaces, for example, the operator $\mathcal{N} : A \longrightarrow B$ takes the input state ρ on A to the output state $\mathcal{N}(\rho)$ on B.

Furthermore, according to Stinespring's factorization theorem [55], if $\mathcal{N} : A \longrightarrow B$ is a CPTP map, then it can be dilated to the isometry $U_{\mathcal{N}} : A \hookrightarrow BW$ with W as the environment system such that $\mathcal{N}(\rho) = \mathrm{Tr}_W(U_{\mathcal{N}}\rho U_{\mathcal{N}}^{\dagger})$ where $\mathrm{Tr}_W(\cdot)$ denotes the partial trace on system W.

The fidelity, which is a measure of closeness, between two states ρ and σ is defined as [56]

$$F(\rho,\sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1 = \mathrm{Tr}\sqrt{\rho^{\frac{1}{2}}\sigma\rho^{\frac{1}{2}}}, \tag{1.1}$$

where the trace norm is defined as

$$\|X\|_1 := \mathrm{Tr}|X| = \mathrm{Tr}\sqrt{X^{\dagger}X} \tag{1.2}$$

It relates to the trace distance in the following well-known way [57]:

$$1 - F(\rho,\sigma) \leq \frac{1}{2}\|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho,\sigma)^2}. \tag{1.3}$$

The *von Neumann entropy* of a quantum state ρ on system A is defined as

$$S(\rho)_A := -\mathrm{Tr}\,\rho\log\rho, \tag{1.4}$$

where throughout this thesis, log denotes by default the binary logarithm, and its inverse function exp, unless otherwise stated, is also to basis 2. $S(\rho)_A$ is also denoted by $S(A)_\rho$. For the diagonalization of ρ, i.e $\rho = \sum_x p_x |v_x\rangle\langle v_x|$ with orthonormal basis $\{|v_x\rangle\}$, the von Neumann entropy reduces to the *Shannon entropy* $H(X)$ of a random variable X with probability distribution p_x:

$$H(X) \coloneqq -\sum_x p_x \log p_x = S(\rho)_A. \qquad (1.5)$$

The von Neumann entropy is always bounded as the following:

$$0 \le S(\rho) \le \log|A|, \qquad (1.6)$$

where $|A|$ is the dimension of the underlying Hilbert space of ρ, i.e. the support of ρ. Moreover, $S(\rho) = 0$ if and only if ρ is a pure state, and $S(\rho) = \log|A|$ if and only if it is a maximally mixed state, i.e. $\rho = \frac{\mathbb{1}}{\log|A|}$.

The mutual information for a state ρ^{AB} on a bipartite Hilbert space $A \otimes B$ is defined as:

$$I(A:B) \coloneqq S(A)_\rho + S(B)_\rho - S(AB)_\rho, \qquad (1.7)$$

which is always non-negative due to sub-additivity of the von Neumann entropy [58] and is equal to 0 if and only if $\rho^{AB} = \rho^A \otimes \rho^B$, that is an uncorrelated state.

Quantum conditional entropy and quantum conditional mutual information, $S(A|B)_\rho$ and $I(A:B|C)_\rho$, respectively, are defined in the same way as their classical counterparts:

$$S(A|B)_\rho \coloneqq S(AB)_\rho - S(B)_\rho, \text{ and}$$
$$I(A:B|C)_\rho \coloneqq S(A|C)_\rho - S(A|BC)_\rho$$
$$= S(AC)_\rho + S(BC)_\rho - S(ABC)_\rho - S(C)_\rho. \qquad (1.8)$$

Quantum conditional entropy can acquire negative values, however, it is always positive if at least one of the systems A or B is classical. Araki-Lieb inequality holds for the conditional entropy as the following [58]:

$$-S(A)_\rho \le S(A|B)_\rho \le S(A)_\rho, \qquad (1.9)$$

where the inequality on the right hand side is known as sub-additivity of the entropy. Quantum conditional mutual information is always positive due to strong sub-additivity of the entropy as the following [59]:

$$S(A|BC)_\rho \le S(A|C)_\rho. \qquad (1.10)$$

8

The quantum *relative entropy* between two quantum states ρ and σ is defined as:

$$D(\rho\|\sigma) := \begin{cases} \mathrm{Tr}\left(\rho(\log\rho - \log\sigma)\right) & \mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma) \\ \infty & \text{otherwise,} \end{cases} \qquad (1.11)$$

which is always non-negative. Pinsker's inequality [19] relates the quantum relative entropy and the trace norm by

$$\|\rho - \sigma\|_1 \leq \sqrt{2\ln 2\, D(\rho\|\sigma)}. \qquad (1.12)$$

Part I

Quantum Source Compression

Chapter 2

Formulation of quantum source compression problems

In this chapter, we first expand on the concept of quantum sources and the literature on that and mathematically define an asymptotic compression task as a general model which include all previously studied asymptotic models. Then, we introduce quantum compression problems with side information and review the literature, and later we proceed with defining the most general asymptotic compression task with side information. Finally at the end of the chapter, we summarize the results that we have accomplished on quantum source compression.

2.1   What is a quantum source?

A statistical quantum source is a quantum system together with correlations with a *reference system*. A criterion of how well a source is reproduced in a communication task is to measure how well the correlations are preserved with the reference system. Without correlation, the information does not make sense because a known quantum state without correlations can be reproduced at the destination without any communication.

A special case is a classical statistical source, which is modeled by a random variable. Since classical information can be copied, a copy of a random variable can be always stored as a reference, and the final processed information is compared with the copy as a reference to analyse the performance of the communication task. However, in the classical information theory literature, the reference is not usually considered explicitly in the description of classical information theory tasks, but arguably it is conceptually necessary in quantum information. This is because it allows us to present the figure of

merit quantifying the decoding error as operationally accessible, for example via the probability of passing a test in the form of a measurement on the combined source and reference systems. This point is made eloquently in the early work of Schumacher on quantum information transmission [11, 13].

To elaborate more on the reference system, consider the source that Schumacher defined in his 1995 paper [8, 60] as an ensemble of pure states $\{p(x), |\psi_x\rangle^A\}$, where the source generates the state $|\psi_x\rangle$ with probability $p(x)$. The figure of merit for the encoding-decoding process is to keep the decoded quantum states *on average* very close to the original states with respect to the fidelity, where the average is taken over the probability distribution $p(x)$. By basic algebra one can show that this is equivalent to preserving the classical-quantum state $\rho^{AX} = \sum_x p(x) |\psi_x\rangle\langle\psi_x|^A \otimes |x\rangle\langle x|^X$, where system A is the quantum system to be compressed and X is the reference system; namely the following fidelity relation holds:

$$\sum_x p(x) F(|\psi_x\rangle\langle\psi_x|^A, \xi_x^{\hat{A}}) = F(\rho^{AX}, \xi^{\hat{A}X}),$$

where $\xi_x^{\hat{A}}$ is the decoded state for the realization x and $\xi^{\hat{A}X} = \sum_x p(x) \xi_x^{\hat{A}} \otimes |x\rangle\langle x|^X$. Another source model that Schumacher considered was the purification of the source ensemble, that is the state $|\psi\rangle^{AR} = \sum_x \sqrt{p(x)} |\psi_x\rangle^A |x\rangle^R$, where the figure of merit for the encoding-decoding process was to preserve the pure state correlations with the reference system R by maintaining a high fidelity between the decoded state and ψ. He showed that both definitions lead to the same compression rate, namely, the von Neumann entropy of the source $S(A)_\rho = S(\rho^A)$, where $\rho^A = \text{Tr}_R \rho^{AR}$. Incidentally, the full proof of optimality in the first model, without any additional restrictions on the encoder, had to wait until [61] (see also [62]); the strong converse, i.e. the optimality of the entropy rate even for constant error bounded away from 1, was eventually given in [30].

Another example of a quantum source is the mixed state source considered by Horodecki [62] and Barnum *et al.* [63], and finally solved by Koashi and Imoto [31], where the source is defined as an ensemble of mixed states $\{p(x), \rho_x^A\}$. Preserving these mixed quantum states, on average, in the process of encoding-decoding, the task is equivalent to preserving the state $\rho^{AX} = \sum_x p(x) \rho_x^A \otimes |x\rangle\langle x|^X$, that is the quantum system A together with its correlation with the classical reference system X.

In this thesis, we consider the most general finite-dimensional source in the realm of quantum mechanics, namely a quantum system A that is correlated with a reference system R in an arbitrary way, described by the overall state ρ^{AR}. In particular, the reference does not necessarily purify the source, nor is it assumed to be classical. The ensemble source and the pure source

defined by Schumacher are special cases of this model, where the reference is a classical system in the former and a purifying system in the latter. So is the source considered by Koashi and Imoto in [31], where the reference system is classical, too.

Understanding the compression of the source ρ^{AR} has paramount importance in the field of quantum information theory and unifies all the models that have been considered in the literature. Schumacher's pure source model in a sense is the most stringent model because it requires preserving the correlations with a purifying reference system which implies that the correlations with any other reference system is preserved which follows from the fact that the fidelity is non-decreasing under quantum channels. However, the converse is not necessarily true: if in a compression task the parties are required to preserve the correlations with a given reference system which does not purify the source state, they might be able to compress more efficiently compared to the scenario where the reference system purifies the source. This is exactly what we show in Chapter 3: we characterise the gap precisely depending on the reference system.

2.2 Mathematical definition of quantum noiseless compression

A source compression task consists of an encoder which maps the source to compressed information which is stored or sent to another party. When it is needed, a decoder maps the compressed information to decoded information, and the aim is to preserve the correlations with the reference system and reconstruct a source which is very close to the original source in some distance measure. In the quantum realm the most general encoding and decoding maps which can be performed on the information is a quantum operation or a CPTP map. The communication means or quantum storage device is assumed to be an ideal channel acting as an identity on the encoded information which can be simulated through various resources such as a qubit channel, sharing entanglement and sending classical information and etc. The resource is the dimension of the Hilbert space of the encoding operation.

Throughout the thesis, we consider the information theoretic asymptotic limit of n copies of a finite dimensional source with state ρ^{AR}, i.e. $\rho^{A^n R^n} = (\rho^{AR})^{\otimes n}$ where system A is the system to be compressed and system R is an inaccessible reference system. We assume that the encoder, Alice, and the decoder, Bob, share initially a maximally entangled state $\Phi_K^{A_0 B_0}$ on registers
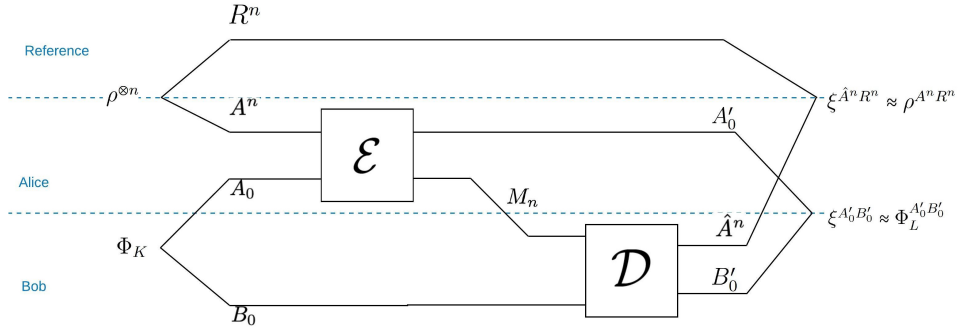
Figure 2.1: Circuit diagram of the compression task: the source is composed of n copies of the state ρ^{AR} where A^n is the system to be compressed and R^n is an inaccessible reference system. Dotted lines are used to demarcate domains controlled by the different participants here the reference, the encoder, Alice and the decoder, Bob. The solid lines represent quantum information registers. The encoder sends the compressed information, i.e. system M_n, to the decoder through a noiseless quantum channel; moreover, they share initial entanglement in the registers A_0 and B_0, respectively. The aim of the compression task is to reconstruct the source at the decoder side, that is the final state $\xi^{\hat{A}^n R^n}$ has the fidelity converging to 1 with the source state $\rho^{A^n R^n}$; this ensures that the correlations between the reconstructed system $\hat{A}^n$ and the reference system R^n are preserved. Furthermore, the encoder and decoder distill entanglement in their registers A_0' and B_0', respectively.

A_0 and B_0 (both of dimension K). The encoder, Alice, performs the encoding compression operation $\mathcal{E}: A^n A_0 \longrightarrow M_n A_0'$ on the system A^n and her part A_0 of the entanglement, which is CPTP map. Alice's encoding operation produces the state $\sigma^{M_n R^n A_0' B_0}$ with M, A_0' and B_0 as the compressed system of Alice, Alice's new entanglement system and Bob's part of the entanglement, respectively. The dimension of the compressed system is without loss of generality not larger than the dimension of the original source, i.e. $|M_n| \le |A|^n$. The system M_n is then sent to Bob via a noiseless quantum channel, who performs a decoding operation $\mathcal{D}: M_n B_0 \longrightarrow \hat{A}^n B_0'$ on the system M_n and his part of the entanglement B_0 where $\hat{A}^n$ and B_0' are the reconstructed source and Bob's new entanglement system. Ideally the encoder and decoder want to distill entanglement in the form of maximally entangled state $\Phi_L^{A_0' B_0'}$ of dimension L in their corresponding registers A_0' and B_0'.

We call $\frac{1}{n}\log(K-L)$ and $\frac{1}{n}\log|M_n|$ the *entanglement rate* and *quantum rate* of the compression protocol, respectively. We say the encoding-decoding

scheme has *fidelity* $1 - \epsilon$, or *error* ϵ, if

$$F\left(\rho^{A^n R^n}, \xi^{\hat{A}^n R^n}\right) \geq 1 - \epsilon, \tag{2.1}$$

where $\xi^{\hat{A}^n R^n} = \mathrm{Tr}_{A'_0 B'_0}\left[\left((\mathcal{D} \circ \mathcal{E}) \otimes \mathrm{id}_{R^n}\right)\rho^{A^n R^n} \otimes \Phi_K^{A_0 B_0}\right]$. Moreover, we say that (E, Q) is an (asymptotically) achievable rate pair if for all n there exist codes (encoders and decoders) such that the fidelity converges to 1, and the entanglement and quantum rates converge to E and Q, respectively. The compression schemes where the error converges to zero are called noiseless compression schemes which we consider throughout the thesis. The rate region is the set of all achievable rate pairs, as a subset of $\mathbb{R} \times \mathbb{R}_{\geq 0}$.

A schematic description of the quantum source and its compression is illustrated in Fig. 2.1 where the system to be compressed and the reference are denoted by A^n and R^n, respectively. This compression problems is addressed in Chapter 3 where we find the optimal trade-off rate region for the entanglement and quantum rates, that is the pairs (E, Q).

2.3 Quantum noiseless compression with side information

Side information in information theory is referred to as extra information, which is correlated with an information source and it is available to encoder, decoder or both of them, and they can use this extra information to use less resources, for example reduce the dimension of the compressed information. Slepian and Wolf for the first time studied the compression of a classical source, i.e a random variable, where a decoder has access to another random variable, which is correlated with the source, and showed that the compression rate is equal to the conditional Shannon entropy [43].

The *visible* paradigm of source compression problems are basically compression problems where an *encoder* has access to side information, i.e. the identity of states from an ensemble generated by a source [27, 60, 61, 64–67]. For example, the source in the visible Schumacher compression [60, 61] is modeled by a classical-quantum state $\rho^{AXX'} = \sum_x p(x)|\psi_x\rangle\langle\psi_x|^A \otimes |x\rangle\langle x|^X \otimes |x\rangle\langle x|^{X'}$, where system A is the system to be compressed, and systems X and X' are the side information system of the encoder and the reference system, respectively. It is shown that both visible model and *blind* model, where the encoder does not have access to system X, lead to the same compression rate, i.e. $S(A)$ [8, 60, 61] whereas this is not the case when system A is composed of mixed states, that is visible and blind models for mixed states lead to different

compression rates. In the visible mixed state compression problem, the source is modeled by many copies of the state $\rho^{AXX'} = \sum_x p(x) \rho_x^A \otimes |x\rangle\langle x|^X \otimes |x\rangle\langle x|^{X'}$ where system A with mixed states is the system to be compressed, and systems X and X' are the side information system of the encoder and the reference system, respectively. Hayashi showed that the optimal compression rate is equal to the regularized entanglement of purification of the source [27] which is different from the blind compression (X is not available to the encoder) rate obtained by Koashi and Imoto [31,42]. The visible compression of the this source when the encoder and decoder share unlimited entanglement is a special case of the remote state preparation considered in [66], and the optimal quantum compression rate is equal to $\frac{1}{2} S(A)$.

Winter in his Phd thesis [30] generalized the notion of correlated sources and side information at the *decoder* to a quantum setting by modeling it as a multipartite quantum source which generates multipartite quantum states where different parties have access to some parts of a source. The first example studied in this context was a hybrid classical-quantum source $\rho^{XARX'} = \sum_x p(x) |x\rangle\langle x|^X \otimes |\psi_x\rangle\langle\psi_x|^{AR} \otimes |x\rangle\langle x|^{X'}$ where an encoder compresses the classical system X, and a decoder aims to reconstruct this system while having access to quantum side information system A such that the correlations with the reference systems R and X' are preserved [29,30]. This example is one of the earliest attempts to find operational meaning to quantum conditional entropy in analogy to the classical conditional Shannon entropy which characterizes the optimal compression rate of a classical source with classical side information at the decoder side, a.k.a. fully classical Slepian-Wolf problem [43].

The compression of a purified source with side information at the decoder is known as state merging or fully quantum Slepian-Wolf (FQSW) and its discovery was an important mile stone in the quantum information field which gave an operational meaning to the quantum conditional entropy [23, 68]; in this task, a source generates many copies of the state $|\psi\rangle^{ABR}$ where an encoder compresses system A and sends it to a decoder who has access to system B and aims to reconstruct system A while preserving the correlations with the reference system R. Depending on the communication means which has been considered shared entanglement with free classical communication or quantum communication, the compression rate is equal to $S(A|B)$ ebits or $\frac{1}{2} I(A:R)$ qubits, respectively [23, 68]. An ensemble version of FQSW is considered in [32] with the source $\rho^{ABX} = \sum_x p(x) |\psi_x\rangle\langle\psi_x|^{AB} \otimes |x\rangle\langle x|^X$ and A, B and X as the system to be compressed, the side information at the decoder and the reference system, respectively; the optimal quantum compression rate is found for some special cases, but the problem has been left open in

18

general.

A generalization of state merging, which is known as quantum state redistribution (QSR), is proposed in [25, 26], where both encoder and decoder have access to side information systems. Namely, a source generates many copies of the state $|\psi\rangle^{ACBR}$, where an encoder compresses system A while having access to side information system C and sends the compressed information to a decoder who has access to system B and aims to reconstruct system A while preserving the correlations with the reference system R; in this compression task systems C and B remain at the disposal of the encoder and decoder, respectively. This gave an operational meaning to the quantum conditional mutual information since the optimal quantum compression rate was obtained to be $\frac{1}{2}I(A:R|C)$.

In the reminder of this section, we define mathematically the most general model for the compression of quantum sources with side information which includes as special cases all the aforementioned side information problems of this section (considering block fidelity defined in Eq. 2.2).

We consider a source generates asymptotic limit of n copies of a finite dimensional state ρ^{ACBR}, i.e. $\rho^{A^nC^nB^nR^n} = (\rho^{ACBR})^{\otimes n}$, and distributes the copies of the systems AC, B and R between an encoder, a decoder and an inaccessible reference system, respectively. We assume that the encoder, Alice, and the decoder, Bob, share initially a maximally entangled state $\Phi_K^{A_0B_0}$ on registers A_0 and B_0 (both of dimension K). The encoder, Alice, performs the encoding compression operation $\mathcal{E}: A^nC^nA_0 \longrightarrow M_n\hat{C}^nA_0'$ on the system A^nC^n and her part A_0 of the entanglement, which is CPTP map. Alice's encoding operation produces the state $\sigma^{M_n\hat{C}^nB^nR^nA_0'B_0}$ with M_n, $\hat{C}^n$, A_0' and B_0 as the compressed system of Alice, a reconstruction of system C^n, Alice's new entanglement system and Bob's part of the entanglement, respectively. The dimension of the compressed system is without loss of generality not larger than the dimension of the original source, i.e. $|M_n| \leq |A|^n$. The system M_n is then sent to Bob via a noiseless quantum channel, who performs a decoding operation $\mathcal{D}: M_nB^nB_0 \longrightarrow \hat{A}^n\hat{B}^nB_0'$ on the compressed information M_n, system B^n and his part of the entanglement B_0 where $\hat{A}^n$, $\hat{B}^n$ and B_0' are the reconstruction of systems A^n, B^n and Bob's new entanglement system, respectively. In this task, the side information systems remain at the disposal of their corresponding parties, that is the encoder and decoder respectively reconstruct systems C^n and B^n after using them as side information. Ideally the encoder and decoder want to distill entanglement in the form of maximally entangled state $\Phi_L^{A_0'B_0'}$ of dimension L in their corresponding registers A_0' and B_0'.

We call $\frac{1}{n}\log(K-L)$ and $\frac{1}{n}\log|M_n|$ the *entanglement rate* and *quantum*
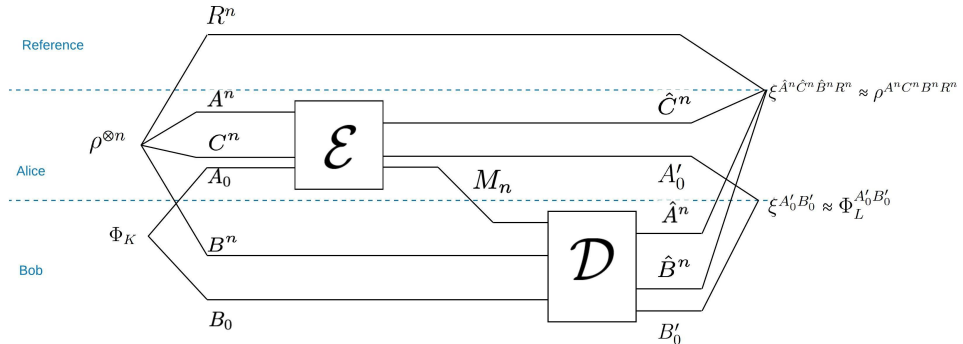
Figure 2.2: Circuit diagram of the compression task with side information: the source is composed of n copies of the state ρ^{ACBR} where A^n is the system to be compressed and R^n is an inaccessible reference system; systems C^n and B^n are the side information available for the encoder and the decoder, respectively. Dotted lines are used to demarcate domains controlled by the different participants here the reference, the encoder, Alice and the decoder, Bob. The solid lines represent quantum information registers. The encoder sends the compressed information, i.e. system M_n, to the decoder through a noiseless quantum channel; moreover, they share initial entanglement in the registers A_0 and B_0, respectively. The aim of the compression task is to reconstruct system A^n at the decoder side while each party reconstructs its own corresponding side information as well, that is the final state $\xi^{\hat{A}^n \hat{C}^n \hat{B}^n R^n}$ has the fidelity converging to 1 with the source state $\rho^{A^n C^n B^n R^n}$; this ensures that the correlations between the reconstructed systems $\hat{A}^n \hat{C}^n \hat{B}^n$ and the reference system R^n are preserved. Furthermore, the encoder and decoder distill entanglement in their registers A_0' and B_0', respectively.

rate of the compression protocol, respectively. We say the encoding-decoding scheme has *block fidelity* $1 - \epsilon$, or *block error* ϵ, if

$$F\left(\rho^{A^n C^n B^n R^n}, \xi^{\hat{A}^n \hat{C}^n \hat{B}^n R^n}\right) \geq 1 - \epsilon, \qquad (2.2)$$

where $\xi^{\hat{A}^n \hat{C}^n \hat{B}^n R^n} = \text{Tr}_{A_0' B_0'}\left[\left((\mathcal{D} \circ \mathcal{E}) \otimes \text{id}_{R^n}\right) \rho^{A^n C^n B^n R^n} \otimes \Phi_K^{A_0 B_0}\right]$. Moreover, we say that (E_b, Q_b) is an (asymptotically) achievable block-error rate pair if for all n there exist codes (encoders and decoders) such that the block fidelity converges to 1, and the entanglement and quantum rates converge to E_b and Q_b, respectively. The rate region is the set of all achievable rate pairs, as a subset of $\mathbb{R} \times \mathbb{R}_{\geq 0}$. A schematic description of the source compression task with side information is illustrated in Fig. 2.2.

We also consider another figure of merit which turns out to be an easier

criterion to evaluate side information problems; we say a code has *per-copy fidelity* $1 - \epsilon$, or *per-copy error* ϵ, if

$$\frac{1}{n} \sum_{j=1}^{n} F(\rho^{ACBR}, \xi^{\hat{A}_j \hat{C}_j \hat{B}_j R_j}) \geq 1 - \epsilon, \qquad (2.3)$$

where $\xi^{\hat{A}_j \hat{C}_j \hat{B}_j R_j} = \text{Tr}_{[n] \setminus j} \xi^{\hat{A}^n \hat{C}^n \hat{B}^n R^n}$, and 'Tr$_{[n] \setminus j}$' denotes the partial trace over all systems with indices in $[n] \setminus j$. Similarly, we say that (E_c, Q_c) is an (asymptotically) achievable per-copy-error rate pair if for all n there exist codes (encoders and decoders) such that the per-copy fidelity converges to 1, and the entanglement and quantum rates converge to E_c and Q_c, respectively. The rate region is the set of all achievable rate pairs, as a subset of $\mathbb{R} \times \mathbb{R}_{\geq 0}$.

This general compression problem has a complex nature; for example, consider the special case of the visible mixed state source by Hayashi [27] with classical reference $R = X'$ and classical side information at the encoder $C = X$, with no side information at the decoder $B = \varnothing$, i.e. $\rho^{AXX'} = \sum_x p(x) \rho_x^A \otimes |x\rangle\langle x|^X \otimes |x\rangle\langle x|^{X'}$; with no entanglement consumption, the optimal block-error quantum rate, i.e. the pair $(0, Q_b^*)$ is equal to the regularized entanglement of purification whereas with free entanglement the optimal block-error quantum rate, i.e. the pair (∞, Q_b^*) is equal to $(\infty, \frac{1}{2} S(A))$ [66]. Therefore, it is insightful to first study the pairs $(0, Q_b^*)$ and (∞, Q_b^*) for some other special cases of the source ρ^{ACBR}.

Moreover as we will show in the subsequent chapters, unlike the classical scenario where conditional entropy characterizes the classical compression rate, for non-pure sources, the quantum conditional entropy or mutual information does not necessary play a role and more complicates functions of the source determine the compression rate. In section 2.5, we briefly go through the special cases of the general source ρ^{ACBR} with side information which we address in this thesis and discuss the challenges of each particular case in its corresponding chapter.

2.4 Distributed noiseless quantum source compression

This thesis mainly focuses on the side information compression problems, however, in chapter 5, aside from a side information problem we study the distributed compression of correlated classical-quantum sources. This motivates us to define a general distributed compression problem, which the side information problem of section 2.3 can be considered a sub-problem of this

distributed scenario since the decoder can use successive decoding, that is it can first decode the information of one of the encoders and treat it as its own side information, and later decode the information of the other encoder.

Here we define the problem for two encoders, however, the definition can be easily extended to three or more encoders. We consider a source generates asymptotic limit of n copies of a finite dimensional state $\rho^{A_1 C_1 A_2 C_2 BR}$, i.e. $\rho^{A_1^n C_1^n A_2^n C_2^n B^n R^n} = \left(\rho^{A_1 C_1 A_2 C_2 BR}\right)^{\otimes n}$, and distributes the copies of the systems $A_1 C_1$, $A_2 C_2$, B and R between encoder 1, encoder 2, a decoder and an inaccessible reference system, respectively. We assume that both encoder 1, Alice and encoder 2, Ava, share initially maximally entangled states $\Phi_{K_1}^{A_{01} B_{01}}$ and $\Phi_{K_2}^{A_{02} B_{02}}$ with the decoder, Bob, respectively (of dimension K_1 and K_2 respectively). The encoder i ($i = 1, 2$) performs the encoding compression operation, i.e. the CPTP map $\mathcal{E}_i : A_i^n C_i^n A_{0i} \longrightarrow M_{i_n} \hat{C}_i^n A_{0i}'$ on the systems $A_i^n C_i^n$ and the entanglement part A_{0i}. The encoding operations are distributed in the sense that each encoder applies her own operation locally without having access to the information of the other encoder. The dimension of the compressed systems are without loss of generality not larger than the dimension of the original sources, i.e. $|M_{i_n}| \leq |A_i|^n$. The systems M_{i_n} ($i = 1, 2$) are then sent to Bob via a noiseless quantum channel, who performs the decoding operation $\mathcal{D} : M_{1_n} M_{2_n} B^n B_{01} B_{02} \longrightarrow \hat{A}_1^{\,n} \hat{A}_2^{\,n} \hat{B}^n B_{01}' B_{02}'$ on the compressed information systems $M_{1_n} M_{2_n}$, system B^n and his parts of the entanglement $B_{01} B_{02}$ where $\hat{A}_1^{\,n} \hat{A}_2^{\,n}$, $\hat{B}^n$ and $B_{01}' B_{02}'$ are the reconstruction of systems $A_1^n A_2^n$, B^n and Bob's new entanglement systems, respectively. In this task, the systems C_1^n, C_2^n and B^n remain at the disposal of their corresponding parties, that is the encoders and the decoder respectively reconstruct systems C_1^n, C_2^n and B^n after using them as side information. Ideally the encoder i ($i = 1, 2$) and the decoder aim to distill entanglement in the form of maximally entangled state $\Phi_{L_i}^{A_{0i}' B_{0i}'}$ of dimension L_i in their corresponding registers A_{0i}' and B_{0i}', respectively.

We call $\frac{1}{n} \log(K_i - L_i)$ and $\frac{1}{n} \log |M_{i_n}|$ the *entanglement rate* and *quantum rate* of the compression protocol, respectively (for $i = 1, 2$). Moreover, we say the encoding-decoding scheme has *block fidelity* $1 - \epsilon$, or *block error* ϵ, if

$$F\left(\rho^{A_1^n C_1^n A_2^n C_2^n B^n R^n}, \xi^{\hat{A}_1^{\,n} \hat{C}_1^{\,n} \hat{A}_2^{\,n} \hat{C}_2^{\,n} \hat{B}^n R^n}\right) \geq 1 - \epsilon, \qquad (2.4)$$

where

$$\xi^{\hat{A}_1^{\,n} \hat{C}_1^{\,n} \hat{A}_2^{\,n} \hat{C}_2^{\,n} \hat{B}^n R^n} =$$
$$\mathrm{Tr}_{A_{01}' B_{01}' A_{02}' B_{02}'} \left[\left((\mathcal{D} \circ (\mathcal{E}_1 \otimes \mathcal{E}_2)) \otimes \mathrm{id}_{R^n} \right) \rho^{A_1^n C_1^n A_2^n C_2^n B^n R^n} \otimes \Phi_{K_1}^{A_{01} B_{01}} \otimes \Phi_{K_2}^{A_{02} B_{02}} \right].$$
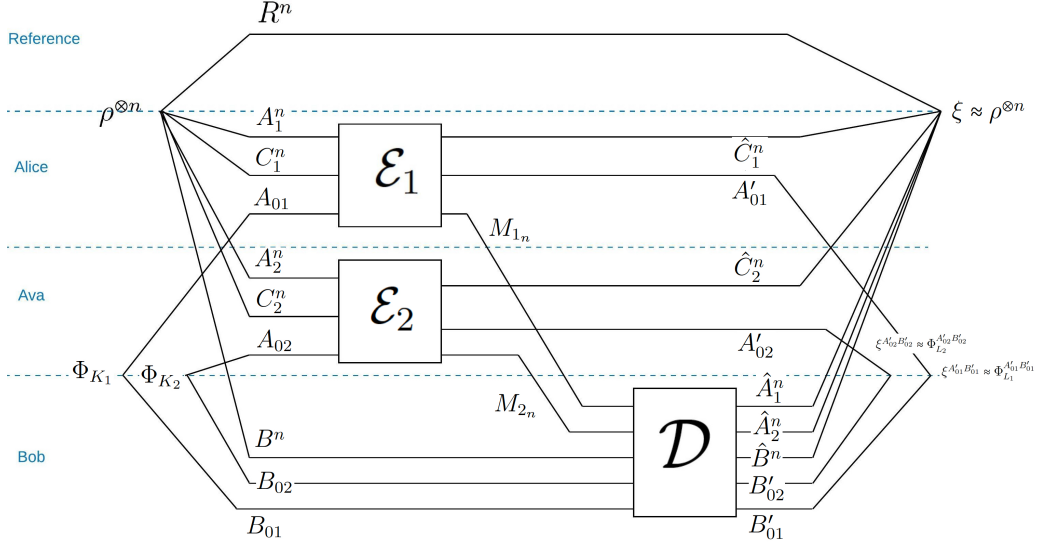
Figure 2.3: Circuit diagram of the distributed compression task: the source is composed of n copies of the state $\rho^{A_1 C_1 A_2 C_2 BR}$ where A_i^n $(i = 1, 2)$ is the system to be compressed and R^n is an inaccessible reference system; systems C_i^n and B^n are the side information available for the encoder i and the decoder, respectively. Dotted lines are used to demarcate domains controlled by the different participants here the reference, the encoders, Alice and Ava, and the decoder, Bob. The solid lines represent quantum information registers. The encoder i sends the compressed information, i.e. system M_{i_n}, to the decoder through a noiseless quantum channel; moreover, they share initial entanglement in the registers A_{0i} and B_{0i}, respectively. The aim of the compression task is to reconstruct systems A_1^n and A_2^n at the decoder side while each party reconstructs its own corresponding side information as well, that is the final state $\xi^{\hat{A_1}^n \hat{C_1}^n \hat{A_2}^n \hat{C_2}^n \hat{B}^n R^n}$ has the fidelity converging to 1 with the source state $\rho^{A_1^n C_1^n A_2^n C_2^n B^n R^n}$; this ensures that the correlations between the reconstructed systems $\hat{A_1}^n \hat{C_1}^n \hat{A_2}^n \hat{C_2}^n \hat{B}^n$ and the reference system R^n are preserved. Furthermore, the encoder i and the decoder distill entanglement in their registers A_{0i}' and B_{0i}', respectively.

Moreover, we say that $(E_{b_1}, E_{b_2}, Q_{b_1}, Q_{b_2})$ is an (asymptotically) achievable block-error rate tuple if for all n there exist codes (encoders and decoders) such that the block fidelity converges to 1, and the ith entanglement and quantum rates converge to E_{b_i} and Q_{b_i} for encoder i, respectively. The rate region is the set of all achievable rate pairs, as a subset of $\mathbb{R} \times \mathbb{R} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$. A schematic description of the source compression task with side information

is illustrated in Fig. 2.3.

In chapter 5, we consider block fidelity, however, the results follow for the per-copy fidelity as well which is defined as follows: we say a code has *per-copy fidelity* $1 - \epsilon$, or *per-copy error* ϵ, if

$$\frac{1}{n} \sum_{j=1}^{n} F\left(\rho^{A_1 C_1 A_2 C_2 BR}, \xi^{\hat{A_{1j}} \hat{C_{1j}} \hat{A_{2j}} \hat{C_{2j}} \hat{B} R}\right) \geq 1 - \epsilon, \tag{2.5}$$

where $\xi^{\hat{A_{1j}} \hat{C_{1j}} \hat{A_{2j}} \hat{C_{2j}} \hat{B} R} = \mathrm{Tr}_{[n] \smallsetminus j} \, \xi^{\hat{A_1}^n \hat{C_1}^n \hat{A_2}^n \hat{C_2}^n \hat{B}^n R^n}$, and '$\mathrm{Tr}_{[n] \smallsetminus j}$' denotes the partial trace over all systems with indices in $[n] \smallsetminus j$. Similarly, we say that $(E_{c_1}, E_{c_2}, Q_{c_1}, Q_{c_2})$ is an (asymptotically) achievable per-copy-error rate tuple if for all n there exist codes such that the per-copy fidelity converges to 1, and the ith entanglement and quantum rates converge to E_{c_i} and Q_{c_i} for encoder i, respectively. The rate region is the set of all achievable rate pairs, as a subset of $\mathbb{R} \times \mathbb{R} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$.

In [69, 70], compression of a pure source $|\psi\rangle^{A_1 A_2 R}$ with side information at the encoders is considered $(C_1, C_2, B = \varnothing)$. The achievable rate region is a convex hull of various points where each point corresponding to an encoder is achieved by applying fully quantum Slepian-Wolf (FQSW) compression and treating the rest of the systems as a reference. The converse bounds are in terms of the multipartite squashed entanglement, which is a measure of multipartite entanglement.

2.5 Summary of our results in quantum source compression and discussion

In this section, we briefly explain the special cases of problems, defined in the previous sections, that we address in this thesis. Notice that in the subsequent chapters we do not necessarily respect the notation A, C, B and R for the system to be compressed, the side information at the encoder, the side information at the decoder and the reference system, however, we clearly define the task and specify the notation for the corresponding registers. Moreover, we specify whether the error criterion is block fidelity or per-copy fidelity.

In chapter 3, we consider the compression of a general mixed state source ρ^{AR} (no side information) and find the optimal trade-off between the entanglement and quantum rates, i.e. the pair (E,Q).

In chapter 4, we unify the visible and blind Schumacher compression by considering an interpolation between them as side information, that is the

source $\rho^{ACX} = \sum_x p(x)|\psi_x\rangle\langle\psi_x|^A \otimes |c_x\rangle\langle c_x|^C \otimes |x\rangle\langle x|^X$ with A, C and X as the system to be compressed, the side information at the encoder and the classical reference system. For this source, we find optimal trade-off between the block-error entanglement and quantum rate pairs (E_b, Q_b).

In chapter 5, we consider quantum source compression with classical side information with the source $\rho^{ARXX'} = \sum_x p(x)|\psi_x\rangle\langle\psi_x|^{AR} \otimes |x\rangle\langle x|^X \otimes |x\rangle\langle x|^{X'}$ and A, X and RX' as the system to be compressed, the side information at the decoder and the hybrid classical-quantum reference systems, respectively. We study the entanglement assisted case (∞, Q_b), the unassisted case $(0, Q_b)$ then distributed scenario considering block fidelity. We find achievable and converse bounds for each scenario and show that the two bounds match for the entanglement assisted quantum block-error rate Q_b up to continuity of a function which appears in the bounds. Finally, considering per-copy fidelity we find the optimal entanglement assisted quantum per-copy-error rate, i.e. the pair (∞, Q_c^*).

In chapter 6, we consider an ensemble generalization of the quantum state redistribution (QSR), i.e. the source $\rho^{ACBR} = \sum_x p(x)|\psi_x\rangle\langle\psi_x|^{ACBR'} \otimes |x\rangle\langle x|^{X'}$ with A, C, B and $R = R'X'$ as the system to be compressed, the side information at the encoder, the side information at the decoder and the hybrid classical-quantum reference systems, respectively. We consider free entanglement scenario and find the optimal quantum per-copy-error rate, i.e. the pair (∞, Q_c^*). With block fidelity, we find achievable and converse bounds which match up to continuity of a function appearing in the bounds.

In summary, for a general mixed state we solve the problem when there is no side information, and the rate region is in terms of an extension of the decomposition of the source state which is discovered by Koashi and Imoto in [42], and later this decomposition extended to a general mixed state in [41]. However, for multipartite states this decomposition does not necessarily preserve the tensor structure over various systems; this turns out to be the main hurdle in dealing with general mixed state problems with side information. This is not an issue for pure or ensemble sources mainly because the structure of maps which preserve these states are well-understood. For these sources the environment systems of the encoding and decoding operations are decoupled from the reconstructed source given the identity of the state from the ensemble. This property is one of the guiding intuitions behind the converse proofs for the side information problems.

Chapter 3

Compression of a general mixed state source

In this chapter, we consider the most general (finite-dimensional) quantum mechanical information source, which is given by a quantum system A that is correlated with a reference system R. The task is to compress A in such a way as to reproduce the joint source state ρ^{AR} at the decoder with asymptotically high fidelity. This includes Schumacher's original quantum source coding problem of a pure state ensemble and that of a single pure entangled state, as well as general mixed state ensembles. Here, we determine the optimal compression rate (in qubits per source system) in terms of the Koashi-Imoto decomposition of the source into a classical, a quantum, and a redundant part. The same decomposition yields the optimal rate in the presence of unlimited entanglement between compressor and decoder, and indeed the full region of feasible qubit-ebit rate pairs. This chapter is based on the papers in [45, 46].

3.1 The source model and the compression task

We consider a general mixed state source ρ^{AR} with A and R as the system to be compressed and the reference system, respectively, where the source generates the information theoretic limit of many copies of the state ρ^{AR}, i.e. $\rho^{A^n R^n} = \left(\rho^{AR}\right)^{\otimes n}$. We assume that the encoder, Alice, and the decoder, Bob, have initially a maximally entangled state $\Phi_K^{A_0 B_0}$ on registers A_0 and B_0 (both of dimension K). The encoder, Alice, performs the encoding compression operation $\mathcal{C} : A^n A_0 \longrightarrow M$ on the system A^n and her part A_0 of the

entanglement, which is a quantum channel, i.e. a completely positive and trace preserving (CPTP) map. Notice that as functions CPTP maps act on the operators (density matrices) over the respective input and output Hilbert spaces, but as there is no risk of confusion, we will simply write the Hilbert spaces when denoting a CPTP map. Alice's encoding operation produces the state $\sigma^{MB_0R^n}$ with M and B_0 as the compressed system of Alice and Bob's part of the entanglement, respectively. The dimension of the compressed system is without loss of generality not larger than the dimension of the original source, i.e. $|M| \leq |A|^n$. We call $\frac{1}{n}\log K$ and $\frac{1}{n}\log|M|$ the *entanglement rate* and *quantum rate* of the compression protocol, respectively. The system M is then sent to Bob via a noiseless quantum channel, who performs a decoding operation $\mathcal{D}: MB_0 \longrightarrow \hat{A}^n$ on the system M and his part of the entanglement B_0. We say the encoding-decoding scheme has *fidelity* $1 - \epsilon$, or *error* ϵ, if

$$F\left(\rho^{A^nR^n}, \xi^{\hat{A}^nR^n}\right) \geq 1 - \epsilon, \tag{3.1}$$

where $\xi^{\hat{A}^nR^n} = ((\mathcal{D} \circ \mathcal{C}) \otimes \mathrm{id}_{R^n})\rho^{A^nR^n}$. Moreover, we say that (E, Q) is an (asymptotically) achievable rate pair if for all n there exist codes such that the fidelity converges to 1, and the entanglement and quantum rates converge to E and Q, respectively. The rate region is the set of all achievable rate pairs, as a subset of $\mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$.

According to Stinespring's theorem [55], a CPTP map $\mathcal{T}: A \longrightarrow \hat{A}$ can be dilated to an isometry $U: A \hookrightarrow \hat{A}E$ with E as an environment system, called an isometric extension of a CPTP map, such that $\mathcal{T}(\rho^A) = \mathrm{Tr}_E(U\rho^A U^\dagger)$. Therefore, the encoding and decoding operations are can in general be viewed as isometries $U_\mathcal{E}: A^nA_0 \hookrightarrow MW$ and $U_\mathcal{D}: MB_0 \hookrightarrow \hat{A}^nV$, respectively, with the systems W and V as the environment systems of Alice and Bob, respectively.

We say a source ω^{BR} is equivalent to a source ρ^{AR} if there are CPTP maps $\mathcal{T}: A \longrightarrow B$ and $\mathcal{R}: B \longrightarrow A$ in both directions taking one to the other:

$$\omega^{BR} = (\mathcal{T} \otimes \mathrm{id}_R)\rho^{AR} \text{ and } \rho^{AR} = (\mathcal{R} \otimes \mathrm{id}_R)\omega^{BR}. \tag{3.2}$$

The rate regions of equivalent sources are the same, because any achievable rate pair for one source is achievable for the other source as well. This follows from the fact that for any code $(\mathcal{C}, \mathcal{D})$ of block length n and error ϵ for ρ^{AR}, concatenating the encoding and decoding operations with $\mathcal{T}$ and $\mathcal{R}$, i.e. letting $\mathcal{C}' = \mathcal{C} \circ \mathcal{R}^{\otimes n}$ and $\mathcal{D}' = \mathcal{T}^{\otimes n} \circ \mathcal{D}$, we get a code of the same error ϵ for ω^{BR}. Analogously we can turn a code for ω^{BR} into one for ρ^{AR}.

28

3.2 The qubit-ebit rate region

The idea behind the compression of the source ρ^{AR} is based on a decomposition of this state introduced in [41], which is a generalization of the decomposition introduced by Koashi and Imoto in [42]. Namely, for any set of quantum states $\{\rho_x\}$, there is a unique decomposition of the Hilbert space describing the structure of CPTP maps which preserve the set $\{\rho_x^A\}$. This idea was generalized in [41] for a general mixed state ρ^{AR} describing the structure of CPTP maps acting on system A which preserve the overall state ρ^{AR}. This was achieved by showing that any such map preserves the set of all possible states on system A which can be obtained by measuring system R, and conversely any map preserving the set of all possible states on system A obtained by measuring system R, preserves the state ρ^{AR}, thus reducing the general case to the case of classical-quantum states

$$\rho^{AY} = \sum_y q(y)\rho_y^A \otimes |y\rangle\langle y|^Y = \sum_y \operatorname{Tr}_R \rho^{AR}(\mathbb{1}_A \otimes M_y^R) \otimes |y\rangle\langle y|^Y,$$

which is the ensemble case considered by Koashi and Imoto. As a matter of fact, looking at the algorithm presented in [42] to compute the decomposition, it is enough to consider an informationally complete POVM (M_y) on R, with no more than $|R|^2$ many outcomes. The properties of this decomposition are stated in the following theorem.

Theorem 3.1 ([41,42])**.** *Associated to the state ρ^{AR}, there are Hilbert spaces C, N and Q and an isometry $U_{KI}: A \hookrightarrow CNQ$ such that:*

1. *The state ρ^{AR} is transformed by U_{KI} as*

$$(U_{KI} \otimes \mathbb{1}_R)\rho^{AR}(U_{KI}^\dagger \otimes \mathbb{1}_R) = \sum_j p_j |j\rangle\langle j|^C \otimes \omega_j^N \otimes \rho_j^{QR} =: \omega^{CNQR}, \quad (3.3)$$

   where the set of vectors $\{|j\rangle^C\}$ form an orthonormal basis for Hilbert space C, and p_j is a probability distribution over j. The states ω_j^N and ρ_j^{QR} act on the Hilbert spaces N and $Q \otimes R$, respectively.

2. *For any CPTP map Λ acting on system A which leaves the state ρ^{AR} invariant, that is $(\Lambda \otimes \operatorname{id}_R)\rho^{AR} = \rho^{AR}$, every associated isometric extension $U : A \hookrightarrow AE$ of Λ with the environment system E is of the following form*

$$U = (U_{KI} \otimes \mathbb{1}_E)^\dagger \left(\sum_j |j\rangle\langle j|^C \otimes U_j^N \otimes \mathbb{1}_j^Q \right) U_{KI}, \quad (3.4)$$

where the isometries $U_j : N \hookrightarrow NE$ satisfy $\mathrm{Tr}\,_E[U_j \omega_j U_j^\dagger] = \omega_j$ for all j. The isometry U_{KI} is unique (up to trivial change of basis of the Hilbert spaces C, N and Q). Henceforth, we call the isometry U_{KI} and the state $\omega^{CNQR} = \sum_j p_j |j\rangle\langle j|^C \otimes \omega_j^N \otimes \rho_j^{QR}$ the Koashi-Imoto (KI) isometry and KI-decomposition of the state ρ^{AR}, respectively.

3. *In the particular case of a tripartite system CNQ and a state ω^{CNQR} already in Koashi-Imoto form (3.3), property 2 says the following: For any CPTP map Λ acting on systems CNQ with $(\Lambda \otimes \mathrm{id}_R)\omega^{CNQR} = \omega^{CNQR}$, every associated isometric extension $U : CNQ \hookrightarrow CNQE$ of Λ with the environment system E is of the form*

$$U = \sum_j |j\rangle\langle j|^C \otimes U_j^N \otimes \mathbb{1}_j^Q, \qquad (3.5)$$

where the isometries $U_j : N \hookrightarrow NE$ satisfy $\mathrm{Tr}\,_E[U_j \omega_j U_j^\dagger] = \omega_j$ for all j.

According to the discussion at the end of Sec. 3.1, the sources ρ^{AR} and ω^{CNQR} are equivalent because there are the isometry U_{KI} and the reversal CPTP map $\mathcal{R} : CNQ \longrightarrow A$, which reverses the action of the KI isometry, such that:

$$
\begin{aligned}
\omega^{CNQR} &= (U_{\mathrm{KI}} \otimes \mathbb{1}_R)\rho^{AR}(U_{\mathrm{KI}}^\dagger \otimes \mathbb{1}_R), \\
\rho^{AR} &= (\mathcal{R} \otimes \mathrm{id}_R)\omega^{CNQR} \\
&= (U_{\mathrm{KI}}^\dagger \otimes \mathbb{1}_R)\omega^{CNQR}(U_{\mathrm{KI}} \otimes \mathbb{1}_R) + \mathrm{Tr}\left[(\mathbb{1}_{CNQ} - \Pi_{CNQ})\omega^{CNQ}\right]\sigma, \quad (3.6)
\end{aligned}
$$

where $\Pi_{CNQ} = U_{\mathrm{KI}}U_{\mathrm{KI}}^\dagger$ is the projection onto the subspace $U_{\mathrm{KI}}A \subset C \otimes N \otimes Q$, and σ is an arbitrary state acting on $A \otimes R$. Henceforth we assume that the source is ω^{CNQR}, which is convenient because our main result is expressed in terms of the systems C and Q. Notice that the source ω^{CNQR} is in turn equivalent to ω^{CQR}, a fact we will exploit in the proof.

Moreover, since the information in C is classical, we can reduce the compression rate even more if the sender and receiver share entanglement, by using dense coding of j. In the following theorem we show the optimal qubit-ebit rate tradeoff for the compression of the source ρ^{AR}.

Theorem 3.2. *For the compression of the source ρ^{AR}, all asymptotically achievable entanglement and quantum rate pairs (E, Q) satisfy*

$$
\begin{aligned}
Q &\geq S(CQ)_\omega - \frac{1}{2}S(C)_\omega, \\
Q + E &\geq S(CQ)_\omega,
\end{aligned}
$$

where the entropies are with respect the KI decomposition of the state ρ^{AR}, i.e. the state ω^{CNQR}. Conversely, all the rate pairs satisfying the above inequalities are asymptotically achievable.

Remark 3.1. *This theorem implies that the optimal asymptotic quantum rates for the compression of the source ρ^{AR} with and without entanglement assistance are $S(CQ)_\omega - \frac{1}{2}S(C)_\omega$ and $S(CQ)_\omega$ qubits, respectively, and $\frac{1}{2}S(C)_\omega$ ebits of entanglement are sufficient and necessary in the entanglement assisted case.*

Remark 3.2. *If in the compression task the parties were required to preserve the correlations with a purifying reference system, then due to Schumacher compression the optimal qubit rate would be $S(A)_\rho = S(CNQ)_\omega$. However, Theorem 3.2 shows that the parties can compress more if they are only required to preserve the correlations with a mixed state reference. This gap can be strictly positive if the redundant system N is mixed given the classical information j in system C, that is $S(CNQ)_\omega - S(CQ)_\omega = S(N|CQ)_\omega > 0$.*
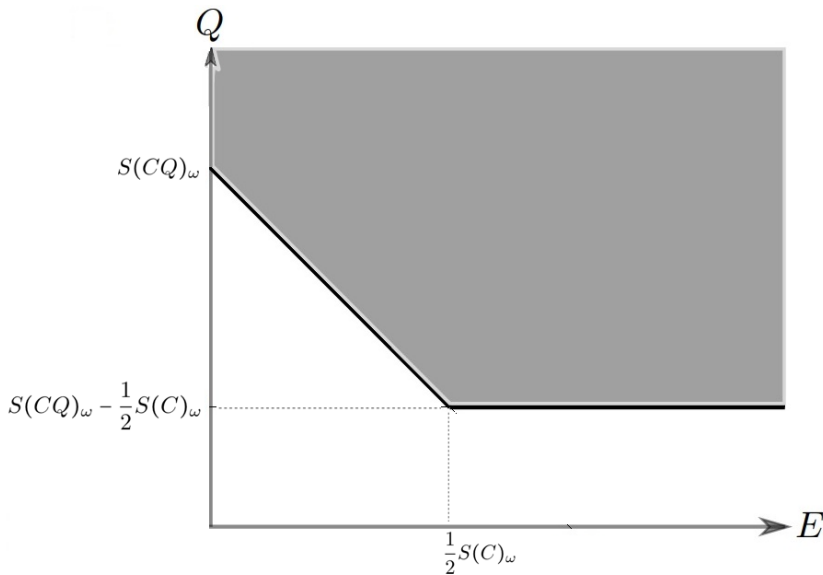
Figure 3.1: The achievable rate region of the entanglement and quantum rates.

Proof. We start with the achievability of these rates. The converse proofs need more tools, so we will leave them to the subsequent sections. Looking at Fig. 3.1, it will be enough to prove the achievability of the corresponding corner points $(E, Q) = (0, S(CQ)_\omega)$ and $(E, Q) = (\frac{1}{2}S(C)_\omega, S(CQ)_\omega - \frac{1}{2}S(C)_\omega)$

31

for the unassisted and entanglement assisted cases, respectively. This is because by definition (and the time-sharing principle) the rate region is convex and upper-right closed. Indeed, all the points on the line $Q + E = S(CQ)_\omega$ for $Q \geq S(CQ)_\omega - \frac{1}{2}S(C)_\omega$ are achievable because one ebit can be distributed by sending a qubit. All other rate pairs are achievable by resource wasting. The rate region is depicted in Fig. 3.1.

As we discussed, we can assume that the source is $(\omega^{CNQR})^{\otimes n} = \omega^{C^n N^n Q^n R^n}$. To achieve the point $(0, S(CQ)_\omega)$, Alice traces out the redundant part N^n of the source, to get the state $\omega^{C^n Q^n R^n}$ and applies Schumacher compression to send the systems $C^n Q^n$ to Bob. Since the Schumacher compression preserves the purification of the systems $C^n Q^n$, it preserves the state $\omega^{C^n Q^n R^n}$ as well. To be more specific, let Λ_S denote the composition of the encoding and decoding operations for the Schumacher compression of the state $|\omega\rangle^{C^n Q^n R^n R'^n}$ where the system R'^n is a purifying reference system which of course the parties do not have access to. The Schumacher compression preserves the following fidelity on the left member of the equation, therefore it preserves the fidelity on the right member:

$$
\begin{aligned}
1 - \epsilon &\leq F\left(\omega^{C^n Q^n R^n R'^n}, (\Lambda_S \otimes \mathrm{id}_{R^n R'^n})\omega^{C^n Q^n R^n R'^n}\right) \\
&\leq F\left(\omega^{C^n Q^n R^n}, (\Lambda_S \otimes \mathrm{id}_{R^n})\omega^{C^n Q^n R^n}\right),
\end{aligned}
$$

where the inequality is due to monotonicity of the fidelity under partial trace. The rate achieved by this scheme is $S(CQ)_\omega$. After applying this scheme, Bob has access to the systems $\hat{C}^n \hat{Q}^n$, which is correlated with the reference system R^n:

$$
\zeta^{\hat{C}^n \hat{Q}^n R^n} = (\Lambda_S \otimes \mathrm{id}_{R^n})\omega^{C^n Q^n R^n}.
$$

Then, to reconstruct the system N^n, Bob applies the CPTP map $\mathcal{N} : CQ \longrightarrow CNQ$ to each copy, which acts as follows:

$$
\mathcal{N}(\rho^{CQ}) = \sum_j (|j\rangle\langle j|^C \otimes \mathbb{1}_Q)\rho^{CQ}(|j\rangle\langle j|^C \otimes \mathbb{1}_Q) \otimes \omega_j^N.
$$

This map satisfies the fidelity criterion of Eq. (3.7) because of monotonicity of the fidelity under CPTP maps:

$$
\begin{aligned}
1 - \epsilon &\leq F\left(\omega^{C^n Q^n R^n}, \zeta^{\hat{C}^n \hat{Q}^n R^n}\right) \\
&\leq F\left((\mathcal{N}^{\otimes n} \otimes \mathrm{id}_{R^n})\omega^{C^n Q^n R^n}, (\mathcal{N}^{\otimes n} \otimes \mathrm{id}_{R^n})\zeta^{\hat{C}^n \hat{Q}^n R^n}\right) \\
&= F\left(\omega^{C^n N^n Q^n R^n}, \tau^{\hat{C}^n \hat{N}^n \hat{Q}^n R^n}\right).
\end{aligned} \tag{3.7}
$$

To achieve the point $(\frac{1}{2}S(C)_\omega, S(CQ)_\omega - \frac{1}{2}S(C)_\omega)$, Alice applies dense coding to send the classical system C^n to Bob which requires $\frac{n}{2}S(C)_\omega$ ebits of initial entanglement and $\frac{n}{2}S(C)_\omega$ qubits [71]. When both Alice and Bob have access to system C^n, Alice can send the quantum system Q^n to Bob by applying Schumacher compression, which requires sending $nS(Q|C)$ qubits to Bob. Therefore, the overall qubit rate is $\frac{1}{2}S(C)_\omega + S(Q|C) = S(CQ)_\omega - \frac{1}{2}S(C)_\omega$. ∎

3.3 Converse

In this section, we will provide the converse bounds for the qubit rate Q and the sum rate $Q + E$ of Theorem 3.2. We obtain these bounds based on the structure of the CPTP maps which preserve the source state ω^{CNQR}. Namely, according to Theorem 3.1 the CPTP maps acting on systems CNQ, which preserve the state ω^{CNQR}, act only on the redundant system N. This implies that the environment systems of such CPTP maps are decoupled from systems QR given the classical information j in the classical system C. This gives us an insight into the structure of the encoding-decoding maps, which preserve the overall state *asymptotically* intact.

To proceed with the proof, we first define two functions that emerge in the converse bounds. Then, we state some important properties of these functions in Lemma 3.1 which we will use to compute the tight asymptotic converse bounds.

Definition 3.1. *For the KI decomposition $\omega^{CNQR} = \sum_j p_j |j\rangle\langle j|^C \otimes \omega_j^N \otimes \rho_j^{QR}$ of the state ρ^{AR} and $\epsilon \geq 0$, define*

$$J_\epsilon(\omega) \coloneqq \max I(\hat{N}E : \hat{C}\hat{Q}|C')_\tau \ s.t.$$

$$\qquad U : CNQ \to \hat{C}\hat{N}\hat{Q}E \ is \ an \ isometry \ with \ F(\omega^{CNQR}, \tau^{\hat{C}\hat{N}\hat{Q}R}) \geq 1 - \epsilon,$$

$$Z_\epsilon(\omega) \coloneqq \max S(\hat{N}E|C')_\tau \ s.t.$$

$$\qquad U : CNQ \to \hat{C}\hat{N}\hat{Q}E \ is \ an \ isometry \ with \ F(\omega^{CNQR}, \tau^{\hat{C}\hat{N}\hat{Q}R}) \geq 1 - \epsilon,$$

where

$$\omega^{CNQRC'} = \sum_j p_j |j\rangle\langle j|^C \otimes \omega_j^N \otimes \rho_j^{QR} \otimes |j\rangle\langle j|^{C'},$$

$$\tau^{\hat{C}\hat{N}\hat{Q}ERC'} = (U \otimes \mathbb{1}_{RC'})\omega^{CNQRC'}(U^\dagger \otimes \mathbb{1}_{RC'}),$$

$$\tau^{\hat{C}\hat{N}\hat{Q}R} = \mathrm{Tr}_{EC'}[\tau^{\hat{C}\hat{N}\hat{Q}ERC'}].$$

In this definition, the dimension of the environment is w.l.o.g. bounded as $|E| \leq (|C||N||Q|)^2$ because the input and output dimensions of the channel

are fixed as $|C|\|N\|\|Q|$; hence, the optimisation is of a continuous function over a compact domain, so we have a maximum rather than a supremum.

Lemma 3.1. *The functions* $Z_\epsilon(\omega)$ *and* $J_\epsilon(\omega)$ *have the following properties:*

1. *They are non-decreasing functions of* ϵ.

2. *They are concave in* ϵ.

3. *They are continuous for* $\epsilon \geq 0$.

4. *For any two states* $\omega_1^{C_1 N_1 Q_1 R_1}$ *and* $\omega_2^{C_2 N_2 Q_2 R_2}$ *and for* $\epsilon \geq 0$,

$$J_\epsilon(\omega_1 \otimes \omega_2) \leq J_\epsilon(\omega_1) + J_\epsilon(\omega_2),$$
$$Z_\epsilon(\omega_1 \otimes \omega_2) \leq Z_\epsilon(\omega_1) + Z_\epsilon(\omega_2).$$

5. *At* $\epsilon = 0$, $Z_0(\omega) = S(N|C)_\omega$ *and* $J_0(\omega) = 0$.

The proof of this lemma follows in the next section. Now we show how it is used to prove the converse (optimality) of Theorem 3.2. As a guide to reading the subsequent proof, we remark that in Eqs. (3.23) and (3.27), the environment systems VW of the encoding-decoding operations appear in the terms $I(\hat{N}^n VW : \hat{C}^n \hat{Q}^n | C'^n)$ and $S(\hat{N}^n VW | C'^n)$, which are bounded by the functions $J_\epsilon(\omega^{\otimes n})$ and $Z_\epsilon(\omega^{\otimes n})$, respectively. As stated in point 4 of Lemma 3.1, these functions are sub-additive, so basically we can single-letterize the terms appearing in the converse. Moreover, from point 3 of Lemma 3.1, we know that these functions are continuous for $\epsilon \geq 0$; therefore, the limit points of these functions are equal to the values of these functions at $\epsilon = 0$. When the fidelity is equal to 1 ($\epsilon = 0$), the structure of the CPTP maps preserving the state ω^{CNQR} in Theorem 3.1 implies that $J_0(\omega) = 0$ and $Z_0(\omega) = S(N|C)_\omega$, as stated in point 5 of Lemma 3.1. Thereby, we conclude the converse bounds in Eqs. (3.26) and (3.30).

Proof of Theorem 3.2 (converse). We first get the following chain of inequal-

ities considering the process of the decoding of the information:

$$nQ + S(B_0) \geq S(M) + S(B_0) \tag{3.8}$$
$$\geq S(MB_0) \tag{3.9}$$
$$= S(\hat{C}^n \hat{N}^n \hat{Q}^n V) \tag{3.10}$$
$$= S(\hat{C}^n \hat{Q}^n) + S(\hat{N}^n V | \hat{C}^n \hat{Q}^n) \tag{3.11}$$
$$\geq nS(CQ) + S(\hat{N}^n V | \hat{C}^n \hat{Q}^n) - n\delta(n, \epsilon) \tag{3.12}$$
$$\geq nS(CQ) + S(\hat{N}^n V | \hat{C}^n \hat{Q}^n C'^n) - n\delta(n, \epsilon) \tag{3.13}$$
$$= nS(CQ) + S(\hat{N}^n V | \hat{C}^n \hat{Q}^n C'^n) - S(\hat{N}^n V | C'^n)$$
$$\quad + S(\hat{N}^n V | C'^n) - n\delta(n, \epsilon)$$
$$= nS(CQ) - I(\hat{N}^n V : \hat{C}^n \hat{Q}^n | C'^n) + S(\hat{N}^n V | C'^n) - n\delta(n, \epsilon)$$
$$\geq nS(CQ) - I(\hat{N}^n VW : \hat{C}^n \hat{Q}^n | C'^n) + S(\hat{N}^n V | C'^n) - n\delta(n, \epsilon) \tag{3.14}$$

where Eq. (3.8) follows because the entropy of a system is bounded by the logarithm of the dimension of that system; Eq. (3.9) is due to sub-additivity of the entropy; Eq. (3.10) follows because the decoding isometry $U_{\mathcal{D}} : MB_0 \hookrightarrow \hat{C}^n \hat{N}^n \hat{Q}^n V$ does not change the entropy; Eq. (3.11) is due to the chain rule; Eq. (3.12) follows from the decodability: the output state on systems $\hat{C}^n \hat{Q}^n$ is $2\sqrt{2\epsilon}$-close to the original state $C^n Q^n$ in trace norm; then the inequality follows by applying the Fannes-Audenaert inequality [72,73], where $\delta(n, \epsilon) = \sqrt{2\epsilon} \log(|C||Q|) + \frac{1}{n} h(\sqrt{2\epsilon})$; Eq. (3.13) is due to strong sub-additivity of the entropy, and system C' is a copy of classical system C; Eq. (3.14) follows from data processing inequality where W is the environment system of the encoding isometry $U_{\mathcal{E}} : C^n N^n Q^n A_0 \hookrightarrow MW$.

Moreover, considering the process of encoding the information, Q is bounded as follows:

$$nQ \geq S(M)$$
$$\geq S(M | WC'^n) \tag{3.15}$$
$$= S(MWC'^n) - S(WC'^n) \tag{3.16}$$
$$= S(C^n N^n Q^n A_0 C'^n) - S(WC'^n) \tag{3.17}$$
$$= S(C^n N^n Q^n C'^n) + S(A_0) - S(WC'^n) \tag{3.18}$$
$$= S(C^n N^n Q^n C'^n) + S(A_0) - S(C'^n) - S(W | C'^n) \tag{3.19}$$
$$= S(C^n N^n Q^n) + S(A_0) - S(C'^n) - S(W | C'^n) \tag{3.20}$$
$$= nS(CQ) + nS(N | CQ) + S(A_0) - nS(C') - S(W | C'^n) \tag{3.21}$$
$$= nS(CQ) + nS(N | C) + S(A_0) - nS(C') - S(W | C'^n), \tag{3.22}$$

where Eq. (3.15) is due to sub-additivity of the entropy; Eq. (3.16) is due to the chain rule; Eq. (3.17) follows because the encoding isometry $U_{\mathcal{E}}$: $C^n N^n Q^n A_0 \hookrightarrow MW$ does not the change the entropy; Eq. (3.18) follows because the initial entanglement A_0 is independent from the source; Eq. (3.19) is due to the chain rule; Eq. (3.20) follows because C' is a copy of the system C, so $S(C'|CNQ) = 0$; Eq. (3.21) is due to the chain rule and the fact that the entropy is additive for product states; Eq. (3.22) follows because conditional on system C the system N is independent from system Q.

Now, we add Eqs. (3.14) and (3.22); the entanglement terms $S(A_0)$ and $S(B_0)$ cancel out, and by dividing by $2n$ we obtain

$$
\begin{aligned}
Q \geq\ & S(CQ) - \frac{1}{2}S(C) + \frac{1}{2}S(N|C) - \frac{1}{2n}I(\hat{N}^n VW : \hat{C}^n \hat{Q}^n | C'^n) \\
& + \frac{1}{2n}S(\hat{N}^n V|C'^n) - \frac{1}{2n}S(W|C'^n) - \frac{1}{2}\delta(n,\epsilon) \\
\geq\ & S(CQ) - \frac{1}{2}S(C) + \frac{1}{2}S(N|C) - \frac{1}{2n}I(\hat{N}^n VW : \hat{C}^n \hat{Q}^n | C'^n) \\
& - \frac{1}{2n}S(\hat{N}^n VW|C'^n) - \frac{1}{2}\delta(n,\epsilon) \hspace{5em} (3.23) \\
\geq\ & S(CQ) - \frac{1}{2}S(C) + \frac{1}{2}S(N|C) - \frac{1}{2n}J_\epsilon(\omega^{\otimes n}) - \frac{1}{2n}Z_\epsilon(\omega^{\otimes n}) - \frac{1}{2}\delta(n,\epsilon) \\
& \hspace{30em} (3.24) \\
\geq\ & S(CQ) - \frac{1}{2}S(C) + \frac{1}{2}S(N|C) - \frac{1}{2}J_\epsilon(\omega) - \frac{1}{2}Z_\epsilon(\omega) - \frac{1}{2}\delta(n,\epsilon), \hspace{2em} (3.25)
\end{aligned}
$$

where Eq. (3.23) follows from strong sub-additivity of the entropy, $S(\hat{N}^n V|C'^n) + S(\hat{N}^n V|WC'^n) \geq 0$; Eq. (3.24) follows from Definition 3.1; Eq. (3.25) is due to point 4 of Lemma 3.1.

In the limit of $\epsilon \to 0$ and $n \to \infty$, the qubit rate is thus bounded by

$$
\begin{aligned}
Q \geq\ & S(CQ) - \frac{1}{2}S(C) + \frac{1}{2}S(N|C) - \frac{1}{2}J_0(\omega) - \frac{1}{2}Z_0(\omega) \\
=\ & S(CQ) - \frac{1}{2}S(C), \hspace{18em} (3.26)
\end{aligned}
$$

where the equality follows from point 5 of Lemma 3.1.

Moreover, from Eq. (3.14) we have:

$$
\begin{aligned}
nQ + S(B_0) =\ & nQ + nE \\
\geq\ & nS(CQ) - I(\hat{N}^n VW : \hat{C}^n \hat{Q}^n | C'^n) + S(\hat{N}^n V|C'^n) - n\delta(n,\epsilon) \\
\geq\ & nS(CQ) - I(\hat{N}^n VW : \hat{C}^n \hat{Q}^n | C'^n) - n\delta(n,\epsilon) \hspace{4em} (3.27) \\
\geq\ & nS(CQ) - J_\epsilon(\omega^{\otimes n}) - n\delta(n,\epsilon) \hspace{9em} (3.28) \\
\geq\ & nS(CQ) - nJ_\epsilon(\omega) - n\delta(n,\epsilon), \hspace{8em} (3.29)
\end{aligned}
$$

where Eq. (3.27) follows because the entropy conditional on a classical system is positive, $S(\hat{N}^n V | C'^n) \geq 0$; Eq. (3.28) follows from Definition 3.1; Eq. (3.29) is due to point 4 of Lemma 3.1.

In the limit of $\epsilon \to 0$ and $n \to \infty$, we thus obtain the following bound on the rate sum:

$$Q + E \geq S(CQ) - J_0(\omega) = S(CQ), \qquad (3.30)$$

where the equality follows from point 5 of Lemma 3.1. ∎

Remark 3.3. *Our lower bound on $Q + E$ in Eq. (3.30) reproduces the result of Koashi and Imoto [31] for the case of a classical-quantum source $\rho^{AX} = \sum_x p(x) \rho_x^A \otimes |x\rangle\langle x|^X$. This is because a code with qubit-ebit rate pair (Q, E) gives rise to a compression code in the sense of Koashi and Imoto using a rate of qubits $Q + E$ and no prior entanglement, simply by first distributing E ebits and then using the entanglement assisted code.*

It is worth noting that conversely, Eq. (3.30) can be obtained from the Koashi-Imoto result, as follows. Any good code for ρ^{AR} is automatically a good code for the classical-quantum source of mixed states

$$\rho^{AY} = \sum_y q(y) \rho_y^A \otimes |y\rangle\langle y|^Y = \sum_y \mathrm{Tr}_R \rho^{AR} (\mathbb{1}_A \otimes M_y^R) \otimes |y\rangle\langle y|^Y,$$

for any POVM (M_y) on R, simply by the monotonicity of the fidelity under CPTP maps. As discussed before, by choosing an informationally complete measurement, the KI-decomposition of the ensemble $\{q(y), \rho_y^A\}$ is identical to that of ρ^{AR} in Theorem 3.1. Thus the unassisted qubit compression rate of ρ^{AY} and of ρ^{AR} are lower bounded by the same quantity, the right hand side of Eq. (3.30).

3.4   Proof of Lemma 3.1

1. The definitions of the functions $J_\epsilon(\omega)$ and $Z_\epsilon(\omega)$ directly imply that they are non-decreasing functions of ϵ.

2. We first prove the concavity of $Z_\epsilon(\omega)$. Let $U_1 : CNQ \hookrightarrow \hat{C}\hat{N}\hat{Q}E$ and $U_2 : CNQ \hookrightarrow \hat{C}\hat{N}\hat{Q}E$ be the isometries attaining the maximum for ϵ_1 and ϵ_2, respectively, which act as follows on the purification $|\omega\rangle^{CNQRC'R'}$ of the previously introduced state $\omega^{CNQRC'}$:

$$|\tau_1\rangle^{\hat{C}\hat{N}\hat{Q}ERC'R'} = (U_1 \otimes \mathbb{1}_{RC'R'}) |\omega\rangle^{CNQRC'R'} \qquad \text{and}$$
$$|\tau_2\rangle^{\hat{C}\hat{N}\hat{Q}ERC'R'} = (U_2 \otimes \mathbb{1}_{RC'R'}) |\omega\rangle^{CNQRC'R'},$$

where $\mathrm{Tr}_{R'}\big[|\omega\rangle\langle\omega|^{CNQRC'R'}\big] = \omega^{CNQRC'}$. For $0 \le \lambda \le 1$, define the isometry $U_0 : CNQ \hookrightarrow \hat{C}\hat{N}\hat{Q}EFF'$ which acts as

$$U_0 := \sqrt{\lambda}U_1 \otimes |11\rangle^{FF'} + \sqrt{1-\lambda}U_2 \otimes |22\rangle^{FF'}, \qquad (3.31)$$

where systems F and F' are qubits, and which leads to the state

$$(U_0 \otimes \mathbb{1}_{RC'R'})|\omega\rangle^{CNQRC'R'}$$
$$= \sqrt{\lambda}|\tau_1\rangle^{\hat{C}\hat{N}\hat{Q}ERC'R'}|11\rangle^{FF'} + \sqrt{1-\lambda}|\tau_2\rangle^{\hat{C}\hat{N}\hat{Q}ERC'R'}|22\rangle^{FF'}.$$

Then, U_0 defines its state τ. for which the reduced state on the systems $\hat{C}\hat{N}\hat{Q}RC'$ is

$$\tau^{\hat{C}\hat{N}\hat{Q}RC'} = \lambda\tau_1^{\hat{C}\hat{N}\hat{Q}RC'} + (1-\lambda)\tau_2^{\hat{C}\hat{N}\hat{Q}RC'}. \qquad (3.32)$$

Therefore, the fidelity for the state τ is bounded as follows:

$$F(\omega^{CNQR}, \tau^{\hat{C}\hat{N}\hat{Q}R})$$
$$= F(\omega^{CNQR}, \lambda\tau_1^{\hat{C}\hat{N}\hat{Q}R} + (1-\lambda)\tau_2^{\hat{C}\hat{N}\hat{Q}R})$$
$$= F(\lambda\omega^{CNQR} + (1-\lambda)\omega^{CNQR}, \lambda\tau_1^{\hat{C}\hat{N}\hat{Q}R} + (1-\lambda)\tau_2^{\hat{C}\hat{N}\hat{Q}R})$$
$$\ge \lambda F(\omega^{CNQR}, \tau_1^{\hat{C}\hat{N}\hat{Q}R}) + (1-\lambda)F(\omega^{CNQR}, \tau_2^{\hat{C}\hat{N}\hat{Q}R})$$
$$\ge 1 - (\lambda\epsilon_1 + (1-\lambda)\epsilon_2). \qquad (3.33)$$

The first inequality is due to simultaneous concavity of the fidelity in both arguments; the last line follows by the definition of the isometries U_1 and U_2. Thus, the isometry U_0 yields a fidelity of at least $1 - (\lambda\epsilon_1 + (1-\lambda)\epsilon_2) =: 1-\epsilon$. Now let $E' = EFF'$ denote the environment of the isometry U_0 defined above. According to Definition 3.1, we obtain

$$Z_\epsilon(\omega) \ge S(\hat{N}E'|C')_\tau$$
$$= S(\hat{N}EFF'|C')_\tau$$
$$= S(F|C')_\tau + S(\hat{N}E|FC')_\tau + S(F'|\hat{N}EFC')_\tau \qquad (3.34)$$
$$\ge S(\hat{N}E|FC')_\tau \qquad (3.35)$$
$$= \lambda S(\hat{N}E|C')_{\tau_1} + (1-\lambda)S(\hat{N}E|C')_{\tau_2} \qquad (3.36)$$
$$= \lambda Z_{\epsilon_1}(\omega) + (1-\lambda)Z_{\epsilon_2}(\omega), \qquad (3.37)$$

where the state τ in the entropies is given in Eq. (3.32); Eq. (3.34) is due to the chain rule; Eq. (3.35) follow because for the state on systems $\hat{N}EFF'C'$ we have $S(F'|C') + S(F'|\hat{N}EFC') \ge 0$ which follows from

strong sub-additivity of the entropy; Eq. (3.36) follows by expanding the conditional entropy on the classical system F; Eq. (3.37) follows from the definitions of the isometries U_1 and U_2.

Moreover, let $U_1 : CNQ \hookrightarrow \hat{C}\hat{N}\hat{Q}E$ and $U_2 : CNQ \hookrightarrow \hat{C}\hat{N}\hat{Q}E$ be the isometries attaining the maximum for ϵ_1 and ϵ_2 in the definition of $J_\epsilon(\omega)$, respectively. Again, define the isometry U_0 as in Eq. (3.31), which leads to the bound on the fidelity as in Eq. (3.33), letting $E' = EFF'$ be the environment of the isometry U_0. According to Definition 3.1, we obtain

$$J_\epsilon(\omega) \geq I(\hat{N}EFF' : \hat{C}\hat{Q}|C')_\tau$$
$$\geq I(\hat{N}EF : \hat{C}\hat{Q}|C')_\tau \qquad\qquad (3.38)$$
$$= I(F : \hat{C}\hat{Q}|C')_\tau + I(\hat{N}E : \hat{C}\hat{Q}|FC')_\tau \qquad\qquad (3.39)$$
$$\geq I(\hat{N}E : \hat{C}\hat{Q}|FC')_\tau \qquad\qquad (3.40)$$
$$= \lambda I(\hat{N}E : \hat{C}\hat{Q}|C')_{\tau_1} + (1 - \lambda)I(\hat{N}E : \hat{C}\hat{Q}|C')_{\tau_2} \qquad\qquad (3.41)$$
$$= \lambda J_{\epsilon_1}(\omega) + (1 - \lambda)J_{\epsilon_2}(\omega), \qquad\qquad (3.42)$$

where Eq. (3.38) follows from data processing; Eq. (3.39) is due to the chain rule for mutual information; Eq. (3.40) follows from strong sub-additivity of the entropy, $I(F : \hat{C}\hat{Q}|C')_\tau \geq 0$; Eq. (3.41) is obtained by expanding the conditional mutual information on the classical system F; finally, Eq. (3.42) follows from the definitions of the isometries U_1 and U_2.

3. The functions are non-decreasing and concave for $\epsilon \geq 0$, so they are continuous for $\epsilon > 0$. The concavity implies furthermore that J_ϵ and Z_ϵ are lower semi-continuous at $\epsilon = 0$. On the other hand, since the fidelity, the conditional entropy and the conditional mutual information are all continuous functions of CPTP maps, and the domain of both optimizations is a compact set, we conclude that $J_\epsilon(\omega)$ and Z_ϵ are also upper semi-continuous at $\epsilon = 0$, so they are continuous at $\epsilon = 0$ [74, Thms. 10.1 and 10.2].

4. We first prove $Z_\epsilon(\omega_1 \otimes \omega_2) \leq Z_\epsilon(\omega_1) + Z_\epsilon(\omega_2)$. In the definition of $Z_\epsilon(\omega_1 \otimes \omega_2)$, let the isometry $U_0 : C_1N_1Q_1C_2N_2Q_2 \hookrightarrow \hat{C}_1\hat{N}_1\hat{Q}_1\hat{C}_2\hat{N}_2\hat{Q}_2E$ be the one attaining the maximum, which acts on the following purified

39

source states with purifying systems R_1' and R_2':

$$|\tau\rangle^{\hat{C}_1\hat{N}_1\hat{Q}_1\hat{C}_2\hat{N}_2\hat{Q}_2ER_1C_1'R_1'R_2C_2'R_2'} \tag{3.43}$$

$$= (U_0 \otimes \mathbb{1}_{R_1C_1'R_1'R_2C_2'R_2'})|\omega_1\rangle^{C_1N_1Q_1R_1C_1'R_1'} \otimes |\omega_2\rangle^{C_2N_2Q_2R_2C_2'R_2'}. \tag{3.44}$$

By definition, the fidelity is bounded by

$$F(\omega_1^{C_1N_1Q_1R_1} \otimes \omega_2^{C_2N_2Q_2R_2}, \tau^{\hat{C}_1\hat{N}_1\hat{Q}_1\hat{C}_2\hat{N}_2\hat{Q}_2R_1R_2}) \geq 1 - \epsilon.$$

Now, we can define an isometry $U_1 : C_1N_1Q_1 \hookrightarrow \hat{C}_1\hat{N}_1\hat{Q}_1E_1$ acting only on systems $C_1N_1Q_1$, by letting $U_1 = (U_0 \otimes \mathbb{1}_{R_2C_2'R_2'})(\mathbb{1}_{C_1N_1Q_1} \otimes |\omega_2\rangle^{C_2N_2Q_2R_2C_2'R_2'})$ and with the environment $E_1 := \hat{C}_2\hat{N}_2\hat{Q}_2ER_2C_2'R_2'$. It has the property that $|\tau\rangle^{\hat{C}_1\hat{N}_1\hat{Q}_1R_1C_1'R_1'E} = (U_1 \otimes \mathbb{1}_{R_1C_1'R_1'})|\omega_1\rangle^{C_1N_1Q_1R_1C_1'R_1'}$ has the same reduced state on $\hat{C}_1\hat{N}_1\hat{Q}_1R_1$ as τ from Eq. (3.44). This isometry preserves the fidelity for ω_1, which follows from monotonicity of the fidelity under partial trace:

$$\begin{aligned}
F(\omega_1^{C_1N_1Q_1R_1}, \tau_1^{\hat{C}_1\hat{N}_1\hat{Q}_1R_1}) &= F(\omega_1^{C_1N_1Q_1R_1}, \tau^{\hat{C}_1\hat{N}_1\hat{Q}_1R_1}) \\
&\geq F(\omega_1^{C_1N_1Q_1R_1} \otimes \omega_2^{C_2N_2Q_2R_2}, \tau^{\hat{C}_1\hat{N}_1\hat{Q}_1\hat{C}_2\hat{N}_2\hat{Q}_2R_1R_2}) \\
&\geq 1 - \epsilon.
\end{aligned}$$

By the same argument, there is the following isometry

$$U_2 : C_2N_2Q_2 \hookrightarrow \hat{C}_1\hat{N}_1\hat{Q}_1\hat{C}_2\hat{N}_2\hat{Q}_2ER_1C_1'R_1',$$

with output system $\hat{C}_2\hat{N}_2\hat{Q}_2$ and environment $E_2 := \hat{C}_1\hat{N}_1\hat{Q}_1ER_1C_1'R_1'$, such that

$$\begin{aligned}
F(\omega_2^{C_2N_2Q_2R_2}, \tau_2^{\hat{C}_2\hat{N}_2\hat{Q}_2R_2}) &= F(\omega_2^{C_2N_2Q_2R_2}, \tau^{\hat{C}_2\hat{N}_2\hat{Q}_2R_2}) \\
&\geq F(\omega_1^{C_1N_1Q_1R_1} \otimes \omega_2^{C_2N_2Q_2R_2}, \tau^{\hat{C}_1\hat{N}_1\hat{Q}_1\hat{C}_2\hat{N}_2\hat{Q}_2R_1R_2}) \\
&\geq 1 - \epsilon.
\end{aligned}$$

Therefore, we obtain:

$$
\begin{aligned}
Z_\epsilon(\omega_1) &+ Z_\epsilon(\omega_2) - Z_\epsilon(\omega_1 \otimes \omega_2) \\
&\geq S(\hat{N}_1 E_1 | C_1')_\tau + S(\hat{N}_2 E_2 | C_2')_\tau - S(\hat{N}_1 \hat{N}_2 E | C_1' C_2')_\tau \qquad (3.45) \\
&= S(\hat{N}_1 E_1 C_1')_\tau + S(\hat{N}_2 E_2 C_2')_\tau - S(\hat{N}_1 \hat{N}_2 E C_1' C_2')_\tau \\
&\qquad\qquad\qquad\qquad\quad - S(C_1') - S(C_2') + S(C_1' C_2') \qquad (3.46) \\
&= S(\hat{N}_1 E_1 C_1')_\tau + S(\hat{N}_2 E_2 C_2')_\tau - S(\hat{N}_1 \hat{N}_2 E C_1' C_2')_\tau \qquad (3.47) \\
&= S(\hat{C}_1 \hat{Q}_1 R_1 R_1') + S(\hat{C}_2 \hat{Q}_2 R_2 R_2') - S(\hat{C}_1 \hat{Q}_1 \hat{C}_2 \hat{Q}_2 R_1 R_1' R_2 R_2') \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (3.48) \\
&= I(\hat{C}_1 \hat{Q}_1 R_1 R_1' : \hat{C}_2 \hat{Q}_2 R_2 R_2') \\
&\geq 0, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad (3.49)
\end{aligned}
$$

where Eq. (3.45) is due to Definition 3.1; Eq. (3.46) is due to the chain rule; Eq. (3.47) because the systems C_1' and C_2' are independent from each other; Eq. (3.48) follows because the overall state on systems $\hat{C}_1 \hat{N}_1 \hat{Q}_1 \hat{C}_2 \hat{N}_2 \hat{Q}_2 E R_1 C_1' R_1' R_2 C_2' R_2'$ is pure; Eq. (3.49) is due to sub-additivity of the entropy.

To prove prove $J_\epsilon(\omega_1 \otimes \omega_2) \leq J_\epsilon(\omega_1) + J_\epsilon(\omega_2)$, let the isometry $U_0 : C_1 N_1 Q_1 C_2 N_2 Q_2 \hookrightarrow \hat{C}_1 \hat{N}_1 \hat{Q}_1 \hat{C}_2 \hat{N}_2 \hat{Q}_2 E$ be the one attaining the maximum in definition of $J_\epsilon(\omega_1 \otimes \omega_2)$, which acts on the following purified source states with purifying systems R_1' and R_2', as in Eq. (3.44). By definition, the fidelity is bounded as

$$
F\left(\omega_1^{C_1 N_1 Q_1 R_1} \otimes \omega_2^{C_2 N_2 Q_2 R_2}, \tau^{\hat{C}_1 \hat{N}_1 \hat{Q}_1 \hat{C}_2 \hat{N}_2 \hat{Q}_2 R_1 R_2}\right) \geq 1 - \epsilon.
$$

Now define $U_1 : C_1 N_1 Q_1 \hookrightarrow \hat{C}_1 \hat{N}_1 \hat{Q}_1 \hat{C}_2 \hat{N}_2 \hat{Q}_2 E R_2 C_2' R_2'$ and $U_2 : C_2 N_2 Q_2 \hookrightarrow \hat{C}_1 \hat{N}_1 \hat{Q}_1 \hat{C}_2 \hat{N}_2 \hat{Q}_2 E R_1 C_1' R_1'$ as in the above discussion, with the environments $E_1 \coloneqq \hat{C}_2 \hat{N}_2 \hat{Q}_2 E R_2 C_2' R_2'$ and $E_2 \coloneqq \hat{C}_1 \hat{N}_1 \hat{Q}_1 E R_1 C_1' R_1'$, respectively. Recall that the fidelity for the states ω_1 and ω_2 is at least $1 - \epsilon$, because of the monotonicity of the fidelity under partial trace. Thus

we obtain

$$
\begin{aligned}
J_\epsilon&(\omega_1) + J_\epsilon(\omega_2) - J_\epsilon(\omega_1 \otimes \omega_2) \\
&\geq I(\hat{N}_1 E_1 : \hat{C}_1 \hat{Q}_1 | C_1')_\tau + I(\hat{N}_2 E_2 : \hat{C}_2 \hat{Q}_2 | C_2')_\tau \\
&\quad - I(\hat{N}_1 \hat{N}_2 E : \hat{C}_1 \hat{Q}_1 \hat{C}_2 \hat{Q}_2 | C_1' C_2')_\tau \\
&= S(\hat{N}_1 E_1 C_1') + S(\hat{C}_1 \hat{Q}_1 C_1') - S(\hat{C}_1 \hat{N}_1 \hat{Q}_1 E_1 C_1') - S(C_1') \\
&\quad + S(\hat{N}_2 E_2 C_2') + S(\hat{C}_2 \hat{Q}_2 C_2') - S(\hat{C}_2 \hat{N}_2 \hat{Q}_2 E_2 C_2') - S(C_2') \\
&\quad - S(\hat{N}_1 \hat{N}_2 E C_1' C_2') - S(\hat{C}_1 \hat{Q}_1 \hat{C}_2 \hat{Q}_2 C_1' C_2') \\
&\quad + S(\hat{C}_1 \hat{N}_1 \hat{Q}_1 \hat{C}_2 \hat{N}_2 \hat{Q}_2 E C_1' C_2') + S(C_1' C_2') \\
&= S(\hat{C}_1 \hat{Q}_1 R_1 R_1') + S(\hat{C}_1 \hat{Q}_1 C_1') - S(R_1 R_1') - S(C_1') \\
&\quad + S(\hat{C}_2 \hat{Q}_2 R_2 R_2') + S(\hat{C}_2 \hat{Q}_2 C_2') - S(R_2 R_2') - S(C_2') \\
&\quad - S(\hat{C}_1 \hat{Q}_1 \hat{C}_2 \hat{Q}_2 R_1 R_1' R_2 R_2') - S(\hat{C}_1 \hat{Q}_1 \hat{C}_2 \hat{Q}_2 C_1' C_2') \\
&\quad + S(R_1 R_1' R_2 R_2') + S(C_1' C_2') \\
&= I(\hat{C}_1 \hat{Q}_1 R_1 R_1' : \hat{C}_2 \hat{Q}_2 R_2 R_2') - I(R_1 R_1' : R_2 R_2') \\
&\quad + I(\hat{C}_1 \hat{Q}_1 C_1' : \hat{C}_2 \hat{Q}_2 C_2') - I(C_1' : C_2') \\
&\geq I(R_1 R_1' : R_2 R_2') - I(R_1 R_1' : R_2 R_2') + I(C_1' : C_2') - I(C_1' : C_2') \\
&= 0,
\end{aligned}
$$

(3.50)
(3.51)
(3.52)
(3.53)

where Eq. (3.50) is due to Definition 3.1; In Eq. (3.51) we expand the mutual informations in terms of entropies; Eq. (3.52) follows because the overall state on systems $\hat{C}_1 \hat{N}_1 \hat{Q}_1 \hat{C}_2 \hat{N}_2 \hat{Q}_2 E R_1 C_1' R_1' R_2 C_2' R_2'$ is pure; Eq. (3.53) is due to data processing.

5. According to Theorem 3.1 [41, 42], any isometry $U : CNQ \to \hat{C}\hat{N}\hat{Q}E$ acting on the state $\omega^{CNQRC'}$ which preserves the reduced state on systems $CNQRC'$ (C' here is considered as a part of the reference system), acts as the following:

$$
(U \otimes \mathbb{1}_{RC'}) \omega^{CNQRC'} (U^\dagger \otimes \mathbb{1}_{RC'}) = \sum_j p_j |j\rangle\langle j|^C \otimes U_j \omega_j^N U_j^\dagger \otimes \rho_j^{QR} \otimes |j\rangle\langle j|^{C'},
$$

where the isometry $U_j : N \to \hat{N}E$ satisfies $\mathrm{Tr}_E[U_j \omega_j^N U_j^\dagger] = \omega_j$. Therefore, in Definition 3.1 for $\epsilon = 0$, the final state is

$$
\tau^{\hat{C}\hat{N}\hat{Q}ERC'} = \sum_j p_j |j\rangle\langle j|^C \otimes U_j \omega_j^N U_j^\dagger \otimes \rho_j^{QR} \otimes |j\rangle\langle j|^{C'}.
$$

42

Thus we can directly evaluate

$$Z_0(\omega) = S(\hat{N}E|C')_\tau = S(N|C)_\omega \text{ and } J_0(\omega) = I(\hat{N}E : \hat{C}\hat{Q}|C')_\tau = 0,$$

concluding the proof. ∎

3.5 Discussion

We have introduced a common framework for all single-source quantum compression problems, i.e. settings without side information at the encoder or the decoder, by defining the compression task as the reproduction of a given bipartite state between the system to be compressed and a reference. That state, which defines the task, can be completely general, and special instances recover Schumacher's quantum source compression (in both variants of a pure state ensemble and of a pure entangled state) [8] and compression of a mixed state ensemble source in the blind variant [31,62].

Our general result gives the optimal quantum compression rate in terms of qubits per source, both in the settings without and with entanglement, and indeed the entire qubit-ebit rate region, reproducing the aforementioned special cases, along with other previously considered problems [48]. Despite the technical difficulties in obtaining it, the end result has a simple and intuitive interpretation. Namely, the given source ρ^{AR} is equivalent to a source in standard Koashi-Imoto form,

$$\omega^{CQR} = \sum_j p_j |j\rangle\langle j|^C \otimes \rho_j^{QR},$$

so that j has to be compressed as classical information, at rate $S(C)$, and Q as quantum information, at rate $S(Q|C)$; in the presence of entanglement, the former rate is halved while the latter is maintained. Indeed, what our Theorem 3.2 shows is that the original source has the same qubit-ebit rate region as the clean classical-quantum mixed source

$$\Omega^{CQRR'C'} = \sum_j p_j |j\rangle\langle j|^C \otimes |\psi_j\rangle\langle\psi_j|^{QRR'} \otimes |j\rangle\langle j|^{C'},$$

where $|\psi_j\rangle^{QRR'}$ purifies ρ_j^{QR}, and $RR'C'$ is considered the reference. In Ω, C is indeed a manifestly classical source, since it is duplicated in the reference system, and conditional on C, Q is a genuinely quantum source since it is purely entangled with the reference system. As $\text{Tr}_{R'C'}\Omega^{CQRR'C'} = \omega^{CQR}$, any code and any achievable rates for Ω are good for ω, and that is how the achievability of the rate region in Theorem 3.2 can be described. The

43

opposite, that a code good for ω should be good for Ω, is far from obvious. Indeed, if that were true, it would not only yield a quick and simple proof of our converse bounds, but would imply that the rate region of Theorem 3.2 satisfies a strong converse! However, as we do not know this reduction to the source Ω, our converse proceeds via a more complicated, indirect route, and yields only a weak converse. Whether the strong converse holds, and what the detailed relation between the sources ω^{CQR} and $\Omega^{CQRR'C'}$ is, remain open questions.

Chapter 4

Unification of the blind and visible Schumacher compression

In this chapter, we ask how the quantum compression of ensembles of pure states is affected by the availability of entanglement, and in settings where the encoder has access to side information. We find the optimal asymptotic quantum rate and the optimal tradeoff (rate region) of quantum and entanglement rates. It turns out that the amount by which the quantum rate beats the Schumacher limit, the entropy of the source, is precisely half the entropy of classical information that can be extracted from the source and side information states without disturbing them at all ("reversible extraction of classical information").

In the special case that the encoder has no side information, or that she has access to the identity of the states, this problem reduces to the known settings of *blind* and *visible* Schumacher compression, respectively, albeit here additionally with entanglement assistance. We comment on connections to previously studied and further rate tradeoffs when also classical information is considered. This chapter is based on the papers in [47, 48].

4.1   The source model

The task of data compression of a quantum source, introduced by Schumacher [8], marks one of the foundations of quantum information theory: not only did it provide an information theoretic interpretation of the von Neumann entropy $S(\rho) = -\operatorname{Tr}\rho\log\rho$ as the minimum compression rate, it also motivated the very concept of the qubit! In the Schumacher modelling, a source is given by an ensemble $\mathcal{E} = \{p(x), |\psi_x\rangle\langle\psi_x|\}$ of pure states $\psi_x = |\psi_x\rangle\langle\psi_x| \in \mathcal{S}(A)$, $|\psi_x\rangle \in A$, with a Hilbert space A of finite dimension $|A| < \infty$; $\mathcal{S}(A)$ denotes

the set of states (density operators). Furthermore, $x \in \mathcal{X}$ ranges over a discrete alphabet, so that we can can describe the source equivalently by the classical-quantum (cq) state $\omega = \sum_x p(x)|x\rangle\langle x|^X \otimes |\psi_x\rangle\langle\psi_x|^A$.

While the achievability of the rate $S(A)_\omega = S(\omega^A)$ was shown in [8,60] (see also [75, Thm. 1.18]), the full (weak) converse was established in [61], a simplified proof being given by M. Horodecki [62]; the strong converse was proved in [30].

In this chapter, we consider a more comprehensive model, where on the one hand the sender/encoder of the compressed data (Alice) has access to side information, namely a pure state σ_x^C in addition to the source state ψ_x^A, and on the other hand, she and the receiver/decoder of the compressed data (Bob) share pure state entanglement in the form of EPR pairs at a certain rate.

Thus, the source is now an ensemble $\mathcal{E} = \{p(x), |\psi_x\rangle\langle\psi_x|^A \otimes |\sigma_x\rangle\langle\sigma_x|^C\}$ of product states, which can be described equivalently by the cqq-state

$$\omega^{XAC} = \sum_{x \in \mathcal{X}} p(x)|x\rangle\langle x|^X \otimes |\psi_x\rangle\langle\psi_x|^A \otimes |\sigma_x\rangle\langle\sigma_x|^C. \qquad (4.1)$$

Yet another equivalent description is via the random variable $X \in \mathcal{X}$, distributed according to p, i.e. $\Pr\{X = x\} = p_x$; this also makes the pure states ψ_X and σ_X random variables.

We will consider the information theoretic limit of many copies of ω, i.e. $\omega^{X^n A^n C^n} = (\omega^{XAC})^{\otimes n}$:

$$\omega^{X^n A^n C^n} = \sum_{x^n \in \mathcal{X}^n} p(x^n)|x^n\rangle\langle x^n|^{X^n} \otimes |\psi_{x^n}\rangle\langle\psi_{x^n}|^{A^n} \otimes |\sigma_{x^n}\rangle\langle\sigma_{x^n}|^{C^n},$$

using the notation

$$x^n = x_1 x_2 \dots x_n, \quad p(x^n) = p(x_1)p(x_2)\cdots p(x_n),$$
$$|x^n\rangle = |x_1\rangle|x_2\rangle\cdots|x_n\rangle, \quad |\psi_{x^n}\rangle = |\psi_{x_1}\rangle|\psi_{x_2}\rangle\cdots|\psi_{x_n}\rangle.$$

4.2   Compression assisted by entanglement

We assume that the encoder, Alice, and the decoder, Bob, have initially a maximally entangled state $\Phi_K^{A_0 B_0}$ on registers A_0 and B_0 (both of dimension K). With probability $p(x^n)$, the source provides Alice with the state $\psi_{x^n}^{A^n} \otimes \sigma_{x^n}^{C^n}$. Then, Alice performs her encoding operation $\mathcal{C} : A^n C^n A_0 \longrightarrow \hat{C}^n C_A$ on the systems A^n, C^n and her part A_0 of the entanglement, which is a quantum channel, i.e. a completely positive and trace preserving (CPTP) map. (Note that our notation is a slight abuse, which we maintain as it

is simpler while it cannot lead to confusions, since channels really are maps between the trace class operators on the involved Hilbert spaces.) The dimension of the compressed system obviously has to be smaller than the original source, i.e. $|C_A| \le |A|^n$. We call $Q = \frac{1}{n} \log |C_A|$ and $E = \frac{1}{n} \log K$ the quantum and entanglement rates of the compression protocol, respectively. The system C_A is then sent to Bob via a noiseless quantum channel, who performs a decoding operation $\mathcal{D} : C_A B_0 \longrightarrow \hat{A}^n$ on the system C_A and his part of entanglement B_0.

According to Stinespring's theorem [55], all these CPTP maps can be dilated to isometries $V_A : A^n C^n A_0 \hookrightarrow \hat{C}^n C_A W_A$ and $V_B : C_A B_0 \hookrightarrow \hat{A}^n W_B$, where the new systems W_A and W_B are the environment systems of Alice and Bob, respectively.

We say the encoding-decoding scheme has fidelity $1 - \epsilon$, or error ϵ, if

$$
\begin{aligned}
\overline{F} &:= F\left(\omega^{X^n \hat{A}^n \hat{C}^n}, \xi^{X^n \hat{A}^n \hat{C}^n} \right) \\
&= \sum_{x^n \in \mathcal{X}^n} p(x^n) F\left(|\psi_{x^n}\rangle\langle\psi_{x^n}|^{A^n} \otimes |\sigma_{x^n}\rangle\langle\sigma_{x^n}|^{C^n}, \xi_{x^n}^{\hat{A}^n \hat{C}^n} \right) \qquad (4.2) \\
&\ge 1 - \epsilon,
\end{aligned}
$$

where $\xi^{X^n \hat{A}^n \hat{C}^n} = \sum_{x^n} p(x^n) |x\rangle\langle x|^{X^n} \otimes \xi_{x^n}^{\hat{A}^n \hat{C}^n}$ and $\xi_{x^n}^{\hat{A}^n \hat{C}^n} = (\mathcal{D} \circ \mathcal{C}) |\psi_{x^n}\rangle\langle\psi_{x^n}|^{A^n} \otimes |\sigma_{x^n}\rangle\langle\sigma_{x^n}|^{C^n} \otimes \Phi_K^{A_0 B_0}$. We say that (E, Q) is an (asymptotically) achievable rate pair if for all n there exist codes such that the fidelity converges to 1, and the entanglement and quantum rates converge to E and Q, respectively. The rate region is the set of all achievable rate pairs, as a subset of $\mathbb{R} \times \mathbb{R}_{\ge 0}$.

Note that this means that we demand not only that Bob can reconstruct the source states ψ_{x^n} with high fidelity on average, but that Alice retains the side information states σ_{x^n} as well with high fidelity.

There are two extreme cases of the side information that have been considered in the literature: If C is a trivial system, or more generally if the states σ_x^C are all identical, then the aforementioned task is the entanglement-assisted version of *blind* Schumacher compression. If $C = X$, or more precisely $|\sigma_x\rangle = |x\rangle$, then Alice has access to classical random variable X, and the task reduces to *visible* Schumacher compression with entanglement assistance. The blind-visible terminology is originally from [61, 64].

Remark 4.1. *In the case of no entanglement being available, i.e. $E = 0$ ($K = 1$), the problem is fully understood: The asymptotic rate $Q = S(A)$ from [8, 60] is achievable without touching the side information, and it is optimal, even in the visible case (which includes all other side informations), by the weak and strong converses of [61, 62] and [30].* ∎

47

4.3 Optimal quantum rate

To formulate the minimum compression rate under unlimited entanglement assistance, we need the following concept.

Definition 4.1. *An ensemble of pure states $\mathcal{E} = \{p(x), |\psi_x\rangle\langle\psi_x|^A \otimes |\sigma_x\rangle\langle\sigma_x|^C\}_{x \in \mathcal{X}}$ is called* reducible *if its states fall into two or more orthogonal subspaces. Otherwise the ensemble $\mathcal{E}$ is called* irreducible. *We apply the same terminology to the source cqq-state ω^{XAC}.*

Notice that a reducible ensemble can be written uniquely as a disjoint union of irreducible ensembles $\mathcal{E} = \biguplus_{y \in \mathcal{Y}} q(y)\mathcal{E}_y$, with a partition $\mathcal{X} = \biguplus_{y \in \mathcal{Y}} \mathcal{X}_y$ and irreducible ensembles

$$\mathcal{E}_y = \{p(x|y), |\psi_x\rangle\langle\psi_x|^A \otimes |\sigma_x\rangle\langle\sigma_x|^C\}_{x \in \mathcal{X}_y},$$

where $q(y)p(x|y) = p(x)$ for $x \in \mathcal{X}_y$ and $q(y) = \sum_{x \in \mathcal{X}_y} p(x)$. We define the subspace spanned by the vectors of each irreducible ensemble as $F_y :=$ span$\{|\psi_x\rangle \otimes |\sigma_x\rangle : x \in \mathcal{X}_y\}$. The irreducible ensembles $\mathcal{E}_y$ are pairwise orthogonal, i.e. $F_{y'} \perp F_y$ for all $y' \neq y$. We may thus introduce the random variable $Y = Y(X)$ taking values in the set $\mathcal{Y}$ with probability distribution $q(y)$; namely, Y is a deterministic function of X such that $\Pr\{X \in \mathcal{X}_Y\} = 1$.

We define the *modified* source as

$$\omega^{XACY} = \sum_x p(x)|x\rangle\langle x|^X \otimes |\psi_x\rangle\langle\psi_x|^A \otimes |\sigma_x\rangle\langle\sigma_x|^C \otimes |y(x)\rangle\langle y(x)|^Y,$$

with side information systems CY. Because there is an isometry $V : AC \to ACY$ which acts as

$$V |\psi_x\rangle^A \otimes |\sigma_x\rangle^C = |\psi_x\rangle^A \otimes |\sigma_x\rangle^C \otimes |y(x)\rangle^Y , \tag{4.3}$$

the extended source ω^{XACY} is equivalent to the original source and side information ω^{XAC} modulo a local operation of Alice.

We first present the optimal asymptotic compression rate in the following theorem and prove the achievability of it, but we leave the converse proof to the end of this section, as it requires introducing further machinery.

Theorem 4.1. *For the given source ω^{XACY}, the optimal asymptotic compression rate assisted by unlimited entanglement is $Q = \frac{1}{2}(S(A) + S(A|CY))$.*

Furthermore, there is a protocol achieving this communication rate with entanglement consumption at rate $E = \frac{1}{2}(S(A) - S(A|CY))$.

Proof. We first show that this rate is achievable. Consider the following purification of ω^{XACY},

$$|\omega\rangle^{XX'ACY} = \sum_x \sqrt{p(x)}\, |x\rangle^X\, |x\rangle^{X'}\, |\psi_x\rangle^A\, |\sigma_x\rangle^C\, |y(x)\rangle^Y,$$

with side information systems CY. This is obtained from

$$|\omega\rangle^{XX'AC} = \sum_x \sqrt{p(x)}\, |x\rangle^X\, |x\rangle^{X'}\, |\psi_x\rangle^A\, |\sigma_x\rangle^C,$$

by Alice applying the isometry V from Eq. (4.3).

We apply quantum state redistribution (QSR) [25, 44] as a subprotocol, where the objective is for Alice to send to Bob A^n, using C^nY^n as side information, while $(XX')^n$ serves as reference system; the figure of merit is the fidelity with the original pure state $(\omega^{XX'ACY})^{\otimes n}$. Denoting the overall encoding-decoding CPTP map $\Lambda : A^nC^nY^n \to \hat{A}^n\hat{C}^n\hat{Y}^n$, QRS gives us the first inequality of the following chain:

$$1 - o(1) \le F\big(\omega^{X^nX'^nA^nC^nY^n}, (\mathrm{id}_{X^nX'^n} \otimes \Lambda)\omega^{X^nX'^nA^nC^nY^n}\big)$$
$$\le F\big(\omega^{X^nA^nC^nY^n}, (\mathrm{id}_{X^n} \otimes \Lambda)\omega^{X^nA^nC^nY^n}\big),$$

where the second inequality follows from monotonicity of the fidelity under partial trace. Thus, the protocol satisfies our fidelity criterion (4.2).

The communication rate we obtain from QSR is $Q = \frac{1}{2}I(A:XX') = \frac{1}{2}(S(A) + S(A|CY))$. Furthermore, QSR guarantees entanglement consumption at the rate $E = \frac{1}{2}I(A:CY) = \frac{1}{2}(S(A) - S(A|CY))$. $\blacksquare$

To prove optimality (the converse), we first need a few preparations. The following definition is inspired by the "reversible extraction of classical information" in [65].

Definition 4.2. *For a source ω^{XAC} and $\epsilon \ge 0$, define*

$$I_\epsilon(\omega) := \max_{V:AC\to\hat{A}\hat{C}W \text{ isometry}} I(X:\hat{C}W)_\xi \ \ s.t. \ \ F(\omega^{XAC}, \xi^{X\hat{A}\hat{C}}) \ge 1 - \epsilon,$$

where

$$\xi^{X\hat{A}\hat{C}W} = (\mathbb{1}_X \otimes V)\omega^{XAC}(\mathbb{1}_X \otimes V^\dagger) = \sum_x p(x)|x\rangle\langle x|^X \otimes |\xi_x\rangle\langle\xi_x|^{\hat{A}\hat{C}W}.$$

In this definition, the dimension of the environment is w.l.o.g. bounded as $|W| \le |A|^2|C|^2$; hence, the optimisation is of a continuous function over a compact domain, so we have a maximum rather than a supremum.

49

Lemma 4.1. *The function $I_\epsilon(\omega)$ has the following properties:*

1. *It is a non-decreasing function of ϵ.*

2. *It is concave in ϵ.*

3. *It is continuous for $\epsilon \geq 0$.*

4. *For any two states $\omega_1^{X_1 A_1 C_1}$ and $\omega_2^{X_2 A_2 C_2}$ and for $\epsilon \geq 0$, $I_\epsilon(\omega_1 \otimes \omega_2) \leq I_\epsilon(\omega_1) + I_\epsilon(\omega_2)$.*

5. *For any state ω^{XAC}, $I_0(\omega) \leq S(CY)$.*

Proof. 1. The definition of $I_\epsilon(\omega)$ directly implies that it is a non-decreasing function of ϵ.

2. To prove the concavity, let $V_1 : AC \to \hat{A}\hat{C}W$ and $V_2 : AC \to \hat{A}\hat{C}W$ be the isometries attaining the maximum for ϵ_1 and ϵ_2, respectively, which act as follows:

$$V_1 |\psi_x\rangle^A |\sigma_x\rangle^C = |\xi_x\rangle^{\hat{A}\hat{C}W} \text{ and } V_2 |\psi_x\rangle^A |\sigma_x\rangle^C = |\zeta_x\rangle^{\hat{A}\hat{C}W} .$$

For $0 \leq \lambda \leq 1$, define the isometry $U : AC \to \hat{A}\hat{C}W RR'$ by letting, for all x,

$$U|\psi_x\rangle^A |\sigma_x\rangle^C := \sqrt{\lambda} |\xi_x\rangle^{\hat{A}\hat{C}W} |00\rangle^{RR'} + \sqrt{1-\lambda} |\zeta_x\rangle^{\hat{A}\hat{C}W} |11\rangle^{RR'} ,$$

where systems R and R' are qubits. Then, the reduced state on the systems $X\hat{A}\hat{C}$ is $\tau^{X\hat{A}\hat{C}} = \sum_x p(x)|x\rangle\langle x|^X \otimes \tau_x^{\hat{A}\hat{C}}$, where $\tau_x^{\hat{A}\hat{C}} = \lambda \xi_x^{\hat{A}\hat{C}} + (1-\lambda)\zeta_x^{\hat{A}\hat{C}}$; therefore, the fidelity is bounded as follows:

$$\begin{aligned}
F(\omega^{XA\hat{C}}, \tau^{X\hat{A}\hat{C}}) &= \sum_x p(x) \sqrt{\langle\psi_x| \left(\lambda \xi_x^{\hat{A}\hat{C}} + (1-\lambda)\zeta_x^{\hat{A}\hat{C}}\right) |\psi_x\rangle} \\
&\geq \lambda \sum_x p(x) \sqrt{\langle\psi_x| \xi_x^{\hat{A}\hat{C}} |\psi_x\rangle} + (1-\lambda) \sum_x p(x) \sqrt{\langle\psi_x| \zeta_x^{\hat{A}\hat{C}} |\psi_x\rangle} \\
&\geq 1 - \left(\lambda \epsilon_1 + (1-\lambda)\epsilon_2\right),
\end{aligned}$$

where the second line follows from the concavity of the function $\sqrt{x}$, and the last line follows by the definition of the isometries V_1 and V_2. Now, define $W' := WRR'$ and let $\epsilon = \lambda \epsilon_1 + (1-\lambda)\epsilon_2$. According to Definition 4.2, we obtain

$$\begin{aligned}
I_\epsilon(\omega) &\geq I(X : \hat{C}W')_\tau \\
&= I(X : R)_\tau + I(X : \hat{C}W|R)_\tau + I(X : R'|\hat{C}WR)_\tau \\
&\geq I(X : \hat{C}W|R)_\tau = \lambda I_{\epsilon_1}(\omega) + (1-\lambda)I_{\epsilon_2}(\omega),
\end{aligned}$$

where the third line is due to strong subadditivity of the quantum mutual information.

3. The function is non-decreasing and concave for $\epsilon \geq 0$, so it is continuous for $\epsilon > 0$. The concavity implies furthermore that I_ϵ is lower semi-continuous at $\epsilon = 0$. On the other hand, since the fidelity and mutual information are both continuous functions of CPTP maps, and the domain of the optimization is a compact set, we conclude that $I_\epsilon(\omega)$ is also upper semi-continuous at $\epsilon = 0$, so it is continuous at $\epsilon = 0$ [74, Thms. 10.1, 10.2].

4. In the definition of $I_\epsilon(\omega_1 \otimes \omega_2)$, let the isometry $V_0 : A_1 C_1 A_2 C_2 \rightarrow \hat{A}_1 \hat{C}_1 \hat{A}_2 \hat{C}_2 W$ be the one attaining the maximum which acts on the purified source state with purifying systems X_1' and X_2' as follows:

$$
\begin{aligned}
&|\xi\rangle^{X_1 X_1' X_2 X_2' \hat{A}_1 \hat{C}_1 \hat{A}_2 \hat{C}_2 W} \\
&\qquad = \left(\mathbb{1}_{X_1 X_1' X_2 X_2'} \otimes V_0 \right) |\omega_1\rangle^{X_1 X_1' A_1 C_1} |\omega_1\rangle^{X_2 X_2' A_2 C_2}.
\end{aligned}
$$

Now, define the isometry $V_1 : A_1 C_1 \rightarrow \hat{A}_1 \hat{C}_1 \hat{A}_2 \hat{C}_2 W X_2 X_2'$ acting only on the systems $A_1 C_1$ with the output state $\hat{A}_1 \hat{C}_1$ and the environment $W_1 := \hat{A}_2 \hat{C}_2 W X_2 X_2'$ as follows:

$$
|\xi\rangle^{X_1 X_1' X_2 X_2' \hat{A}_1 \hat{C}_1 \hat{A}_2 \hat{C}_2 W} = \left(\mathbb{1}_{X_1 X_1'} \otimes V_1 \right) |\omega_1\rangle^{X_1 X_1' A_1 C_1}.
$$

Hence, we obtain

$$
\begin{aligned}
F\left(\omega_1^{X_1 A_1 C_1}, \xi^{X_1 \hat{A}_1 \hat{C}_1} \right) &\geq F\left(\omega_1^{X_1 A_1 C_1} \otimes \omega_2^{X_2 A_2 C_2}, \xi^{X_1 X_2 \hat{A}_1 \hat{C}_1 \hat{A}_2 \hat{C}_2} \right) \\
&\geq 1 - \epsilon,
\end{aligned}
$$

where the first inequality is due to monotonicity of the fidelity under CPTP maps, and the second inequality follows by the definition of V_0. Consider the isometry $V_2 : A_2 C_2 \rightarrow \hat{A}_1 \hat{C}_1 \hat{A}_2 \hat{C}_2 W X_1 X_1'$ defined in a similar way, with the output state $\hat{A}_2 \hat{C}_2$ and the environment $W_2 := \hat{A}_1 \hat{C}_1 W X_1 X_1'$. Therefore, we obtain

$$
\begin{aligned}
I_\epsilon(\omega_1) + I_\epsilon(\omega_2) &\geq I(X_1 : \hat{C}_1 W_1) + I(X_2 : \hat{C}_2 W_2) \\
&\geq I(X_1 : \hat{C}_1 \hat{C}_2 W) + I(X_2 : \hat{C}_1 \hat{C}_2 W X_1) \\
&= I(X_1 X_2 : \hat{C}_1 \hat{C}_2 W) = I_\epsilon(\omega_1 \otimes \omega_2),
\end{aligned}
$$

where the second line is due to data processing.

5. In the definition of $I_0(\omega)$ let $V_0 : AC \rightarrow \hat{A}\hat{C}W$ be the isometry attaining

the maximum with $F(\omega^{XAC}, \xi^{X\hat{A}\hat{C}}) = 1$. Hence, we obtain

$$
\begin{aligned}
I_0(\omega) &= I(X : \hat{C}W) = I(XY : \hat{C}W) \\
&= I(Y : \hat{C}W) + I(X : \hat{C}W|Y) \\
&\leq S(Y) + I(X : \hat{C}W|Y) \\
&= S(Y) + I(X : W|Y) + I(X : \hat{C}|WY) \\
&\leq S(Y) + I(X : W|Y) + S(C|WY) \\
&\leq S(Y) + I(X : W|Y) + S(C|Y),
\end{aligned}
$$

where the first line follows because Y is a function of X. The second and fourth line are due to the chain rule. The third line follows because for the classical system Y the conditional entropy $S(Y|\hat{C}W)$ is non-negative. The penultimate line follows because for any x the state on the system $\hat{C}$ is pure. The last line is due to strong sub-additivity of the entropy. Furthermore, for every y, the ensemble $\mathcal{E}_y$ is irreducible; hence, the conditional mutual information $I(X : W|Y) = 0$ which follows from the detailed discussion on page 2028 of [65]. ∎

Proof of the converse part of Theorem 4.1. We start by observing

$$
nQ + S(B_0) \geq S(C_A) + S(B_0) \geq S(C_A B_0) = S(\hat{A}^n W_B),
$$

where the second inequality is due to subadditivity of the entropy, and the equality follows because the decoding isometry V_B does not change the entropy. Hence, we get

$$
\begin{aligned}
nQ + S(B_0) &\geq S(\hat{A}^n) + S(W_B|\hat{A}^n) \\
&\geq S(\hat{A}^n) + S(W_B|\hat{A}^n X^n) \\
&\geq S(A^n) + S(W_B|\hat{A}^n X^n) - n\delta(n, \epsilon) \\
&= S(A^n) + S(\hat{A}^n W_B|X^n) - S(\hat{A}^n|X^n) - n\delta(n, \epsilon) \\
&= S(A^n) + S(\hat{C}^n W_A|X^n) - S(\hat{A}^n|X^n) - n\delta(n, \epsilon) \\
&\geq S(A^n) + S(\hat{C}^n W_A|X^n) - 3n\delta(n, \epsilon), \tag{4.4}
\end{aligned}
$$

where in the first and second line we use the chain rule and subadditivity of entropy. The inequality in the third line follows from the decodability of the system A^n: the fidelity criterion (4.2) implies that the output state on systems $\hat{A}^n$ is $2\sqrt{2\epsilon}$-close to the original state A^n in trace norm; then apply the Fannes-Audenaert inequality [72, 73] where $\delta(n, \epsilon) = \sqrt{2\epsilon} \log |A| + \frac{1}{n}h(\sqrt{2\epsilon})$. The equalities in the fourth and the fifth line are due to the chain

52

rule and the fact that for any x^n the overall state of $\hat{A}^n\hat{C}^nW_AW_B$ is pure. In the last line, we use the decodability of the systems X^nA^n, that is the output state on systems $X^n\hat{A}^n$ is $2\sqrt{2\epsilon}$-close to the original states X^nA^n in trace norm, then we apply the Alicki-Fannes inequality [76, 77].

Moreover, we bound Q as follows:

$$
\begin{aligned}
nQ \geq S(C_A) &\geq S(C_A|\hat{C}^nW_A) \\
&= S(A^nC^nA_0) - S(\hat{C}^nW_A) \\
&= S(A^nC^nY^n) + S(A_0) - S(\hat{C}^nW_A),
\end{aligned}
\tag{4.5}
$$

where the first equality follows because the encoding isometry $V_A : A^nC^nA_0 \to C_A\hat{C}^nW_A$ does not the change the entropy. Adding Eqs. (4.4) and (4.5), we thus obtain

$$
\begin{aligned}
Q &\geq \frac{1}{2}(S(A) + S(ACY)) - \frac{1}{2n}I(\hat{C}^nW_A : X^n) - \frac{3}{2}\delta(n,\epsilon) \\
&\geq \frac{1}{2}(S(A) + S(ACY)) - \frac{1}{2n}I(\hat{C}^nW_AW_B : X^n) - \frac{3}{2}\delta(n,\epsilon) \\
&\geq \frac{1}{2}(S(A) + S(ACY)) - \frac{1}{2n}I_\epsilon(\omega^{\otimes n}) - \frac{3}{2}\delta(n,\epsilon) \\
&\geq \frac{1}{2}(S(A) + S(ACY)) - \frac{1}{2}I_\epsilon(\omega) - \frac{3}{2}\delta(n,\epsilon)
\end{aligned}
$$

where the second line is due to data processing. The third line follows from Definition 4.2. The last line follows from point 4 of Lemma 4.1. In the limit of $\epsilon \to 0$ and $n \to \infty$, the rate is bounded by

$$
\begin{aligned}
Q &\geq \frac{1}{2}(S(A) + S(ACY)) - \frac{1}{2}I_0(\omega) \\
&\geq \frac{1}{2}(S(A) + S(ACY)) - \frac{1}{2}S(CY) \\
&= \frac{1}{2}(S(A) + S(A|CY)),
\end{aligned}
$$

where the first line follows from point 3 of Lemma 4.1 stating that $I_\epsilon(\omega)$ is continuous at $\epsilon = 0$. The second line is due to point 5 of Lemma 4.1. ∎

4.4   Complete rate region

In this section, we find the complete rate region of achievable rate pairs (E, Q).
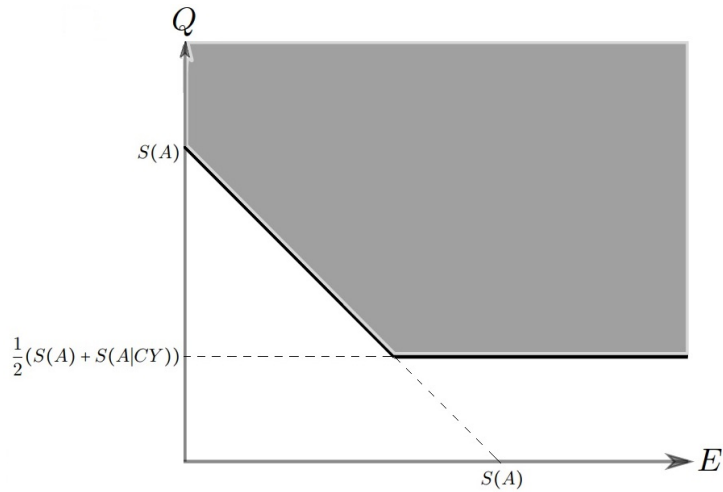
Figure 4.1: The optimal rate region of quantum and entanglement rates.

Theorem 4.2. *For the source ω^{XACY}, all asymptotically achievable entanglement and quantum rate pairs (E, Q) satisfy*

$$Q \geq \frac{1}{2}(S(A) + S(A|CY)),$$
$$Q + E \geq S(A).$$

Conversely, all the rate pairs satisfying the above inequalities are achievable.

Proof. The first inequality comes from Theorem 4.1. For the second inequality, consider any code with quantum communication rate R and entanglement rate E. By using an additional communication rate E, Alice and Bob can distribute the entanglement first, and then apply the given code, converting it into one without preshared entanglement and communication rate $Q + E$, having exactly the same fidelity. By Remark 4.1, $Q + E \geq S(A)$.

As for the achievability, the corner point $(\frac{1}{2}I(A:CY), \frac{1}{2}(S(A) + S(A|CY)))$ is achievable, because QSR which is used as the achievability protocol in Theorem 4.1 uses $\frac{1}{2}I(A:CY)$ ebits of entanglement between Alice and Bob. Furthermore, all the points on the line $Q + E = S(A)$ for $Q \geq \frac{1}{2}(S(A) + S(A|CY))$ are achievable because one ebit can be distributed by sending a qubit. All other rate pairs are achievable by resource wasting. The rate region is depicted in Fig. 4.1 ∎

54

4.5 Discussion

First of all, let us look what our result tell us in the cases of blind and visible compression.

Corollary 4.1. *In blind compression (i.e. if C is trivial, or more generally the states σ_x are all identical), the compression of the source ω^{XACY} reduces to the entanglement-assisted Schumacher compression for which Theorem 4.1 gives the optimal asymptotic quantum rate*

$$Q = \frac{1}{2}(S(A) + S(A|Y)) = S(A) - \frac{1}{2}S(Y).$$

This implies that if the source is irreducible, then this rate is equal to the Schumacher limit $S(A)$. In other words, the entanglement does not help the compression. Moreover, due to Theorem 4.2, a rate $\frac{1}{2}S(Y)$ of entanglement is consumed in the compression, and $E + Q \geq S(A)$ in general. ∎

The blind compression of a source ω^{XAY} is also considered in [65], but there instead of entanglement, a noiseless classical channel was assumed in addition to the quantum channel. It was shown that the optimal quantum rate assisted with free classical communication is equal to $S(A) - S(Y)$, while a rate $S(Y)$ of classical communication suffices. By sending the classical information using dense coding [71], spending $\frac{1}{2}$ ebit and $\frac{1}{2}$ qubit per cbit, we can recover the quantum and entanglement rates of Corollary 4.1. This means that our converse implies the optimality of the quantum rate from [65].

Thus we are motivated to look at a modified compression model where the resources used are classical communication and entanglement. Namely, we let Alice and Bob share entanglement at rate E and use classical communication at rate C, but otherwise the objective is the same as in Section 4.2; define the rate region as the set of all asymptotic achievable classical communication and entanglement rate pairs (C, E), such that the decoding fidelity asymptotically converges to 1.

Theorem 4.3. *For a source ω^{XAY}, a rate pair (C, E) is achievable if and only if*

$$C \geq 2S(A) - S(Y), \ \ E \geq S(A) - S(Y).$$

Proof. We start with the converse. The first inequality follows from Theorem 4.1, because with unlimited entanglement shared between Alice and Bob, $\frac{1}{2}(S(A) + S(A|Y)) = S(A) - \frac{1}{2}S(Y)$ qubits of quantum communication is equivalent to $2S(A) - S(Y)$ bits of classical communication due to teleportation [78] and dense coding [71]. The second inequality follows from [65], because with free classical communication, the quantum rate is lower bounded
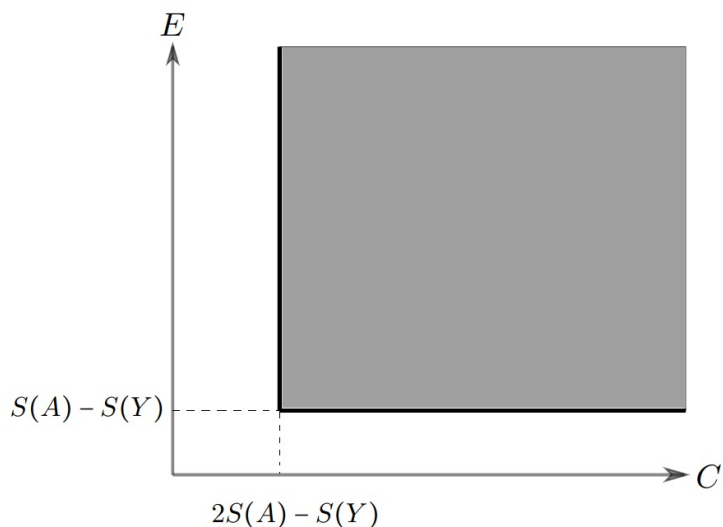
Figure 4.2: The optimal rate region of classical and entanglement rates.

by $S(A) - S(Y)$ which, due to teleportation [78], is equivalent to sharing $S(A) - S(Y)$ ebits when classical communication is for free.

The achievability of the corner point $(2S(A)-S(Y), S(A)-S(Y))$ follows from [65] because the compression protocol uses $S(A)-S(Y)$ qubits and $S(Y)$ bits of classical communication which is equivalent to using $S(A)-S(Y)$ ebits of entanglement and $2S(A) - 2S(Y) + S(Y)$ bits of classical communication, due to dense coding [71]. Other rate pairs are achievable by resource wasting. The rate region is depicted in Fig. 4.2. ∎

Corollary 4.1. *In the visible case, our compression problem reduces to the visible version of Schumacher compression with entanglement assistance. In this case, according to Theorem 4.1 the optimal asymptotic quantum rate is $Q = \frac{1}{2}S(A)$. Moreover, a rate $E = \frac{1}{2}S(A)$ of entanglement is consumed in the compression scheme, and $E + Q \geq S(A)$ in general.* ∎

We remark that the visible compression assisted by unlimited entanglement is also a special case of remote state preparation considered in [66], from which we know that the rate $Q = \frac{1}{2}S(A)$ is achievable and optimal.

The visible analogue of [65], of compression using qubit and cbit resources, was treated in [67], where the achievable region was determined as the union of all all pairs (C, Q) such that $Q \geq S(A|Z)$ and $C \geq I(X:Z)$, for any random variable Z forming a Markov chain Z—X—A. Compare to the complicated boundary of this region the much simpler one of Corollary 4.1, which consists of two straight lines.

We close by discussing several open questions for future work: First, the final discussion of different pairs of resources to compress suggests that an interesting target would be the characterisation of the full triple resource tradeoff region for Q, C and E together.

Secondly, we recall that our definition of successful decoding included preservation of the side information σ_x^C with high fidelity. What is the optimal compression rate Q if the side information does not have to be preserved? For an example where this change has a dramatic effect on the optimal communication rate, consider the ensemble $\mathcal{E}$ consisting of the three two-qubit states $|0\rangle^A |0\rangle^C$, $|1\rangle^A |0\rangle^C$ and $|+\rangle^A |+\rangle^C$ (where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$), with probabilities $\frac{1}{2} - t$, $\frac{1}{2} - t$ and $2t$, respectively. Note that $\mathcal{E}$ is irreducible, hence for $t \approx 0$, we get an optimal quantum rate of $Q \approx 1$, because $S(A) \approx S(A|C) \approx 1$. However, by applying a CNOT unitary (with A as control and C as target), the ensemble is transformed into $\mathcal{E}'$ consisting of the states $|0\rangle^A |0\rangle^{C'}$, $|1\rangle^A |1\rangle^{C'}$ and $|+\rangle^A |+\rangle^{C'}$. The state of A is not changed, only the side information, which is why we denote it C'. Hence we can apply Theorem 4.1 to get a quantum rate $Q \approx \frac{1}{2}$, because $S(A) \approx 1$, $S(A|C) \approx 0$.

Thirdly, note that the lower bound $Q + E \geq S(A)$ in Theorem 4.2 holds with a strong converse (see the proof and [30]). But does $Q \geq \frac{1}{2}(S(A) + S(A|CY))$ hold as a strong converse rate with unlimited entanglement? Likewise, in the setting of [65] with unlimited classical communication, is $Q \geq S(A) - S(Y)$ a strong converse bound for the quantum rate?

Chapter 5

Distributed compression of correlated classical-quantum sources

In this chapter, we resume the investigation of the problem of independent local compression of correlated quantum sources, the classical case of which is covered by the celebrated Slepian-Wolf theorem. We focus specifically on classical-quantum (cq) sources, for which one edge of the rate region, corresponding to the compression of the classical part, using the quantum part as side information at the decoder, was previously determined by Devetak and Winter [Phys. Rev. A 68, 042301 (2003)]. Whereas the Devetak-Winter protocol attains a rate-sum equal to the von Neumann entropy of the joint source, here we show that the full rate region is much more complex, due to the partially quantum nature of the source. In particular, in the opposite case of compressing the quantum part of the source, using the classical part as side information at the decoder, typically the rate sum is strictly larger than the von Neumann entropy of the total source.

We determine the full rate region in the *generic* case, showing that, apart from the Devetak-Winter point, all other points in the achievable region have a rate sum strictly larger than the joint entropy. We can interpret the difference as the price paid for the quantum encoder being ignorant of the classical side information. In the general case, we give an achievable rate region, via protocols that are built on the decoupling principle, and the protocols of quantum state merging and quantum state redistribution. Our achievable region is matched almost by a single-letter converse, which however still involves asymptotic errors and an unbounded auxiliary system. This chapter is based on the papers in [49, 50].

5.1 The source and the compression model

The Slepian-Wolf problem of two sources correlated in a known way, but subject to separate, local compression [43] has proved to provide a unifying principle for much of Shannon theory, giving rise to natural information theoretic interpretations of entropy and conditional entropy, and exhibiting deep connections with error correction, channel capacities and mutual information (cf. [79]). The quantum case has been investigated for two decades, starting with the second author's PhD thesis [30] and subsequently in [29], up to the systematic study [32], and while we still do not have a complete understanding of the rate region, it has become clear that the problem is of much higher complexity than the classical case. The quantum Slepian-Wolf problem, and specifically quantum data compression with side information at the decoder, has resulted in many fundamental advances in quantum information theory, including the protocols of quantum state merging [23,68] and quantum state redistribution [25], which have given operational meaning to the conditional von Neumann entropy, the mutual information and the conditional quantum mutual information, respectively.

A variety of resource models and different tasks have been considered over the years: The source and its recovery was either modelled as an ensemble of pure states (following Schumacher [8]), or as a pure state between the encoders and a reference system; the communication resource required was either counted in qubits communicated, in addition either allowing or disallowing entanglement, or it was counted in ebits shared between the agents, but with free classical communication. While this latter model has led to the most complete picture of the general rate region, in the present chapter we will go back to the original idea [8,30] of quantifying the communication, counted in qubits, between the encoders and the decoder.

Source model. The source model we shall consider is a hybrid classical-quantum one, with two agents, Alice and Bob, whose task is is to compress the classical and quantum parts of the source, respectively. They then send their shares to a decoder, Debbie, who has to reconstruct the classical information with high probability and the quantum information with high (average) fidelity.

In detail, the source is characterised by a classical source, i.e. a probability distribution $p(x)$ on a discrete (in fact: finite) alphabet $\mathcal{X}$ which is observed by Alice, and a family of quantum states ρ_x on a quantum system B, given by a Hilbert space of finite dimension $|B|$. To define the problem of independent local compression (and decompression) of such a correlated classical-quantum source, we shall consider purifications ψ_x^{BR} of the ρ_x, i.e. $\rho_x^B = \operatorname{Tr}_R \psi_x^{RB}$. Thus

the source can be described compactly by the cq-state

$$\omega^{XBR} = \sum_{x \in \mathcal{X}} p(x) \, |x\rangle\langle x|^X \otimes |\psi_x\rangle\langle\psi_x|^{BR} \, .$$

We will be interested in the information theoretic limit of many copies of ω, i.e.

$$\begin{aligned}
\omega^{X^n B^n R^n} &= \left(\omega^{XBR}\right)^{\otimes n} \\
&= \sum_{x^n \in \mathcal{X}^n} p(x^n) \, |x^n\rangle\langle x^n|^{X^n} \otimes |\psi_{x^n}\rangle\langle\psi_{x^n}|^{B^n R^n},
\end{aligned}$$

where we use the notation

$$\begin{aligned}
x^n &= x_1 x_2 \ldots x_n, \\
|x^n\rangle &= |x_1\rangle |x_2\rangle \cdots |x_n\rangle , \\
p(x^n) &= p(x_1)p(x_2)\cdots p(x_n), \text{ and} \\
|\psi_{x^n}\rangle &= |\psi_{x_1}\rangle |\psi_{x_2}\rangle \cdots |\psi_{x_n}\rangle .
\end{aligned}$$

Alice and Bob, receiving their respective parts of the source, separately encode these using the most general allowed quantum operations; the compressed quantum information, living on a certain number of qubits, is passed to the decoder who has to output, again acting with a quantum operation, an element of $\mathcal{X}^n$ and a state on B^n, in such a way as to attain a low error probability for x^n and a high-fidelity approximation of the conditional quantum source state, $\psi_{x^n}^{B^n R^n}$. We consider two models: unassisted and entanglement-assisted, which we describe formally in the following (see Figs. 5.1 and 5.2).

Unassisted model. With probability $p(x^n)$, the source provides Alice and Bob respectively with states $|x^n\rangle^{X^n}$ and $|\psi_{x^n}\rangle^{B^n R^n}$. Alice and Bob then perform their respective encoding operations $\mathcal{E}_X : X^n \longrightarrow C_X$ and $\mathcal{E}_B : B^n \longrightarrow C_B$, respectively, which are quantum operations, i.e. completely positive and trace preserving (CPTP) maps. Of course, as functions they act on the operators (density matrices) over the respective input and output Hilbert spaces. But as there is no risk of confusion, we will simply write the Hilbert spaces when denoting a CPTP map. Note that since X is a classical random variable, $\mathcal{E}_X$ is entirely described by a cq-channel. We call $R_X = \frac{1}{n} \log |C_X|$ and $R_B = \frac{1}{n} \log |C_B|$ the quantum rates of the compression protocol. Since Alice and Bob are required to act independently, the joint encoding operation is $\mathcal{E}_X \otimes \mathcal{E}_B$. The systems C_X and C_B are then sent to Debbie who performs a decoding operation $\mathcal{D} : C_X C_B \longrightarrow \hat{X}^n \hat{B}^n$. $\hat{X}^n$ and $\hat{B}^n$ are output systems with Hilbert spaces $\hat{X}^n$ and $\hat{B}^n$ which are isomorphic to Hilbert spaces X^n

and B^n, respectively. We define the extended source state

$$
\omega^{X^n X'^n B^n R^n}
$$
$$
= \left(\omega^{XX'BR} \right)^{\otimes n}
$$
$$
= \sum_{x^n \in \mathcal{X}^n} p(x^n) |x^n\rangle\langle x^n|^{X^n} \otimes |x^n\rangle\langle x^n|^{X'^n} \otimes |\psi_{x^n}\rangle\langle\psi_{x^n}|^{B^n R^n}, \tag{5.1}
$$

and say the encoding-decoding scheme has average fidelity $1 - \epsilon$ if

$$
\overline{F} \coloneqq F\left(\omega^{X^n X'^n B^n R^n}, \xi^{\hat{X}^n X'^n \hat{B}^n R^n} \right) \ge 1 - \epsilon, \tag{5.2}
$$

where

$$
\xi^{\hat{X}^n X'^n \hat{B}^n R^n} = \left(\mathcal{D} \circ \left(\mathcal{E}_X \otimes \mathcal{E}_B \right) \otimes \mathrm{id}_{X'^n R^n} \right) \omega^{X^n X'^n B^n R^n},
$$

and $\mathrm{id}_{X'^n R^n}$ is the identity (ideal) channel acting on $X'^n R^n$. By the above fidelity definition and the linearity of CPTP maps, the average fidelity defined in (5.2) can be expressed equivalently as

$$
\overline{F} = \sum_{x^n \in \mathcal{X}^n} p(x^n) F\left(|x^n\rangle\langle x^n|^{X^n} \otimes |\psi_{x^n}\rangle\langle\psi_{x^n}|^{B^n R^n}, \xi_{x^n}^{\hat{X}^n \hat{B}^n R^n} \right)
$$

where

$$
\xi_{x^n}^{\hat{X}^n \hat{B}^n R^n} =
$$
$$
\left(\mathcal{D} \circ \left(\mathcal{E}_X \otimes \mathcal{E}_B \right) \otimes \mathrm{id}_{R^n} \right) |x^n\rangle\langle x^n|^{X^n} \otimes |\psi_{x^n}\rangle\langle\psi_{x^n}|^{B^n R^n}.
$$

We say that (R_X, R_B) is an (asymptotically) achievable rate pair if there exist codes $(\mathcal{E}_X, \mathcal{E}_B, \mathcal{D})$ as above for every n, with fidelity $\overline{F}$ converging to 1, and classical and quantum rates converging to R_X and R_B, respectively. The rate region is the set of all achievable rate pairs, as a subset of $\mathbb{R}^2_{\ge 0}$.

It is shown by Devetak and Winter in [29, Theorem 1] and [30, Corollary IV.13] that the rate pair

$$
(R_X, R_B) = (S(X|B), S(B)) \tag{5.3}
$$

is achievable and optimal. The optimality is two-fold; first, the rate sum achieved, $R_X + R_B = S(XB)$ is minimal, and secondly, even with unlimited R_B, $R_X \ge S(X|B)$. This shows that the Devetak-Winter point is an extreme point of the rate region. Interestingly, Alice can achieve the rate $S(X|B)$ using only classical communication. However, we will prove the converse theorems considering a quantum channel for Alice, which are obviously stronger statements. In Theorem 5.8, we show that our system model

Figure 5.1: Circuit diagram of the unassisted model. Dotted lines are used to demarcate domains controlled by the different participants. The solid lines represent quantum information registers.

is equivalent to the model considered in [29, 30], which implies the achievability and optimality of this rate pair in our system model. We remark that in [29], the rate $R_B = S(B)$ was not explicitly discussed, but it is clear that it can always be achieved by Schumacher's quantum data compression [8], introducing an arbitrarily small additional error.

Entanglement-assisted model. This model generalizes the unassisted model, and it is basically the same, except that we let Bob and Debbie share entanglement and use it in encoding and decoding, respectively. In addition, we take care of any possible entanglement that is produced in the process. Consequently, while Alice's encoding $\mathcal{E}_X : X^n \longrightarrow C_X$ remains the same, the Bob's encoding and the decoding map now act as $\mathcal{E}_B : B^n B_0 \longrightarrow C_B B_0'$ and $\mathcal{D} : C_X C_B D_0 \longrightarrow \hat{X}^n \hat{B}^n D_0'$, respectively, where B_0 and D_0 are K-dimensional quantum registers of Bob and Debbie, respectively, designated to hold the initially shared entangled state, and B_0' and D_0' are L-dimensional registers for the entanglement produced by the protocol. Ideally, both initial and final entanglement are given by maximally entangled states Φ_K and Φ_L, respectively. Correspondingly, we say that the encoding-decoding scheme has average fidelity $1 - \epsilon$ if

$$\overline{F} := F\left(\omega^{X^n X'^n B^n R^n} \otimes \Phi_L^{B_0' D_0'}, \xi^{\hat{X}^n X'^n \hat{B}^n R^n B_0' D_0'}\right)$$
$$\geq 1 - \epsilon, \tag{5.4}$$

63

where

$$\xi^{\hat{X}^n X'^n \hat{B}^n R^n B_0' D_0'} = \left(\mathcal{D} \circ \left(\mathcal{E}_X \otimes \mathcal{E}_{BB_0} \otimes \mathrm{id}_{D_0}\right) \otimes \mathrm{id}_{X'^n R^n}\right)$$
$$\omega^{X^n X'^n B^n R^n} \otimes \Phi_L^{B_0' D_0'}.$$

We call $E = \frac{1}{n}(\log K - \log L)$ the entanglement rate of the scheme. The CPTP map $\mathcal{E}_B$ takes the input systems $B^n B_0$ to the compressed system C_B plus Bob's share of the output entanglement, B_0'. Debbie applies the decoding operation $\mathcal{D}$ on the received systems $C_X C_B$ and her part of the initial entanglement D_0, to produce an output state on systems $\hat{X}^n \hat{B}^n$ plus her share of the output entanglement, D_0'. Similar to the unassisted model, $\hat{X}^n$ and $\hat{B}^n$ are output systems with Hilbert spaces $\hat{X}^n$ and $\hat{B}^n$ which are isomorphic to Hilbert spaces X^n and B^n, respectively. We say (R_X, R_B, E) is an (asymptotically) achievable rate triple if for all n there exist entanglement-assisted codes as before, such that the fidelity $\overline{F}$ converges to 1, and the classical, quantum and entanglement rates converge to R_X, R_B and E, respectively. The rate region is the set of all achievable rate pairs, as a subset of $\mathbb{R}^2_{\geq 0} \times \mathbb{R}$. In the following we will be mostly interested in the projection of this region onto the first two coordinates, R_X and R_B, corresponding to unlimited entanglement assistance.

It is a simple consequence of the time sharing principle that the rate regions, both for the unassisted and the entanglement-assisted model, are closed convex regions. Furthermore, since one can always waste rate, the rate regions are open to the "upper right". This means that the task of characterizing the rate regions boils down to describing the lower boundary, which can be achieved by convex inequalities. In the Slepian-Wolf problem, they are in fact linear inequalities, and we will find analogues of these in the present investigation.

Stinespring's dilation theorem [55] states that any CPTP map can be built from the basic operations of isometry and reduction to a subsystem by tracing out the environment system [55]. Thus, the encoders and the decoder are without loss of generality isometries

$$U_X : X^n \longrightarrow C_X W_X,$$
$$U_B : B^n B_0 \longrightarrow C_B B_0' W_B,$$
$$V : C_X C_B D_0 \longrightarrow \hat{X}^n \hat{B}^n D_0' W_D,$$

where the new systems W_X, W_B and W_D are the environment systems of Alice, Bob and Debbie, respectively. They simply remain locally in possession of the respective party.
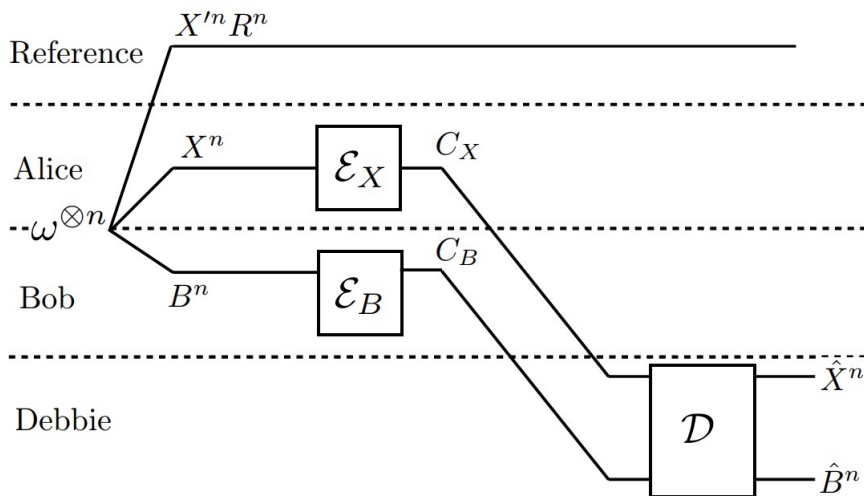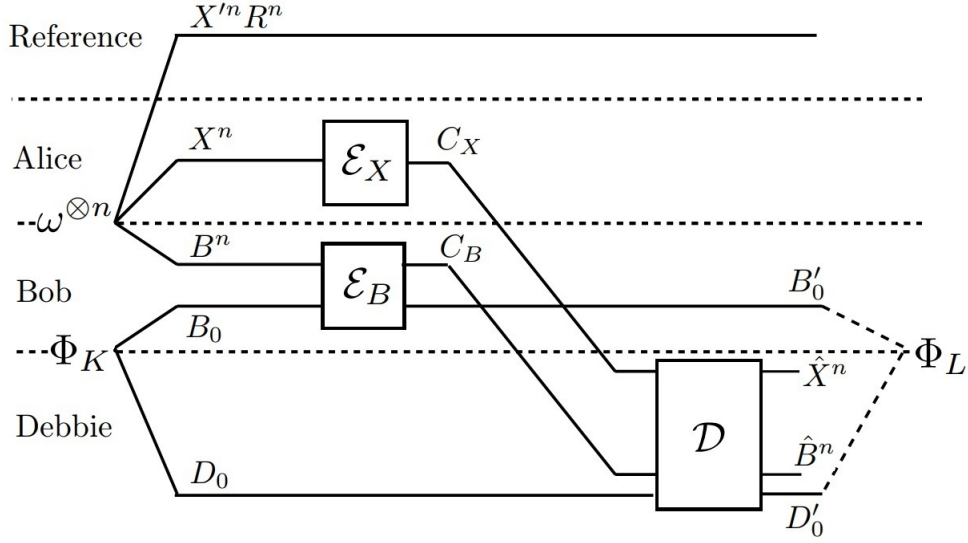
Figure 5.2: Circuit diagram of the entanglement-assisted model. Dotted lines are used to demarcate domains controlled by the different participants. The solid lines represent quantum information registers.

The following lemma states that for a code of block length n and error ϵ, the environment parts of the encoding and decoding isometries, i.e. W_X, W_B and W_D, as well as the entanglement output registers B_0' and D_0', are decoupled from the reference R^n, conditioned on X^n. This lemma plays a crucial role in the proofs of converse theorems.

Lemma 5.1. *(Decoupling condition) For a code of block length n and error ϵ in the entanglement-assisted model, let W_X, W_B and W_D be the environments of Alice's and Bob's encoding and of Debbie's decoding isometries, respectively. Then,*

$$I(W_X W_B W_D B_0' D_0' : \hat{X}^n \hat{B}^n R^n | X'^n)_\xi \le n\delta(n,\epsilon),$$

where $\delta(n,\epsilon) = 4\sqrt{6\epsilon}\log(|X||B|) + \frac{2}{n}h(\sqrt{6\epsilon})$, with the binary entropy $h(\epsilon) = -\epsilon\log\epsilon - (1-\epsilon)\log(1-\epsilon)$; the conditional mutual information is with respect to the state

$$\xi^{X'^n \hat{X}^n \hat{B}^n B_0' D_0' W_X W_B W_D R^n}$$
$$= \left(V \circ (U_X \otimes U_B \otimes \mathbb{1}_{D_0}) \otimes \mathbb{1}_{X'^n R^n}\right)$$
$$\left(\omega^{X^n X'^n B^n R^n} \otimes \Phi_K^{B_0 D_0}\right)$$
$$\left(V \circ (U_X \otimes U_B \otimes \mathbb{1}_{D_0}) \otimes \mathbb{1}_{X'^n R^n}\right)^\dagger.$$

65

Proof. We show that the fidelity criterion (5.4) implies that given x^n, the environments W_X, W_B and W_D of Alice's, Bob's and Debbie's isometries are decoupled from the the rest of the output systems.

The parties share n copies of the state $\omega^{X'XBR}$, where Alice and Bob have access to systems X^n and B^n, respectively, and X'^n and R^n are the reference systems. Alice and Bob apply the following isometries to encode their systems, respectively:

$$U_X : X^n \longrightarrow C_X W_X,$$
$$U_B : B^n B_0 \longrightarrow C_B B_0' W_B,$$

where Alice and Bob send respectively their compressed information C_X and C_B to Debbie and keep the environment parts W_X and W_B of their respective isometries for themselves. Debbie applies the decoding isometry $V : C_X C_B D_0 \longrightarrow \hat{X}^n \hat{B}^n D_0' W_D$ to the systems $C_X C_B$ and her part of the entanglement D_0, to generate the output systems $\hat{X}^n \hat{B}^n D_0'$, with W_D the environment of her isometry. This leads to the following final state after decoding:

$$\xi^{X'^n \hat{X}^n \hat{B}^n B_0' D_0' W_X W_B W_D R^n}$$
$$= \sum_{x^n} p(x^n) |x^n\rangle\langle x^n|^{X'^n} \otimes |\xi_{x^n}\rangle\langle \xi_{x^n}|^{\hat{X}^n \hat{B}^n B_0' D_0' W_X W_B W_D R^n},$$

where

$$|\xi_{x^n}\rangle^{\hat{X}^n \hat{B}^n B_0' D_0' W_X W_B W_D R^n}$$
$$= V^{C_X C_B D_0 \to \hat{X}^n \hat{B}^n D_0' W_D}$$
$$\left(U_X^{X^n \to C_X W_X} |x^n\rangle^{X^n} \otimes U_B^{B^n B_0 \to C_B B_0' W_B} \right.$$
$$\left. \left(|\psi_{x^n}\rangle^{B^n R^n} |\Phi_K\rangle^{B_0 D_0} \right) \right).$$

The fidelity defined in Eq. (5.4) is now bounded as follows:

$$\overline{F} = F\left(\omega^{X'^n X^n B^n R^n} \otimes \Phi_L^{B_0' D_0'}, \xi^{X'^n \hat{X}^n \hat{B}^n B_0' D_0' R^n} \right)$$
$$\leq F\left(\omega^{X'^n X^n B^n R^n}, \xi^{X'^n \hat{X}^n \hat{B}^n R^n} \right)$$
$$= \sum_{x^n \in \mathcal{X}^n} p(x^n) F\left(|x^n\rangle\langle x^n|^{X^n} \otimes |\psi_{x^n}\rangle\langle \psi_{x^n}|^{B^n R^n}, \xi_{x^n}^{\hat{X}^n \hat{B}^n R^n} \right)$$
$$= \sum_{x^n} p(x^n) \sqrt{\langle x^n| \langle \psi_{x^n}|^{B^n R^n} \xi_{x^n}^{\hat{X}^n \hat{B}^n R^n} |x^n\rangle |\psi_{x^n}\rangle^{B^n R^n}}$$
$$\leq \sum_{x^n} p(x^n) \sqrt{\|\xi_{x^n}^{\hat{X}^n \hat{B}^n R^n}\|}, \tag{5.5}$$

where in the first line $\xi^{X'^n \hat{X}^n \hat{B}^n B_0' D_0' R^n} = \left(\mathcal{D} \circ \left(\mathrm{id}_{X^n D_0} \otimes \mathcal{E}_B \right) \otimes \mathrm{id}_{X'^n R^n} \right) \omega^{X^n X'^n B^n R^n} \otimes \Phi_K^{B_0 D_0}$. The inequality in the second line is due to the monotonicity of fidelity under partial trace, and $\| \xi_{x^n}^{\hat{X}^n \hat{B}^n R^n} \|$ denotes the operator norm, which in this case of a positive semidefinite operator is the maximum eigenvalue of $\xi_{x^n}^{\hat{X}^n \hat{B}^n R^n}$. Now, consider the Schmidt decomposition of the state $|\xi_{x^n}\rangle^{\hat{X}^n \hat{B}^n B_0' D_0' W_X W_B W_D R^n}$ with respect to the partition $\hat{X}^n \hat{B}^n R^n : B_0' D_0' W_X W_B W_D$, i.e.

$$|\xi_{x^n}\rangle^{\hat{X}^n \hat{B}^n B_0' D_0' W_X W_B W_D R^n}$$
$$= \sum_i \sqrt{\lambda_{x^n}(i)} \, |v_{x^n}(i)\rangle^{\hat{X}^n \hat{B}^n R^n} |w_{x^n}(i)\rangle^{B_0' D_0' W_X W_B W_D}.$$

High average fidelity $\overline{F} \geq 1 - \epsilon$ implies that *on average* the above states are approximately product states. In other words, the two subsystems are nearly decoupled on average:

$$\sum_{x^n} p(x^n) F\left(|\xi_{x^n}\rangle\langle\xi_{x^n}| \,, \xi_{x^n}^{\hat{X}^n \hat{B}^n R^n} \otimes \xi_{x^n}^{B_0' D_0' W_X W_B W_D} \right)$$
$$= \sum_{x^n} p(x^n) \sqrt{\langle \xi_{x^n} | \xi_{x^n}^{\hat{X}^n \hat{B}^n R^n} \otimes \xi_{x^n}^{B_0' D_0' W_X W_B W_D} | \xi_{x^n} \rangle}$$
$$= \sum_{x^n} p(x^n) \sum_i \lambda_{x^n}(i)^{\frac{3}{2}}$$
$$\geq \sum_{x^n} p(x^n) \| \xi_{x^n}^{\hat{X}^n \hat{B}^n R^n} \|^{\frac{3}{2}}$$
$$\geq \left(\sum_{x^n} p(x^n) \sqrt{\| \xi_{x^n}^{\hat{X}^n \hat{B}^n R^n} \|} \right)^3$$
$$\geq (1 - \epsilon)^3 \geq 1 - 3\epsilon, \tag{5.6}$$

where in the first line $|\xi_{x^n}\rangle\langle\xi_{x^n}|$ is a state on systems $\hat{X}^n \hat{B}^n B_0' D_0' W_X W_B W_D R^n$. The inequality in the fifth line follows from the convexity of x^3 for $x \geq 0$, and in the sixth line we have used Eq. (5.5). Based on the relation between fidelity and trace distance (Lemma A.6), we thus obtain for the product ensemble

$$\zeta^{X'^n \hat{X}^n \hat{B}^n B_0' D_0' W_X W_B W_D R^n}$$
$$:= \sum_{x^n} p(x^n) |x^n\rangle\langle x^n|^{X'^n} \otimes \xi_{x^n}^{\hat{X}^n \hat{B}^n R^n} \otimes \xi_{x^n}^{B_0' D_0' W_X W_B W_D},$$

67

that

$$\|\xi - \zeta\|_1$$
$$= \sum_{x^n} p(x^n)$$
$$\left\| |\xi_{x^n}\rangle\langle\xi_{x^n}|^{\hat{X}^n\hat{B}^n B_0' D_0' W_X W_B W_D R^n} - \xi_{x^n}^{\hat{X}^n\hat{B}^n R^n} \otimes \xi_{x^n}^{B_0' D_0' W_X W_B W_D} \right\|_1$$
$$\leq 2\sqrt{6\epsilon}.$$

By the Alicki-Fannes inequality (Lemma A.11), this implies

$$I(\hat{X}^n\hat{B}^n R^n : B_0' D_0' W_X W_B W_D | X'^n)_\xi$$
$$= S(\hat{X}^n\hat{B}^n R^n | X'^n)_\xi$$
$$\qquad - S(\hat{X}^n\hat{B}^n R^n | X'^n B_0' D_0' W_X W_B W_D)_\xi$$
$$\leq 2\sqrt{6\epsilon}\log(|X|^n|B|^n|R|^n) + 2h(\sqrt{6\epsilon})$$
$$\leq 2\sqrt{6\epsilon}\log(|X|^{2n}|B|^{2n}) + 2h(\sqrt{6\epsilon})$$
$$=: n\delta(n,\epsilon), \tag{5.7}$$

where we note in the second line that $S(\hat{X}^n\hat{B}^n R^n | X'^n B_0' D_0' W_X W_B W_D)_\zeta = S(\hat{X}^n\hat{B}^n R^n)_\zeta = S(\hat{X}^n\hat{B}^n R^n)_\xi$, and in the forth line that we can without loss of generality assume $|R| \leq |X||B|$, since that is the maximum possible dimension of the support of ω^R. ∎

5.2 Quantum data compression with classical side information

In this section, we assume that Alice sends her information to Debbie at rate $R_X = \log|\mathcal{X}|$ such that Debbie can decode it perfectly, and we ask how much Bob can compress his system given that the decoder has access to classical side information X^n. This problem is a special case of the *classical-quantum Slepian-Wolf problem* (CQSW problem), and we call it quantum data compression with classical side information at the decoder, in analogy to the problem of classical data compression with quantum side information at the decoder which is addressed in [29, 30]. Note we do not speak about the compression and decompression of the classical part at all, and the decoder may depend directly on x^n. Of course, by Shannon's data compression theorem [1], X can always be compressed to a rate $R_X = H(X)$, introducing an arbitrarily small error probability.

We know from previous section that Bob's encoder, in the entanglement-assisted model, is without loss of generality an isometry $U \equiv U_B : B^n B_0 \longrightarrow CWB_0'$, taking B^n and Bob's part of the entanglement B_0 to systems $C \otimes W \otimes B_0'$, where $C \equiv C_B$ is the compressed information of rate $R_B = \frac{1}{n} \log |C|$; $W \equiv W_B$ is the environment of Bob's encoding CPTP map, and B_0' is the register carrying Bob's share of the output entanglement (in this section, we drop subscript B from C_B and W_B). Having access to side information X^n, Debbie applies the decoding isometry $V : X^n C D_0 \to \hat{X}^n \hat{B}^n W_D D_0'$ to generate the output systems $\hat{X}^n \hat{B}^n$ and entanglement share D_0', and where W_D is the environment of the isometry. We call this encoding-decoding scheme a side information code of block length n and error ϵ for the entanglement-assisted model if the average fidelity (5.4) is at least $1 - \epsilon$. Similarly, we define a side information code for the unassisted model by removing the corresponding systems of entanglement in the encoding and decoding isometries, that is systems B_0, B_0', D_0 and D_0'.

To state our lower bound on the necessary compression rate, we introduce the following quantity, which emerges naturally from the converse proof.

Definition 5.1. *For the state $\omega^{XBR} = \sum_x p(x) |x\rangle\langle x|^X \otimes |\psi_x\rangle\langle\psi_x|^{BR}$ and $\delta \geq 0$, define*

$$I_\delta(\omega) := \sup_{\mathcal{T}} I(X:W)_\sigma$$

$$s.t. \ \mathcal{T} : B \to W \ \ CPTP \ with \ I(R:W|X)_\sigma \leq \delta,$$

where the mutual informations are understood with respect to the state $\sigma^{XWR} = (\mathrm{id}_{XR} \otimes \mathcal{T})\omega$ and W ranges over arbitrary finite dimensional quantum systems. Furthermore, let $\widetilde{I}_0 := \lim_{\delta \searrow 0} I_\delta = \inf_{\delta > 0} I_\delta$.

Note that the system W is not restricted in any way, which is the reason why in this definition we have a supremum and an infimum, rather than a maximum and a minimum. (It is a simple consequence of compactness of the domain of optimisation, together with the continuity of the mutual information, that if we were to impose a bound on the dimension of W in the above definition, the supremum in I_δ would be attained, and for the infimum in $\widetilde{I}_0$, it would hold that $\widetilde{I}_0 = I_0$.)

Lemma 5.2. *The function $I_\delta(\omega)$ introduced in Definition 5.1, has the following properties:*

1. *It is a non-decreasing function of δ.*

2. *It is a concave function of δ.*

3. *It is continuous for $\delta > 0$.*

4. *For any two states $\omega_1^{X_1 B_1 R_1}$ and $\omega_2^{X_2 B_2 R_2}$ and for $\delta, \delta_1, \delta_2 \geq 0$*

$$I_\delta(\omega_1 \otimes \omega_2) = \max_{\delta_1 + \delta_2 = \delta} \left(I_{\delta_1}(\omega_1) + I_{\delta_2}(\omega_2) \right).$$

5. $I_{n\delta}(\omega^{\otimes n}) = n I_\delta(\omega)$.

6. I_0 *and* $\widetilde{I}_0$ *are additive:*

$$I_0(\omega_1 \otimes \omega_2) = I_0(\omega_1) + I_0(\omega_2) \quad and$$
$$\widetilde{I}_0(\omega_1 \otimes \omega_2) = \widetilde{I}_0(\omega_1) + \widetilde{I}_0(\omega_2).$$

Proof. 1) The non-decrease with δ is evident from the definition.

2) For this consider $\delta_1, \delta_2 \geq 0$, $0 < p < 1$, and let $\delta = p\delta_1 + (1 - p)\delta_2$. Let furthermore channels $\mathcal{T}_i : B \to W_i$ be given $(i = 1, 2)$ such that for the states $\sigma_i^{X W_i R} = (\mathrm{id}_{XR} \otimes \mathcal{T}_i)\omega$, $I(R : W_i | X)_{\sigma_i} \leq \delta_i$.

Now define $W := W_1 \oplus W_2$, so that W_1 and W_2 can be considered mutually orthogonal subspaces of W, and define the new channel $\mathcal{T} := p\mathcal{T}_1 + (1 - p)\mathcal{T}_2 : B \to W$. By the chain rule for the mutual information, one can check that w.r.t. $\sigma^{XWR} = (\mathrm{id}_{XR} \otimes \mathcal{T})\omega$,

$$\begin{aligned}
I(R : W | X)_\sigma &= p I(R : W_1 | X)_{\sigma_1} + (1 - p) I(R : W_2 | X)_{\sigma_2} \\
&\leq p\delta_1 + (1 - p)\delta_2 \\
&= \delta,
\end{aligned}$$

and likewise

$$I(X : W)_\sigma = p I(X : W_1)_{\sigma_1} + (1 - p) I(X : W_2)_{\sigma_2}.$$

Hence, $I_\delta \geq p I(X : W_1)_{\sigma_1} + (1 - p) I(X : W_2)_{\sigma_2}$; by maximizing over the channels, the concavity follows.

3) Properties 1 and 2 imply that it is continuous for $\delta > 0$.

4) First, we prove that $I_\delta(\omega_1 \otimes \omega_2) \leq \max_{\delta_1 + \delta_2 = \delta} \left(I_{\delta_1}(\omega_1) + I_{\delta_2}(\omega_2) \right)$; the other direction of the inequality is trivial from the definition. Let $\mathcal{T} : B_1 B_2 \to W$ be a CPTP map such that

$$\begin{aligned}
\delta &\geq I(W : R_1 R_2 | X_1 X_2) \\
&= I(W : R_1 | X_1 X_2) + I(W : R_2 | X_1 R_1 X_2) \quad\quad (5.8) \\
&= I(W X_2 : R_1 | X_1) + I(W X_1 R_1 : R_2 | X_2),
\end{aligned}$$

70

where the first line is to chain rule, and the second line is due to the independence of ω_1 and ω_2. We now define the new systems $W_1 \coloneqq WX_2$ and $W_2 \coloneqq WX_1R_1$. Then we have,

$$
\begin{aligned}
I(W:X_1X_2) &= I(W:X_2) + I(W:X_1|X_2) && (5.9)\\
&= I(W:X_2) + I(WX_2:X_1)\\
&\le I(\underbrace{WX_1R_1}_{W_2}:X_2) + I(\underbrace{WX_2}_{W_1}:X_1),
\end{aligned}
$$

where the second equality is due to the independence of X_1 and X_2. The inequality follows from data processing. From Eq. (5.8) we know that $I(W_1 : R_1|X_1) \le \delta_1$ and $I(W_2 : R_2|X_2) \le \delta_2$ for some $\delta_1 + \delta_2 = \delta$. Thereby, from Eq. (5.9) we obtain

$$
\begin{aligned}
I_\delta(\omega_1 \otimes \omega_2) &\le I_{\delta_1}(\omega_1) + I_{\delta_2}(\omega_2)\\
&\le \max_{\delta_1+\delta_2=\delta} I_{\delta_1}(\omega_1) + I_{\delta_2}(\omega_2).
\end{aligned}
$$

5) Now, the multi-copy additivity follows easily from property 4: According to the first statement of the lemma, we have

$$
I_{n\delta}(\omega^{\otimes n}) = \max_{\delta_1+\ldots+\delta_n=n\delta} I_{\delta_1}(\omega) + \ldots + I_{\delta_n}(\omega).
$$

Here, the right hand side is clearly $\ge nI_\delta(\omega)$ since we can choose all $\delta_i = \delta$. By the concavity of $I_\delta(\omega)$ in δ, on the other hand, we have for any $\delta_1+\ldots+\delta_n = n\delta$ that

$$
\frac{1}{n}\big(I_{\delta_1}(\omega) + \ldots + I_{\delta_n}(\omega)\big) \le I_\delta(\omega),
$$

so the maximum is attained at $\delta_i = \delta$ for all $i = 1,\ldots,n$.

6) The property 4 of the lemma also implies that I_0 and $\widetilde{I}_0$ are additive. ∎

Remark 5.1. *There is a curious resemblance of our function I_δ with the so-called* information bottleneck function *introduced by Tishby et al. [80], whose generalization to quantum information theory is recently being discussed [81, 82]. Indeed, the concavity and additivity properties of the two functions are proved by the same principles, although it is not evident to us, what –if any–, the information theoretic link between I_δ and the information bottleneck is.*

5.2.1 Converse bound

In this subsection, we use the properties of the function $I_\delta(\omega)$ (Lemma 5.2) to prove a lower bound on Bob's quantum communication rate.

Theorem 5.1. *In the entanglement-assisted model, consider any side information code of block length n and error ϵ. Then, Bob's quantum communication rate is lower bounded as*

$$R_B \geq \frac{1}{2}\left(S(B) + S(B|X) - I_{\delta(n,\epsilon)} - \delta(n,\epsilon)\right),$$

where $\delta(n,\epsilon) = 4\sqrt{6\epsilon}\log(|X||B|) + \frac{2}{n}h(\sqrt{6\epsilon})$. Any asymptotically achievable rate R_B is consequently lower bounded

$$R_B \geq \frac{1}{2}\left(S(B) + S(B|X) - \widetilde{I}_0\right).$$

Proof. As already discussed in the introduction to this section, the encoder of Bob is without loss of generality an isometry $U : B^n B_0 \longrightarrow CWB'_0$. The existence of a high-fidelity decoder using X^n as side information implies that systems WB'_0 are decoupled from system R^n conditional on X^n; indeed, by Lemma 5.1, $I(R^n : WB'_0|X'^n) \leq n\delta(n,\epsilon)$. The first part of the converse reasoning is as follows:

$$\begin{aligned}
nR_B = \log|C| &\geq S(C) \\
&\geq S(CWB'_0) - S(WB'_0) \\
&= S(B^n) + S(B_0) - S(WB'_0),
\end{aligned}$$

where the second inequality is a version of subadditivity, and the equality in the last line holds because the encoding isometry U does not change the entropy; furthermore, B^n and B_0 are initially independent. Moreover, the decoder can be dilated to an isometry $V : X^n C D_0 \longrightarrow \hat{X}^n \hat{B}^n D'_0 W_D$, where W_D and D'_0 are the environment of Debbie's decoding operation and the output of Debbie's entanglement, respectively. Using the decoupling condition of Lemma 5.1 once more, we have

$$\begin{aligned}
nR_B + S(D_0) = \log|C| + S(D_0) &\geq S(C) + S(D_0) \\
&\geq S(CD_0) \\
&\geq S(X^n C D_0|X'^n) \\
&= S(\hat{X}^n \hat{B}^n D'_0 W_D|X'^n) \\
&= S(WB'_0 R^n|X'^n) \\
&\geq S(R^n|X'^n) + S(WB'_0|X'^n) - n\delta(n,\epsilon) \\
&= S(B^n|X^n) + S(WB'_0|X'^n) - n\delta(n,\epsilon),
\end{aligned}$$

where the third and fourth line are by subadditivity of the entropy; the fifth line follows because the decoding isometry V does not change the entropy.

72

The sixth line holds because for any given x^n the overall state of the systems $\hat{X}^n \hat{B}^n B_0' D_0' W W_D R^n$ is pure. The penultimate line is due to the decoupling condition (Lemma 5.1), and the last line follows because for a given x^n the overall state of the systems $B^n R^n$ is pure. Adding these two relations and dividing by $2n$, we obtain

$$R_B \geq \frac{1}{2}(S(B) + S(B|X)) - \frac{1}{2n} I(X'^m : W B_0') - \frac{1}{2}\delta(n, \epsilon),$$

where the terms $S(B_0)$ and $S(D_0)$ cancel out each other because B_0 and D_0 are K-dimensional quantum registers with maximally entangled states Φ_K.

In the above inequality, the mutual information on the right hand side is bounded as

$$I(X'^m : W B_0') \leq I_{n\delta(n,\epsilon)}(\omega^{\otimes n}) = n I_{\delta(n,\epsilon)}(\omega),$$

To see this, define the CPTP map $\mathcal{T} : B^n \longrightarrow \widetilde{W} := W B_0'$ as $\mathcal{T}(\rho) := \operatorname{Tr}_{C D_0}(U \otimes \mathbb{1})(\rho \otimes \Phi_K^{B_0 D_0})(U \otimes \mathbb{1})^\dagger$. Then we have $I(R^n : \widetilde{W} | X'^m) \leq n\delta(n, \epsilon)$, and hence the above inequality follows directly from Definition 5.1.

The second statement of the theorem follows because $\delta(n, \epsilon)$ tends to zero as $n \to \infty$ and $\epsilon \to 0$. ∎

Remark 5.2. *Notice that the term $\frac{1}{n}I(X'^m : W B_0')$ is not necessarily small. For example, suppose that the source is of the form $|\psi_x\rangle^{BR} = |\psi_x\rangle^{B'R} \otimes |\psi_x\rangle^{B''}$ for all x; clearly it is possible to perform the coding task by coding only B' and trashing B'' (i.e. putting it into W), because by having access to x the decoder can reproduce $\psi_x^{B''}$ locally. In this setting, characteristically $\frac{1}{n}I(X'^m : W B_0')$ does not go to zero because B''^n ends up in W.*

5.2.2 Achievable rates

In this subsection, we provide achievable rates both for the unassisted and entanglement-assisted model.

Theorem 5.2. *In the unassisted model, there exists a sequence of side information codes that compress Bob's system B^n at the asymptotic qubit rate*

$$R_B = \frac{1}{2}\left(S(B) + S(B|X)\right).$$

Proof. We recall that in a side information code, Bob aims to send his system B^n to Debbie while she has access to side information system X^n as explained at the beginning of this section. We can use the fully quantum Slepian-Wolf protocol (FQSW), also called coherent state merging protocol ([68]

73

section 7), as a subprotocol since it considers the entanglement fidelity as the decodability criterion, which is more stringent than the average fidelity defined in (5.2). Namely, let

$$|\Omega\rangle^{XX'BR} = \sum_{x\in\mathcal{X}} \sqrt{p(x)} \, |x\rangle^X \, |x\rangle^{X'} \, |\psi_x\rangle^{BR}$$

be the source in the FQSW problem, where B is the system to be compressed, X is the side information at the decoder, R and X' are the reference systems. Bob applies the corresponding encoding map of the FQSW protocol $\mathcal{E}_B : B^n \longrightarrow C$ and sends system C to Debbie who then applies the decoding map of the FQSW protocol $\mathcal{D} : X^n C \longrightarrow X^n \hat{B}^n$ to her side information system X^n and the compressed information C to reconstruct system $\hat{B}^n$. These encoding and decoding operations preserve the entanglement fidelity F_e which is the decodability criterion of the FQSW problem:

$$
\begin{aligned}
F_e &= F\big(\Omega^{X^n X'^n B^n R^n}, (\mathcal{D} \circ (\mathrm{id}_{X^n} \otimes \mathcal{E}_B) \otimes \mathrm{id}_{X'^n R^n})\Omega^{X^n X'^n B^n R^n}\big) \\
&\leq F\big(\omega^{X^n X'^n B^n R^n}, (\mathcal{D} \circ (\mathrm{id}_{X^n} \otimes \mathcal{E}_B) \otimes \mathrm{id}_{X'^n R^n})\omega^{X^n X'^n B^n R^n}\big) \\
&= \overline{F},
\end{aligned}
$$

where the inequality is due to the monotonicity of fidelity under CPTP maps, namely the projective measurement on system X' in the computational basis $\{|x\rangle\langle x|\}$). Therefore, if an encoding-decoding scheme attains an entanglement fidelity for the FQSW problem going to 1, then it will have the average fidelity for the CQSW problem going to 1 as well. Hence, the FQSW rate

$$R_B = \frac{1}{2} I(B : X'R)_\Omega = \frac{1}{2}(S(B)_\omega + S(B|X)_\omega),$$

is achievable. ∎

Remark 5.3. *Notice that for the source considered at the end of the previous subsection in Remark 5.2, where $|\psi_x\rangle^{BR} = |\psi_x\rangle^{B'R} \otimes |\psi_x\rangle^{B''}$ for all x, we can achieve a rate strictly smaller than the rate stated in the above theorem. The reason is that R is only entangled with B', so clearly it is possible to perform the coding task by coding only B' and trashing B'' because by having access to x the decoder can reproduce the state $\psi_x^{B''}$ locally. Thereby, the rate $\frac{1}{2}(S(B') + S(B'|X))$ is achievable by applying coherent state merging as above.*

The previous observation shows that in general, the rate $\frac{1}{2}(S(B)+S(B|X))$ from Theorem 5.2 is not optimal. By looking for a systematic way of obtaining better rates, we have the following result in the entanglement-assisted model.

Theorem 5.3. *In the entanglement-assisted model, there exists a sequence of side information codes with the following asymptotic entanglement and qubit rates:*

$$E = \frac{1}{2}\left(I(C:W)_\sigma - I(C:X)_\sigma\right) \quad and$$

$$R_B = \frac{1}{2}\left(S(B)_\omega + S(B|X)_\omega - I(X:W)_\sigma\right),$$

where C and W are, respectively, the system and environment of an isometry $V : B \to CW$ on ω^{XBR} producing the state $\sigma^{XCWR} = (\mathbb{1}_{XR} \otimes V)\omega^{XBR}(\mathbb{1}_{XR} \otimes V)^\dagger$, such that $I(W : R|X)_\sigma = 0$.

Proof. Notice that there is always an isometry $V : B \to CW$ with $I(W : R|X)_\sigma = 0$, and the trivial example is the isometry $V : B \to BW$ where system W is a trivial system with state $|0\rangle\langle 0|^W$.

First, Bob applies the isometry V to each copy of the n systems $B_1, \ldots, B_n$:

$$\sigma^{XX'CWR}$$
$$= (V^{B \to CW} \otimes \mathbb{1}_{XX'R})\omega^{XX'BR}(V^{B \to CW} \otimes \mathbb{1}_{XX'R})^\dagger$$
$$= \sum_x p(x)\, |x\rangle\langle x|^X \otimes |x\rangle\langle x|^{X'} \otimes |\phi_x\rangle\langle\phi_x|^{CWR}.$$

Now consider the following source state from which the state $\sigma^{XX'CWR}$ is obtained by applying projective measurement on system X' in the computational basis $\{|x\rangle\langle x|\}$,

$$|\Sigma\rangle^{XX'CWR} = \sum_{x \in \mathcal{X}} \sqrt{p(x)}\, |x\rangle^X |x\rangle^{X'} |\phi_x\rangle^{CWR}.$$

For this source, consider Bob and Debbie respectively hold the CW and X systems, and Bob wishes to send system C to Debbie while keeping W for himself. For many copies of the above state, the parties can apply the quantum state redistribution (QSR) protocol [26, 44] for transmitting C, having access to system W as side information at the encoder and to X as side information at the decoder. According to this protocol, Bob needs exactly the rate of $R_B = \frac{1}{2}I(C : X'R|X)_\Sigma = \frac{1}{2}(S(B)_\omega + S(B|X)_\omega - I(X : W)_\sigma)$ qubits of communication. The protocol requires the rate of $\frac{1}{2}I(C : W)_\Sigma = \frac{1}{2}I(C : W)_\sigma$ ebits of entanglement shared between the encoder and decoder, and at the end of the protocol the rate of $\frac{1}{2}I(C : X)_\Sigma = \frac{1}{2}I(C : X)_\sigma$ ebits of entanglement is distilled between the encoder and the decoder (see equations (1) and (2) in [26]). This protocol attains high fidelity for the state $\Sigma^{X^n X'^n C^n W^n R^n}$, and

75

consequently for the state $\sigma^{X^n X'^n C^n W^n R^n}$ due to the monotonicity of fidelity under CPTP maps:

$$1-\epsilon \leq F\left(\Sigma^{X^n X'^n C^n W^n R^n} \otimes \Phi_L^{B_0' D_0'}, \hat{\Sigma}^{\hat{X}^n X'^n \hat{C}^n \hat{W}^n R^n B_0' D_0'}\right)$$

$$\leq F\left(\sigma^{X^n X'^n C^n W^n R^n} \otimes \Phi_L^{B_0' D_0'}, \hat{\sigma}^{\hat{X}^n X'^n \hat{C}^n \hat{W}^n R^n B_0' D_0'}\right), \qquad (5.10)$$

where

$$\hat{\Sigma}^{\hat{X}^n X'^n \hat{C}^n \hat{W}^n R^n B_0' D_0'}$$

$$= (\mathcal{D} \circ (\mathrm{id}_{X^n D_0} \otimes \mathcal{E}_{CWB_0}) \otimes \mathrm{id}_{X'^n R^n}) \Sigma^{X^n X'^n C^n W^n R^n} \otimes \Phi_K^{B_0 D_0},$$

and

$$\hat{\sigma}^{\hat{X}^n X'^n \hat{C}^n \hat{W}^n R^n B_0' D_0'}$$

$$= (\mathcal{D} \circ (\mathrm{id}_{X^n D_0} \otimes \mathcal{E}_{CWB_0}) \otimes \mathrm{id}_{X'^n R^n}) \sigma^{X^n X'^n C^n W^n R^n} \otimes \Phi_K^{B_0 D_0},$$

and $\mathcal{E}_{CWB_0}$ and $\mathcal{D}$ are respectively the encoding and decoding operations of the QSR protocol. The condition $I(W:R|X)_\sigma = 0$ implies that for every x the systems W and R are decoupled:

$$\phi_x^{WR} = \phi_x^W \otimes \phi_x^R.$$

By Uhlmann's theorem [56, 83], there exist isometries $V_x : C \to VB$ for all $x \in \mathcal{X}$, such that

$$\left(\mathbb{1} \otimes V_x^{C \to VB}\right) |\phi_x\rangle^{CWR} = |\nu_x\rangle^{VW} \otimes |\psi_x\rangle^{BR}.$$

After applying the decoding operation $\mathcal{D}$ of QSR, Debbie applies the isometry $V_x : C \to VB$ for each x, which does not change the fidelity (5.10). By tracing out the unwanted systems $V^n W^n$, due to the monotonicity of the fidelity under partial trace, the fidelity defined in (5.4) will go to 1 in this encoding-decoding scheme. ∎

Remark 5.4. *In Theorem 5.3, the smallest achievable rate, when unlimited entanglement is available, is equal to $\frac{1}{2}(S(B) + S(B|X) - I_0)$. This rate resembles the converse bound $R_B \geq \frac{1}{2}(S(B) + S(B|X) - \widetilde{I}_0)$, except that $\widetilde{I}_0 \geq I_0$. In the definition of $\widetilde{I}_0$, it seems unlikely that we can take the limit of δ going to 0 directly because there is no dimension bound on the systems C and W, so compactness cannot be used directly to prove that $\widetilde{I}_0$ and I_0 are equal.*

76

Remark 5.5. *Looking again at the entanglement rate in Theorem 5.3, $E = \frac{1}{2}\left(I(C:W)_\sigma - I(C:X)_\sigma\right)$, we reflect that there may easily be situations where $E \leq 0$, meaning that no entanglement is consumed, and in fact no initial entanglement is necessary. In this case, the theorem improves the rate of Theorem 5.2 by the amount $\frac{1}{2}I(X:W)$. This motivates the definition of the following variant of I_0,*

$$I_{0-}(\omega) \coloneqq \sup I(X:W) \; s.t. \; I(R:W|X) = 0,$$
$$I(C:W) - I(C:X) \leq 0,$$

where the supremum is over all isometries $V:B \to CW$.

As a corollary to these considerations, in the unassisted model the rate $\frac{1}{2}\left(S(B) + S(B|X) - I_{0-}\right)$ is achievable.

5.2.3 Optimal compression rate for generic sources

In this subsection, we find the optimal compression rate for *generic* sources, by which we mean any source except for a submanifold of lower dimension within the set of all sources. Concretely, we will consider sources where there is at least one x for which the reduced state $\psi_x^B = \operatorname{Tr}_R |\psi_x\rangle\langle\psi_x|^{BR}$ has full support on B. In this setting, coherent state merging as a subprotocol gives the optimal compression rate, so not only does the protocol not use any initial entanglement, but some entanglement is distilled at the end of the protocol.

Theorem 5.4. *In both unassisted and entanglement-assisted models, for any side information code of a generic source, the asymptotic compression rate R_B of Bob is lower bounded*

$$R_B \geq \frac{1}{2}\left(S(B) + S(B|X)\right),$$

so the protocol of Theorem 5.2 has optimal rate for a generic source. Moreover, in that protocol no prior entanglement is needed and a rate $\frac{1}{2}I(X:B)$ ebits of entanglement is distilled between the encoder and decoder.

Proof. The converse bound of Theorem 5.1 states that the asymptotic quantum communication rate of Bob is lower bounded as

$$R_B \geq \frac{1}{2}\left(S(B) + S(B|X) - \widetilde{I_0}\right),$$

where $\widetilde{I_0}$ comes from Definition 5.1. We will show that for generic sources, $\widetilde{I_0} = I_0 = 0$. Moreover, Theorem 5.2 states that using coherent state merging, the asymptotic qubit rate of $\frac{1}{2}(S(B) + S(B|X))$ is achievable, that no prior

entanglement is required and a rate of $\frac{1}{2}I(X:B)$ ebits of entanglement is distilled between the encoder and the decoder.

We show that for any CPTP map $\mathcal{T} : B \to W$, which acts on a generic ω^{XBR} and produces state $\sigma^{XWR} = (\mathrm{id}_{XR} \otimes \mathcal{T})\omega^{XBR}$ such that $I(R:W|X)_\sigma \leq \delta$ for $\delta \geq 0$, the quantum mutual information $I(X:W)_\sigma \leq \delta' \log|X| + 2h(\frac{1}{2}\delta')$ where δ' is defined in Eq. (5.12) below. Thus, we obtain

$$\widetilde{I}_0 = \lim_{\delta \searrow 0} I_\delta = 0.$$

To show this claim, we proceed as follows. From $I(R:W|X)_\sigma \leq \delta$ we have

$$I(R:W|X=x)_\sigma \leq \frac{\delta}{p(x)} \qquad \forall x \in \mathcal{X},$$

so by Pinsker's inequality [19] we obtain

$$\left\| \phi_x^{WR} - \phi_x^W \otimes \phi_x^R \right\|_1 \leq \sqrt{\frac{2\delta \ln 2}{p(x)}} \qquad \forall x \in \mathcal{X}.$$

By Uhlmann's theorem (Lemma A.8 and Lemma A.9), there exists an isometry $V_x : C \to BV$ such that

$$\left\| (V_x \otimes \mathbb{1}_{WR}) \phi_x^{CWR} (V_x \otimes \mathbb{1}_{WR})^\dagger - \theta_x^{WV} \otimes \psi_x^{BR} \right\|_1$$
$$\leq \sqrt{\sqrt{\frac{\delta \ln 2}{2p(x)}} \left(2 - \sqrt{\frac{\delta \ln 2}{2p(x)}} \right)}, \qquad (5.11)$$

where θ_x^{WV} is a purification of ϕ_x^W. Since the source is generic by definition there is an x, say $x = 0$, for which ψ_0^B has full support on $\mathcal{L}(\mathcal{H}_B)$, i.e. $\lambda_0 := \lambda_{\min}(\psi_0^B) > 0$. By Lemma A.12 in Appendix A, for any $|\psi_x\rangle^{BR}$ there is an operator T_x acting on the reference system such that

$$|\psi_x\rangle^{BR} = (\mathbb{1}_B \otimes T_x) |\psi_0\rangle^{BR}.$$

Using this fact, we show that the decoding isometry V_0 in Eq. (5.11) works

for all states:

$$\left\| (V_0 \otimes \mathbb{1}_{WR}) \phi_x^{CWR} (V_0^\dagger \otimes \mathbb{1}_{WR}) - \theta_0^{WV} \otimes \psi_x^{BR} \right\|_1$$

$$= \left\| (V_0 \otimes \mathbb{1}_{WR})(\mathbb{1}_{CW} \otimes T_x) \phi_0^{CWR} (\mathbb{1}_{CW} \otimes T_x)^\dagger (V_0^\dagger \otimes \mathbb{1}_{WR}) \right.$$
$$\left. - \theta_0^{WV} \otimes (\mathbb{1}_B \otimes T_x) \psi_0^{BR} (\mathbb{1}_B \otimes T_x)^\dagger \right\|_1$$

$$= \left\| (\mathbb{1}_{BVW} \otimes T_x)(V_0 \otimes \mathbb{1}_{WR}) \phi_0^{CWR} (V_0^\dagger \otimes \mathbb{1}_{WR})(\mathbb{1}_{BVW} \otimes T_x^\dagger) \right.$$
$$\left. - (\mathbb{1}_{BVW} \otimes T_x) \theta_0^{WV} \otimes \psi_0^{BR} (\mathbb{1}_{BVW} \otimes T_x^\dagger) \right\|$$

$$\le \left\| \mathbb{1}_{BVW} \otimes T_x \right\|_\infty^2$$
$$\left\| (V_0 \otimes \mathbb{1}_{WR}) \phi_0^{CWR} (V_0^\dagger \otimes \mathbb{1}_{WR}) - \theta_0^{WV} \otimes \psi_0^{BR} \right\|_1$$

$$\le \frac{1}{\lambda_0} \sqrt{ \sqrt{\frac{\delta \ln 2}{2p(0)}} \left(2 - \sqrt{\frac{\delta \ln 2}{2p(0)}} \right) },$$

where the last two inequalities follow from Lemma A.3 and Lemma A.12, respectively. By tracing out the systems VBR in the above chain of inequalities, we get

$$\left\| \phi_x^W - \phi_0^W \right\|_1 \le \frac{1}{\lambda_0} \sqrt{ \sqrt{\frac{\delta \ln 2}{2p(0)}} \left(2 - \sqrt{\frac{\delta \ln 2}{2p(0)}} \right) } =: \delta'. \qquad (5.12)$$

Thus, by triangle inequality we obtain

$$\left\| \underbrace{\sum_x p(x) |x\rangle\langle x|^X \otimes \phi_x^W}_{\sigma^{XW}} - \underbrace{\sum_x p(x) |x\rangle\langle x|^X \otimes \phi_0^W}_{=: \sigma_0^{XW}} \right\|_1$$

$$\le \sum_x p(x) \left\| \phi_x^W - \phi_0^W \right\|_1$$

$$\le \frac{1}{\lambda_0} \sqrt{ \sqrt{\frac{\delta \ln 2}{2p(0)}} \left(2 - \sqrt{\frac{\delta \ln 2}{2p(0)}} \right) } = \delta'. \qquad (5.13)$$

By applying the Alicki-Fannes inequality in the form of Lemma A.11, to Eq. (5.13), we have

$$I(X:W)_\sigma = S(X)_\sigma - S(X|W)_\sigma + S(X|W)_{\sigma_0} - S(X|W)_{\sigma_0}$$
$$= S(X|W)_{\sigma_0} - S(X|W)_\sigma$$
$$\le \delta' \log |X| + 2h\left(\frac{1}{2} \delta' \right),$$

and the right hand side of the above inequality vanishes for $\delta \to 0$. ∎
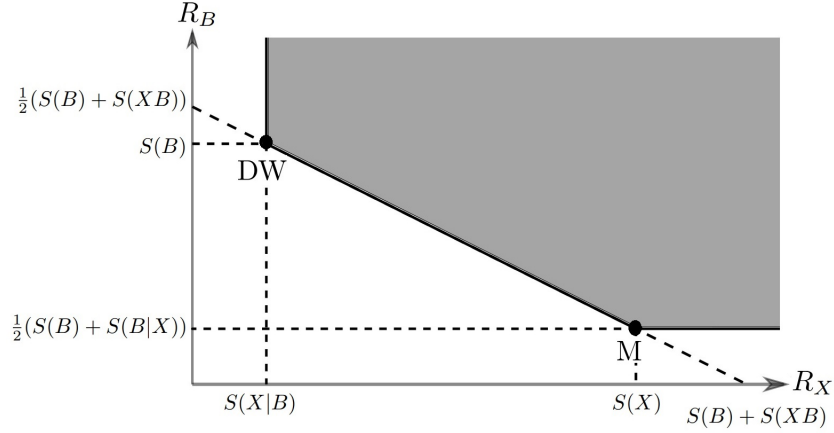
Figure 5.3: The region of all pairs (R_X, R_B) satisfying the three conditions of Eq. (5.14); it is the upper-right convex closure of the Devetak-Winter (DW) and the merging (M) point. All of these points are achievable in the unassisted model.

5.3 Towards the full rate region

In this section, we consider the full rate region of the distributed compression of a classical-quantum source.

Theorem 5.5. *In the unassisted model, for distributed compression of a classical-quantum source, the rate pairs satisfying the following inequalities are achievable:*

$$
\begin{aligned}
R_X &\geq S(X|B), \\
R_B &\geq \frac{1}{2}\left(S(B) + S(B|X)\right), \\
R_X + 2R_B &\geq S(B) + S(XB).
\end{aligned}
\tag{5.14}
$$

Proof. From the Devetak-Winter code, Eq. (5.3), and the code based on state merging, Theorem 5.2, two rate points in the unassisted (and hence also in the unlimited entanglement-assisted) rate region are:

$$
\begin{aligned}
(R_X, R_B) &= (S(X|B), S(B)), \\
(R_X, R_B) &= \left(S(X), \frac{1}{2}(S(B) + S(B|X))\right).
\end{aligned}
$$

Their upper-right convex closure is hence an inner bound to the rate region, depicted schematically in Fig. 5.3. ∎

80

For generic sources we find that this is in fact the rate region. However, in general, we only present some outer bounds and inner bounds (achievable rates), which show the rate region to be much more complicated than the rate region of the classical Slepian-Wolf problem.

5.3.1 General converse bounds

For distributed compression of a classical-quantum source in general, we start with a general converse bound.

Theorem 5.6. *In the entanglement-assisted model, the asymptotic rate pairs for distributed compression of a classical-quantum source are lower bounded as*

$$R_X \geq S(X|B),$$
$$R_B \geq \frac{1}{2}\left(S(B) + S(B|X) - \widetilde{I}_0\right), \qquad (5.15)$$
$$R_X + 2R_B \geq S(B) + S(BX) - \widetilde{I}_0.$$

In the unassisted model, in addition to the above lower bounds, the asymptotic rate pairs are bounded as

$$R_X + R_B \geq S(XB).$$

Proof. The individual lower bounds have been established already: $R_X \geq S(X|B)$ is from [29, 30], in a slightly different source model. However, it also holds in our system model if Bob sends his information using unlimited communication such that Debbie can decode it perfectly. Namely, notice that the fidelity (5.2) is more stringent than the decoding criterion of [29,30], so any converse bound considering the decoding criterion of [29, 30] is also a converse bound in our system model. The bound $R_B \geq \frac{1}{2}(S(B)+S(B|X)-\widetilde{I}_0)$ is from Theorem 5.4. These two bounds hold in the unassisted, as well as the entanglement-assisted model.

In the unassisted model, the rate sum lower bound $R_X + R_B \geq S(XB)$ has been argued in [29, 30], too. As a matter of fact, for any distributed compression scheme for the source, $\mathcal{E}_X \otimes \mathcal{E}_B$ jointly describes a Schumacher compression scheme with asymptotically high fidelity. Thus, its rate must be asymptotically lower bounded by the joint entropy of the source, $S(XB)$ [8, 30, 60, 61].

This leaves the bound $R_X + 2R_B \geq S(B) + S(BX) - \widetilde{I}_0$ to be proved in the entanglement-assisted model, which we tackle now. The encoders of Alice and Bob are isometries $U_X : X^n \to C_X W_X$ and $U_B : B^n B_0 \to C_B W_B B'_0$,

respectively. They send their respective compressed systems C_X and C_W to Debbie and keep the environment parts W_X and W_B for themselves. Then, Debbie applies the decoding isometry $V : C_X C_B D_0 \to \hat{X}^n \hat{B}^n W_D D_0'$, where $\hat{X}^n \hat{B}^n D_0'$ are the output systems, and W_D and D_0' are the environment of Debbie's decoding isometry and her output entanglement, respectively. We first bound the following sum rate:

$$
\begin{aligned}
& nR_X + nR_B + S(D_0) \\
& \geq S(C_X) + S(C_B) + S(D_0) \\
& \geq S(C_X C_B D_0) \\
& = S(\hat{X}^n \hat{B}^n W_D D_0') \\
& = S(\hat{X}^n \hat{B}^n) + S(W_D D_0' | \hat{X}^n \hat{B}^n) \\
& \geq S(\hat{X}^n \hat{B}^n) + S(W_D D_0' | \hat{X}^n \hat{B}^n X'^m) \\
& \geq S(X^n B^n) + S(W_D D_0' | \hat{X}^n \hat{B}^n X'^m) \\
& \qquad\qquad\qquad - n\sqrt{2\epsilon}\log(|X||B|) - h(\sqrt{2\epsilon}) \\
& \geq S(X^n B^n) + S(W_D D_0' | X'^m) - 2n\delta(n,\epsilon) \\
& \geq S(X^n B^n) + S(W_X W_B B_0' | X'^m) \\
& \qquad\qquad\qquad - S(R^n \hat{B}^n \hat{X}^n | X'^m) - 2n\delta(n,\epsilon) \\
& \geq S(X^n B^n) + S(W_X W_B B_0' | X'^m) - 2n\delta(n,\epsilon) - n\delta'(n,\epsilon) \\
& = S(X^n B^n) + S(W_X | X'^m) + S(W_B B_0' | X'^m) \\
& \qquad\qquad\qquad - 2n\delta(n,\epsilon) - n\delta'(n,\epsilon) \\
& \geq S(X^n B^n) + S(W_B B_0' | X'^m) - 2n\delta(n,\epsilon) - n\delta'(n,\epsilon), \qquad (5.16)
\end{aligned}
$$

where the third line is by subadditivity, the equality in the third line follows because the decoding isometry V does not change the entropy. Then, in the fifth and sixth line we use the chain rule and strong subadditivity of entropy. The inequality in the seventh line follows from the decodability of the systems $X^n B^n$: the fidelity criterion (5.4) implies that the output state on systems $\hat{X}^n \hat{B}^n$ is $2\sqrt{2\epsilon}$-close to the original state $X^n B^n$ in trace norm; then apply the Fannes inequality (Lemma A.10). The eighth line follows from the decoupling condition (Lemma 5.1), which implies that $I(W_D D_0' : \hat{X}^n \hat{B}^n | X'^m) \leq n\delta(n,\epsilon) = 4n\sqrt{6\epsilon}\log(|X||B|) + 2h(\sqrt{6\epsilon})$. In the ninth line, we use that for any given x^n, the overall state of $W_X W_B W_D B_0' D_0' R^n \hat{B}^n \hat{X}^n$ is pure, and invoking subadditivity. In line tenth, we use the decoding fidelity (5.4) once more, saying that the output state on systems $\hat{X}^n \hat{B}^n R^n X'^m$ is $2\sqrt{2\epsilon}$-close to the original state $X^n B^n R^n X'^m$ in trace norm; then apply the Alicki-Fannes inequality (Lemma A.11) in the following equation; notice that

given x^n the state on systems $X^n B^n R^n$ is pure, therefore $S(X^n B^n R^n | X'^n) = 0$, and we obtain:

$$
\begin{aligned}
\left| S(\hat{X}^n \hat{B}^n R^n | X'^n) \right. & \left. - S(X^n B^n R^n | X'^n) \right| \\
&= S(\hat{X}^n \hat{B}^n R^n | X'^n) \\
&\leq 2n\sqrt{2\epsilon} \log |X||B||R| + (1 + \sqrt{2\epsilon})h\left(\frac{\sqrt{2\epsilon}}{1 + \sqrt{2\epsilon}}\right) \\
&\leq 4n\sqrt{2\epsilon} \log |X||B| + (1 + \sqrt{2\epsilon})h\left(\frac{\sqrt{2\epsilon}}{1 + \sqrt{2\epsilon}}\right) \\
&\coloneqq \delta'(n,\epsilon),
\end{aligned}
\tag{5.17}
$$

where in the penultimate line, we can without loss of generality assume $|R| \leq |X||B|$. The equality in the twelfth line of Eq. (5.16) follows because for a given x^n the encoded states of Alice and Bob are independent.

Moreover, we bound R_B as follows:

$$
\begin{aligned}
nR_B &\geq S(C_B) \\
&\geq S(C_B | W_B B_0') \\
&= S(C_B W_B B_0') - S(W_B B_0') \\
&= S(B^n B_0) - S(W_B B_0') \\
&= S(B^n) + S(B_0) - S(W_B B_0').
\end{aligned}
\tag{5.18}
$$

Adding Eqs. (5.16) and (5.18), and after cancellation of $S(B_0) = S(D_0)$, we get

$$
\begin{aligned}
R_X &+ 2R_B \\
&\geq S(B) + S(XB) - \frac{1}{n} I(X'^n : W_B B_0') - 2\delta(n,\epsilon) - \delta'(n,\epsilon) \\
&\geq S(B) + S(XB) - \frac{1}{n} I_{n\delta(n,\epsilon)}(\omega^{\otimes n}) - 2\delta(n,\epsilon) - \delta'(n,\epsilon) \\
&= S(B) + S(XB) - I_{\delta(n,\epsilon)}(\omega) - 2\delta(n,\epsilon) - \delta'(n,\epsilon),
\end{aligned}
\tag{5.19}
$$

where given that $I(R^n : B_0' W_B | X'^n) \leq \delta(n,\epsilon)$, which we have from the decoupling condition (Lemma 5.1), the second equality follows directly from Definition 5.1, just as in the proof of Theorem 5.1. The equality in the last line follows from Lemma 5.2. In the limit of $n \to \infty$ and $\epsilon \to 0$, we have $\delta(n,\epsilon) \to 0$ and $\delta'(n,\epsilon) \to 0$, and so $I_{\delta(n,\epsilon)}$ converges to $\widetilde{I}_0$. ∎

5.3.2 General achievability bounds

For general, non-generic sources, the achievability bounds of Theorem 5.5 and the outer bounds of Theorem 5.6 do not match. Here we present several more

general achievability results that go somewhat towards filling in the unknown area in between, without, however, resolving the question completely.

Theorem 5.7. *In the entanglement-assisted model, for distributed compression of a classical-quantum source, any rate pairs satisfying the following inequalities are achievable: with $\alpha = \frac{2I(X:B)}{I(X:B)+I_0}$,*

$$R_X \geq S(X|B),$$
$$R_B \geq \frac{1}{2}\left(S(B) + S(B|X) - I_0\right), \qquad (5.20)$$
$$R_X + \alpha R_B \geq S(X|B) + \alpha S(B).$$

More generally, for any auxiliary random variable Y such that $Y\!-\!X\!-\!B$ is a Markov chain, all the following rate pairs (and hence also their upper-right convex closure) are achievable:

$$R_X = I(X:Y) + S(X|BY) = S(X|B) + I(Y:B),$$
$$R_B = \frac{1}{2}(S(B) + S(B|Y) - I(Y:W))$$
$$= S(B) - \frac{1}{2}\left(I(Y:B) + I(Y:W)\right),$$

where C and W are the system and environment of an isometry $V : B \to CW$ with $I(W:R|Y) = 0$.

Proof. The region described by Eq. (5.20) is precisely the upper-right convex closure of the two corner points $(S(X|B), S(B))$ and $(S(X), \frac{1}{2}(S(B) + S(B|X) - I_0))$. Their achievability follows from Theorems 5.8 and 5.3.

We use the following two achievable points to show the second statement:

$$(S(X|B), S(B)) \quad \text{and} \quad \left(S(X), \frac{1}{2}(S(B) + S(B|X) - I_0)\right).$$

Namely, Alice and Debbie (the receiver) use the Reverse Shannon Theorem to simulate the channel taking X to Y in i.i.d. fashion, which costs $I(X:Y)$ bits of classical communication [20]. Now we are in a situation that we know, Bob has to encode B^n with side information Y^n at the decoder, which can be done at the rate $\frac{1}{2}(S(B) + S(B|Y) - I(Y:W))$, by the quantum state redistribution protocol of Theorem 5.3. Then Alice has to send some more information to allow the receiver to decode X^n which is an instance of classical compression of X with quantum side information BY that is already at the decoder, hence costing another $S(X|BY)$ bits in communication, by the Devetak-Winter protocol [29, 30]. For $Y = X$, we recover the rate point $\left(S(X), \frac{1}{2}(S(B) + S(B|X) - I_0)\right)$, and for $Y = \varnothing$ we recover $(S(X|B), S(B))$. ∎
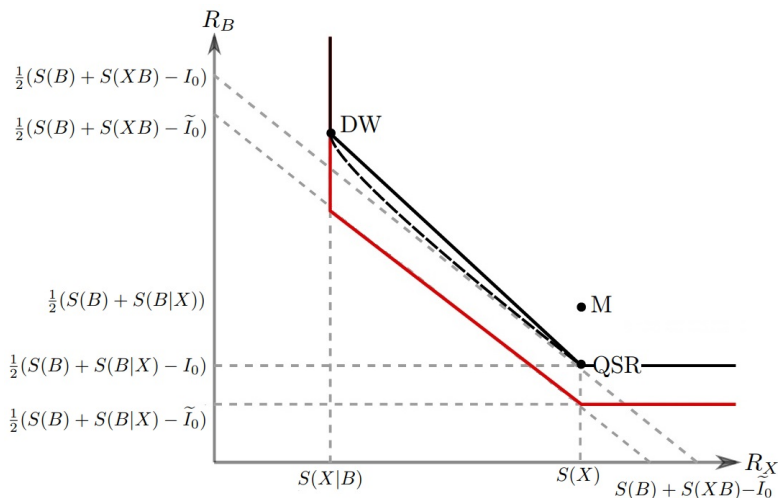
Figure 5.4: General outer (converse) bound, in red, and inner (achievable) bounds, in black, on the entanglement-assisted rate region, assuming unlimited entanglement. In general, our achievable points, the one from Devetak-Winter (DW), and the ones using merging (M) and quantum state redistribution (QSR) are no longer on the boundary of the outer bound. The achievable region is potentially slightly larger than the upper-right convex closure of the points DW and QSR, connected by a solid black straight line; indeed, the second part of Theorem 5.7 allows us to interpolate between DW and QSR along the black dashed curve.

In Fig. 5.4, we show the situation for a general source, depicting the most important inner and outer bounds on the rate region in the entanglement-assisted model.

5.3.3 Rate region for generic sources

In this subsection, we find the complete rate region for generic sources, generalizing the insight of Theorem 5.4 for the subproblem of quantum compression with classical side information at the decoder.

Theorem 5.8. *In both unassisted and entanglement-assisted models, for a generic classical-quantum source, in particular one where there is an x such that ψ_x^B has full support, the optimal asymptotic rate region for distributed*

85

compression is the set of rate pairs satisfying

$$R_X \geq S(X|B),$$
$$R_B \geq \frac{1}{2}\left(S(B) + S(B|X)\right),$$
$$R_X + 2R_B \geq S(B) + S(XB).$$

Moreover, there are protocols achieving these bounds requiring no prior entanglement.

Proof. We have argued the achievability already at the start of this section (Theorem 5.5). As for the converse, we have shown in Theorem 5.4 that for a generic source, $\widetilde{I}_0 = 0$, hence the claim follows from the outer bounds of Theorem 5.6. ∎

This means that for generic sources, which we recall are the complement of a set of measure zero, the rate region has the shape of Fig. 5.3.

5.4 Discussion and open problems

After seeing no progress for over 15 years in the problem of distributed compression of quantum sources, we have decided to take a fresh look at the classical-quantum sources considered in [29, 30]. There, the problem of compressing the classical source using the quantum part as side information at the decoder was solved; here we analyzed the full rate region, in particular we were interested in the other extreme of compressing the quantum source using the classical part as side information at the decoder. Like in the classical Slepian-Wolf coding, the former problem exhibits no rate loss, in that the quantum part of the source is compressed to the Schumacher rate, the local entropy, and the sum rate equals the joint entropy of the source. Interestingly, this is not the case for the latter problem: clearly, if the classical side information were available both at the encoder and the decoder, the optimal compression rate would be the conditional entropy $S(B|X)$, which would again imply no sum rate loss. However, since the classical side information is supposed to be present only at the decoder, we have shown that in general the rate sum is strictly larger, in fact generically by $\frac{1}{2}I(X : B)$, and with this additional rate there is always a coding scheme achieving asymptotically high fidelity. This additional rate could be called "the price of ignorance", as it corresponds to the absence of the side information at the encoder.

To deal with general classical-quantum sources, we introduced information quantities I_0 and $\widetilde{I}_0$ (Definition 5.1), to upper and lower bound the

optimal quantum compression rate as

$$\frac{1}{2}\big(S(B)+S(B|X)-\widetilde{I}_0\big)\leq R_B^*\leq\frac{1}{2}\big(S(B)+S(B|X)-I_0\big),$$

when unlimited entanglement is available. For generic sources, $I_0 = \widetilde{I}_0 = 0$, but in general we do not understand these quantities very well, and the first set of open problems that we would like to mention is about them: is $I_0 = \widetilde{I}_0$ in general, or are there examples of gaps? How can one calculate either one of these quantities, given that a priori the auxiliary register W is unbounded? In fact, can one without loss of generality put a finite bound on the dimension of W, for either optimization problem?

Further open problems concern the need for prior shared entanglement to achieve the optimal quantum compression rate R_B^*. As a matter of fact, it would already be interesting to know whether the rate $\frac{1}{2}\left(S(B) + S(B|X) - I_0\right)$ requires in general pre-shared entanglement.

The full rate region inherits these features: while it is simple, and in fact generated by the optimal codes for the two compression-with-side-information problems (quantum compression with classical side information, and classical compression with quantum side information), in the generic case, in general the picture is very complicated, and we have only been able to give several outer and inner bounds on the rate region, whose determination remains an open problem.

We also would like to comment on the source model that we consider in this chapter, and its relation to the classical Slepian-Wolf coding. Our classical-quantum source is characterised by a classical source, the random variable X, and a quantum source B, which is described by a density matrix ρ_x^B, but realized as quantum correlation with a purifying reference system R: $\rho_x^B = \text{tr}_R |\psi_x\rangle\langle\psi_x|^{BR}$. A source code in our sense reproduces the states $|\psi_x\rangle^{BR}$ with high fidelity on average, which implies that, for any ensemble decomposition $\rho_x^B = \sum_y p(y|x) |\psi_{xy}\rangle\langle\psi_{xy}|^B$, it reproduces the states $|\psi_{xy}\rangle^B$ with high fidelity on average (with respect to the ensemble probabilities $p(x)p(y|x)$). If we only demand the latter, there is no need for the purifying system R, and the source can be described compactly by the cccq-state

$$\sigma^{X'XYB} =$$
$$\sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x)p(y|x)|x\rangle\langle x|^{X'}\otimes|x\rangle\langle x|^X\otimes|y\rangle\langle y|^Y\otimes|\psi_{xy}\rangle\langle\psi_{xy}|^B, \qquad (5.21)$$

where X' and Y are reference systems with which the correlation is preserved in a compression protocol. This now includes the well-known classical correlated source considered by Slepian and Wolf [43], namely if the system B

87

is classical with orthonormal states $|\psi_{xy}\rangle = |y\rangle$. In the Schumacher's single compression problem [8], both source models, that is, the ensemble source and the purified source, lead to the same compression rate. However, when there is side information or more generally in the distributed setting, different source models, albeit sharing the reduced states on XB, do not lead to the same compression rate [30]. Our results provide a clear manifestation of this: recall that the minimum compression rate of Bob in the Slepian-Wolf setting is $S(B|X)$, with the ensemble fidelity criterion. On the other hand, if the distributions $p(y|x)$ have pairwise overlapping support, or theorem regarding generic sources applies, resulting in the strictly larger minimum rate $\frac{1}{2}(S(B) + S(B|X))$ when the average entanglement fidelity criterion is used. The difference can be attributed to the harder task of maintaining the entanglement with the reference system, rather than "only" classical correlation.

More broadly, a quantum source can be defined as a quantum system together with correlations with a reference system, in our case any state ρ^{ABR}. The compression task is to reproduce this state with high fidelity by coding and decoding of A and B. While this problem is far from understood in the general case, what we saw here is that the compression rate may depend on the concrete correlation with the reference system. In the present chapter, we have considered both a globally purifying quantum system and an ensemble of purifications, and in this final discussion, implicitly looked at a classical system keeping track of an ensemble of states subject to a probability distribution.

Finally, we mention that both models of quantum data compression with classical side information with partially purified source of Eq. (5.1) and the ensemble model defined in Eq. (5.21) are special cases of the model that we consider in the next chapter. There we define an ensemble extension of the QSR source, namely the ensemble $\{p(x), |\psi_x\rangle\langle\psi_x|^{ACBR}\}$ with corresponding cqqqq-state $\sum_x p(x)|x\rangle\langle x|^X \otimes |\psi_x\rangle\langle\psi_x|^{ACBR}$ where Alice who has access to side information system C wants to compress system A and send it, via a noiseless quantum channel, to Bob who has access to side information system B. We let the encoder and decoder share free entanglement and consider two decodability critera: per-copy fidelity and block fidelity where in the former the fidelity is preserved for each copy of the source while in the latter the fidelity is preserved for the whole block of n systems similar to the fidelity defined in Eq. (5.2). For the former criterion we find the optimal quantum communication rate and for the latter criterion we find a converse bound and an achievable rate which match up to an asymptotic error and an unbounded auxiliary system. Our new results imply that in the compression of system B with classical side information at the decoder X in the source model of Eq. (5.1), the converse bound of Theorem 5.1, i.e. the following rate is

optimal in the entanglement-assisted model with per-copy fidelity:

$$R_B = \frac{1}{2}\left(S(B) + S(B|X) - \widetilde{I}_0\right).$$

Chapter 6

Quantum state redistribution for ensemble sources

In this chapter, we consider a generalization of the quantum state redistribution task, where pure multipartite states from an ensemble source are distributed among an encoder, a decoder and a reference system. The encoder, Alice, has access to two quantum systems: system A which she compresses and sends to the decoder, Bob, and the side information system C which she wants to keep at her site. Bob has access to quantum side information in a system B, wants to decode the compressed information in such a way to preserve the correlations with the reference system on average.

As figures of merit, we consider both block error (which is the usual one in source coding) and per-copy error (which is more akin to rate-distortion theory), and find the optimal compression rate for the second criterion, and achievable and converse bounds for the first. The latter almost match in general, up to an asymptotic error and an unbounded auxiliary system; for so-called irreducible sources they are provably the same. This chapter is based on the publications in [51, 52].

6.1   The source model

Quantum state redistribution (QSR) is a source compression task where both encoder and decoder have access to side information systems [25, 26, 44]. Namely, Alice, Bob and a reference system share asymptotically many copies of a pure state $|\psi\rangle^{ACBR}$, where Alice aims to compress the quantum system A and send it to Bob via a noiseless quantum channel, while she has access to a side information quantum system C, and Bob has access to the side information quantum system B. Bob upon receiving the compressed information

reconstructs system A, and the figure of merit of this task is to preserve the entanglement fidelity between the reconstructed systems and the purifying reference system R.

Quantum state redistribution generalizes Schumacher's compression, which is recovered as the extreme case that neither encoder nor decoder have any side information [8]: the source is simply described by a pure state $|\psi\rangle^{AR}$ shared between the encoder and a reference system. However, besides this model, and originally, Schumacher considered a source generating an ensemble of pure states, i.e. $\mathcal{E} = \{p(x), |\psi_x\rangle\langle\psi_x|^A\}$, and showed both source models lead to the same optimal compression rate (cf. Barnum *et al.* [61], as well as [30]), namely the von Neumann entropy of the reduced or average state of A, respectively.

In the presence of side information systems though, an ensemble model and a purified source model can lead to different compression rates. An example of this is the classical-quantum Slepian-Wolf problem considered in [49, 50], where the compression rate can be strictly smaller than that of the corresponding purified source.

The general correlated ensemble source $\mathcal{E} = \{p(x), |\psi_x\rangle\langle\psi_x|^{AB}\}$ was considered first in [30] and then developed in [29] and by Ahn *et al.* [32], with A the system to be compressed and B the side information system at the decoder. It is an ensemble version of the coherent state merging task introduced in [68, 84]. In [29], the source is $|\psi_x\rangle^{AB} = |f(x)\rangle^A |\phi_x\rangle^B$. The optimal compression rate for an irreducible source of product states and a source generating Bell states is found in [32], however, in general case the problem had been left open.

In the present chapter, we consider an even more general ensemble source where both encoder and decoder have access to side information systems, and which thus constitutes an ensemble generalization of the pure QSR source. More precisely, we consider a source which is given by an ensemble $\mathcal{E} = \{p(x), |\psi_x\rangle\langle\psi_x|^{ACBR}\}$ of pure states $\psi_x = |\psi_x\rangle\langle\psi_x| \in \mathcal{S}(A \otimes C \otimes B \otimes R)$, $|\psi_x\rangle \in A \otimes C \otimes B \otimes R$, with a Hilbert space $A \otimes C \otimes B \otimes R$, which in this chapter we assume to be of finite dimension $|A| \cdot |C| \cdot |B| \cdot |R| < \infty$; $\mathcal{S}(A \otimes C \otimes B \otimes R)$ denotes the set of states (density operators). Furthermore, $x \in \mathcal{X}$ ranges over a discrete alphabet, so we can describe the source equivalently by the classical-quantum (cq) state $\omega^{ACBRX} = \sum_x p(x)|\psi_x\rangle\langle\psi_x|^{ACBR} \otimes |x\rangle\langle x|^X$. In this model, A and C are Alice's information to be sent and side information system, respectively. System B is the side information of Bob, and R and X are inaccessible reference systems used only to define the task.

The ensemble model of the previous chapter as well as those models that have been considered in [30, 32, 48–50] are all special cases of the model that we consider here. We find the optimal compression rate under the per-copy

fidelity criterion, and achievable and converse rates under the block-fidelity criterion which almost match, up to an asymptotic error and an unbounded auxiliary system. In the generic case of so-called *irreducible* ensembles, they are provably the same.

6.2   The compression task

We consider the information theoretic setting of many copies of the source ω^{ACBRX}, i.e. $\omega^{A^nC^nB^nR^nX^n} = (\omega^{ACBRX})^{\otimes n}$:

$$\omega^{A^nC^nB^nR^nX^n} = \sum_{x^n \in \mathcal{X}^n} p(x^n)|\psi_{x^n}\rangle\langle\psi_{x^n}|^{A^nC^nB^nR^n} \otimes |x^n\rangle\langle x^n|^{X^n},$$

using the notation

$$x^n = x_1 x_2 \ldots x_n, \qquad p(x^n) = p(x_1)p(x_2)\cdots p(x_n),$$
$$|x^n\rangle = |x_1\rangle|x_2\rangle\cdots|x_n\rangle, \quad |\psi_{x^n}\rangle = |\psi_{x_1}\rangle|\psi_{x_2}\rangle\cdots|\psi_{x_n}\rangle.$$

We assume that the encoder, Alice, and the decoder, Bob, have initially a maximally entangled state $\Phi_K^{A_0B_0}$ on registers A_0 and B_0 (both of dimension K). Alice, who has access to A^n and the side information system C^n, performs the encoding compression operation $\mathcal{E} : A^nC^nA_0 \longrightarrow M\hat{C}^n$ on A^nC^n and her part A_0 of the entanglement, which is a quantum channel, i.e. a completely positive and trace preserving (CPTP) map. Notice that as functions, CPTP maps act on the operators (density matrices) over the respective input and output Hilbert spaces, but as there is no risk of confusion, we will simply write the Hilbert spaces when denoting a CPTP map. Alice's encoding operation produces the state $\sigma^{M\hat{C}^nB^nB_0R^nX^n}$ with M, $\hat{C}^n$ and B_0 as the compressed system of Alice, the reconstructed side information system of Alice and Bob's part of the entanglement, respectively. The dimension of the compressed system is without loss of generality not larger than the dimension of the original source, i.e. $|M| \le |A|^n$. The system M is then sent via a noiseless quantum channel to Bob, who performs a decoding operation $\mathcal{D} : MB^nB_0 \longrightarrow \hat{A}^n\hat{B}^n$ on the compressed system M, his side information B^n and his part of the entanglement B_0, to reconstruct the original systems, now denoted $\hat{A}^n$ and $\hat{B}^n$. We call $\frac{1}{n}\log|M|$ the *quantum rate* of the compression protocol. We say an encoding-decoding scheme (or code, for short) has *block fidelity* $1 - \epsilon$, or *block error* ϵ, if

$$F := F\big(\omega^{A^nC^nB^nR^nX^n}, \xi^{\hat{A}^n\hat{C}^n\hat{B}^nR^nX^n}\big)$$
$$= \sum_{x^n} p(x^n)F\left(\psi_{x^n}^{A^nC^nB^nR^n}, \xi_{x^n}^{\hat{A}^n\hat{C}^n\hat{B}^nR^n}\right) \ge 1 - \epsilon, \tag{6.1}$$

93

where

$$\xi^{\hat{A}^n \hat{C}^n \hat{B}^n R^n X^n} = \sum_{x^n \in \mathcal{X}^n} p(x^n) \xi_{x^n}^{\hat{A}^n \hat{C}^n \hat{B}^n R^n} \otimes |x^n\rangle\langle x^n|^{X^n}$$
$$= \left((\mathcal{D} \circ \mathcal{E}) \otimes \mathrm{id}_{R^n X^n} \right) \omega^{A^n C^n B^n R^n X^n}.$$

We say a code has *per-copy fidelity* $1 - \epsilon$, or *per-copy error* ϵ, if

$$\overline{F} := \frac{1}{n} \sum_{i=1}^n F\big(\omega^{A_i C_i B_i R_i X^n}, \xi^{\hat{A}_i \hat{C}_i \hat{B}_i R_i X^n}\big)$$
$$= \sum_{x^n} p(x^n) \frac{1}{n} \sum_{i=1}^n F\left(\psi_{x_i}^{ACBR}, \xi_{x^n}^{\hat{A}_i \hat{C}_i \hat{B}_i R_i}\right) \geq 1 - \epsilon. \qquad (6.2)$$

By the monotonicity of the fidelity under the partial trace (over $X_{[n]\smallsetminus i}$), this implies the easier to verify condition

$$\widetilde{F} := \frac{1}{n} \sum_{i=1}^n F\left(\omega^{ACBRX}, \xi^{\hat{A}_i \hat{C}_i \hat{B}_i R_i X_i}\right) \geq 1 - \epsilon, \qquad (6.3)$$

where $\xi^{\hat{A}_i \hat{C}_i \hat{B}_i R_i X_i} = \mathrm{Tr}_{[n]\smallsetminus i} \xi^{\hat{A}^n \hat{C}^n \hat{B}^n R^n X^n}$, and '$\mathrm{Tr}_{[n]\smallsetminus i}$' denotes the partial trace over all systems with indices in $[n] \smallsetminus i$.

Conversely, Eq. 6.3 can be shown to imply the criterion 6.2 with $(1-\epsilon)^2 \geq 1 - 2\epsilon$ on the right hand side. Indeed, note that

$$F\left(\omega^{ACBRX}, \xi^{\hat{A}_i \hat{C}_i \hat{B}_i R_i X_i}\right)$$
$$= \sum_{x_i} p(x_i) F\left(\psi_{x_i}^{ACBR}, \sum_{x_{[n]\smallsetminus i}} p(x_{[n]\smallsetminus i}) \xi_{x^n}^{\hat{A}_i \hat{C}_i \hat{B}_i R_i}\right).$$

Thus, by the convexity of the square function and Jensen's inequality,

$$(1-\epsilon)^2 \leq \left(\frac{1}{n} \sum_{i=1}^n F\big(\omega^{ACBRX}, \xi^{\hat{A}_i \hat{C}_i \hat{B}_i R_i X_i}\big)\right)^2$$
$$\leq \frac{1}{n} \sum_{i=1}^n \sum_{x^n} p(x^n) F\left(\psi_{x_i}^{ACBR}, \xi_{x^n}^{\hat{A}_i \hat{C}_i \hat{B}_i R_i}\right),$$

and the last line is the left hand side of Eq. 6.2.

Correspondingly, we say Q_b and Q_c are an asymptotically achievable block-error rate and an asymptotically achievable per-copy-error rate, respectively, if for all n there exist codes such that the block fidelity and per-copy fidelity converge to 1, and the quantum rate converges to Q_b and Q_c, respectively. Because of the above demonstrated relations $\widetilde{F}^2 \leq \overline{F} \leq \widetilde{F}$ it doesn't matter which of the two version of per-copy fidelity we take.

According to Stinespring's theorem [55], the encoding and decoding CPTP maps $\mathcal{E}$ and $\mathcal{D}$ can be dilated respectively to the isometries $U_{\mathcal{E}} : A^n C^n A_0 \hookrightarrow M \hat{C}^n W_n$ and $U_{\mathcal{D}} : M B^n B_0 \hookrightarrow \hat{A}^n \hat{B}^n V_n$, with W_n and V_n as the environment systems of the encoder and decoder, respectively.

6.3   Main Results

In Theorem 6.1 we obtain the main results of this chapter concerning optimal (minimum) block-error rate Q_b^* and optimal per-copy-error rate Q_c^*. These rates are expressed in terms of the following single-letter function.

Definition 6.1. *For a state* $\omega^{ACBRX} = \sum_x p(x)|\psi_x\rangle\langle\psi_x|^{ACBR} \otimes |x\rangle\langle x|^X$ *and* $\epsilon \geq 0$ *define:*

$$Q(\epsilon) := \inf \frac{1}{2} I(Z : RXX'|B)_\sigma \text{ over CPTP maps}$$
$$\mathcal{E}_\epsilon : AC \to Z\hat{C} \text{ and } \mathcal{D}_\epsilon : ZB \to \hat{A}\hat{B} \text{ s.t.}$$
$$F(\omega^{ACBRX}, \xi^{\hat{A}\hat{C}\hat{B}RX}) \geq 1 - \epsilon,$$

where

$$\sigma^{Z\hat{C}BRX} := (\mathcal{E}_\epsilon \otimes \mathrm{id}_{BRX})\omega^{ACBRX} = \sum_x p(x)\sigma_x^{Z\hat{C}BR} \otimes |x\rangle\langle x|^X,$$
$$\xi^{\hat{A}\hat{C}\hat{B}RX} := (\mathcal{D}_\epsilon \otimes \mathrm{id}_{\hat{C}RX})\sigma^{Z\hat{C}BRX} = \sum_x p(x)\xi_x^{\hat{A}\hat{C}\hat{B}R} \otimes |x\rangle\langle x|^X.$$

Moreover, define $\widetilde{Q}(0) := \lim_{\epsilon \to 0+} Q(\epsilon)$.

The function $Q(\epsilon)$ is defined for the specific source ω^{ACBRX}; this dependency is dropped to simplify the notation.

Theorem 6.1. *The minimum asymptotically achievable rate with per-copy error is*

$$Q_c^* = \widetilde{Q}(0).$$

Instead, the minimum asymptotically achievable rate with block error is bounded from above and below as follows:

$$\widetilde{Q}(0) \leq Q_b^* \leq Q(0).$$

Proof. We prove the achievability here and leave the converse proof to the next section.

Let $U_0 : AC \hookrightarrow Z\hat{C}W$ and $\widetilde{U}_0 : ZB \hookrightarrow \hat{A}\hat{B}V$ be respectively the isometric extension of the CPTP maps $\mathcal{E}_0$ and $\mathcal{D}_0$ in Definition 6.1 with fidelity 1 (i.e. $\epsilon = 0$). To achieve the block-error rate $Q_b = Q(0)$, Alice applies the isometry U_0, after which the purified state shared between the parties is

$$|\sigma_0\rangle^{Z\hat{C}WBRXX'} = \sum_x \sqrt{p(x)} |\sigma_0(x)\rangle^{Z\hat{C}WBR} \otimes |x\rangle^X \otimes |x\rangle^{X'}.$$

95

Then the parties apply the QSR protocol to many copies of the above source where Alice sends system M to Bob and systems $\hat{C}$ and W are her side information. The rate achieved by the QSR protocol is

$$Q_b = \frac{1}{2} I(Z : RXX'|B)_{\sigma_0}.$$

After executing the QSR protocol, Bob has Z^n, and the state shared between the parties is $\hat{\sigma}_0^{Z^n \hat{C}^n W^n B^n R^n X^n X'^n}$, which satisfies the following entanglement fidelity:

$$F\left((\sigma_0^{Z\hat{C}WBRXX'})^{\otimes n}, \hat{\sigma}_0^{Z^n \hat{C}^n W^n B^n R^n X^n X'^n}\right) \to 1, \qquad (6.4)$$

as $n \to \infty$. Then, Bob applies to each system the CPTP map $\mathcal{D}_0 : ZB \longrightarrow \hat{A}\hat{B}$. Due to the monotonicity of the fidelity under CPTP maps, we obtain from Eq. (6.4)

$$F\left((\mathcal{D}_0^{\otimes n} \otimes \mathrm{id})(\sigma_0^{Z\hat{C}BRX})^{\otimes n}, (\mathcal{D}_0^{\otimes n} \otimes \mathrm{id})\hat{\sigma}_0^{Z^n \hat{C}^n B^n R^n X^n}\right) \to 1 \qquad (6.5)$$

as $n \to \infty$, where the identity channel id acts on systems $\hat{C}^n R^n X^n$. Notice that by the definition of $\mathcal{D}_0$,

$$(\omega^{ACBRX})^{\otimes n} = (\mathcal{D}_0^{\otimes n} \otimes \mathrm{id}_{\hat{C}^n R^n X^n})(\sigma_0^{Z\hat{C}BRX})^{\otimes n}.$$

Thus, the block fidelity criterion of Eq. (6.1) holds.

Now, let $U_\epsilon : AC \hookrightarrow Z\hat{C}W$ and $\widetilde{U}_\epsilon : ZB \hookrightarrow \hat{A}\hat{B}V$ be respectively the isometric extension of the CPTP maps $\mathcal{E}_\epsilon$ and $\mathcal{D}_\epsilon$ in Definition 6.1 with fidelity $1 - \epsilon$. To achieve the per-copy-error rate $Q_c^\star$, to each copy of the source Alice applies the isometry U_ϵ. Then the purified state shared between the parties is

$$|\sigma_\epsilon\rangle^{Z\hat{C}WBRXX'} = \sum_x \sqrt{p(x)}\, |\sigma_\epsilon(x)\rangle^{Z\hat{C}WBR} \otimes |x\rangle^X \otimes |x\rangle^{X'}.$$

The parties apply the QSR protocol to many copies of the above source where Alice sends system Z to Bob and systems $\hat{C}$ and W are her side information. The rate achieved by the QSR protocol is

$$Q_c = \frac{1}{2} I(M : RXX'|B)_{\sigma_\epsilon}.$$

After executing the QSR protocol, Bob has Z^n, and the state shared between the parties is $\hat{\sigma}_\epsilon^{Z^n \hat{C}^n W^n B^n R^n X^n X'^n}$, which satisfies the following entanglement fidelity:

$$F\left((\sigma_\epsilon^{Z\hat{C}WBRXX'})^{\otimes n}, \hat{\sigma}_\epsilon^{Z^n \hat{C}^n W^n B^n R^n X^n X'^n}\right) \to 1$$

as $n \to \infty$. Due to monotonicity of the fidelity under partial trace, we obtain the per-copy fidelity,

$$F(\sigma_\epsilon^{Z\hat{C}BRX}, \hat{\sigma}_\epsilon^{Z_i\hat{C}_iB_iR_iX_i}) \to 1, \tag{6.6}$$

for all $i \in [n]$ and $n \to \infty$. Then, to each system i, Bob applies the CPTP map $\mathcal{D}_\epsilon$. We obtain

$$F\Big((\mathcal{D}_\epsilon \otimes \mathrm{id}_{\hat{C}RX})\sigma_\epsilon^{Z\hat{C}BRX}, (\mathcal{D}_\epsilon \otimes \mathrm{id}_{\hat{C}RX})\hat{\sigma}_\epsilon^{Z_i\hat{C}_iB_iR_iX_i}\Big) \to 1 \tag{6.7}$$

for all $i \in [n]$ and $n \to \infty$, which follows from Eq. (6.6) due to monotonicity of the fidelity under CPTP maps. On the other hand, the state $\xi_\epsilon^{\hat{A}\hat{C}\hat{B}RX} = (\mathcal{D}_\epsilon \otimes \mathrm{id}_{\hat{C}RX})\sigma_\epsilon^{Z\hat{C}BRX}$ has high fidelity with the original source state, directly from the definition of $\mathcal{D}_\epsilon$:

$$F(\xi_\epsilon^{\hat{A}\hat{C}\hat{B}RX}, \omega^{ACBRX}) \to 1.$$

Therefore, from the above fidelity and Eq. (6.7) we obtain

$$F\Big(\omega^{ACBRX}, (\mathcal{D}_\epsilon \otimes \mathrm{id}_{\hat{C}RX})\hat{\sigma}_\epsilon^{M_i\hat{C}_iB_iR_iX_i}\Big) \to 1$$

for all $i \in [n]$ and $n \to \infty$, which satisfies the per-copy fidelity criterion in Eq. (6.3). $\blacksquare$

Now, we define a new single-letter function which then we use to obtain simplified rates in Lemma 6.1 and Corollary 6.1 which both are proved in [51].

Definition 6.2. *For a state $\omega^{ACBRX} = \sum_x p(x)|\psi_x\rangle\langle\psi_x|^{ACBR} \otimes |x\rangle\langle x|^X$ and $\epsilon \geq 0$ define:*

$$K_\epsilon(\omega) := \sup I(W : X|\hat{C})_\sigma \text{ over isometries}$$
$$U : AC \to Z\hat{C}W \text{ and } \widetilde{U} : ZB \to \hat{A}\hat{B}V \text{ s.t.}$$
$$F(\omega^{ACBRX}, \xi^{\hat{A}\hat{C}\hat{B}RX}) \geq 1 - \epsilon,$$

where

$$\sigma^{Z\hat{C}WBRX} := (U \otimes \mathbb{1}_{BRX})\omega^{ACBRX}(U \otimes \mathbb{1}_{BRX})^\dagger$$
$$= \sum_x p(x)|\sigma_x\rangle\langle\sigma_x|^{Z\hat{C}WBR} \otimes |x\rangle\langle x|^X,$$
$$\xi^{\hat{A}\hat{C}\hat{B}WVRX} := (\widetilde{U} \otimes \mathbb{1}_{\hat{C}WRX})\sigma^{Z\hat{C}WBRX}(\widetilde{U} \otimes \mathbb{1}_{\hat{C}WRX})^\dagger$$
$$= \sum_x p(x)|\xi_x\rangle\langle\xi_x|^{\hat{A}\hat{C}\hat{B}WVR} \otimes |x\rangle\langle x|^X,$$
$$\xi^{\hat{A}\hat{C}\hat{B}RX} := \mathrm{Tr}_{VW}\,\xi^{\hat{A}\hat{C}\hat{B}WVRX}.$$

Moreover, define $\widetilde{K}_0 := \lim_{\epsilon \to 0+} K_\epsilon(\omega)$.

97

Remark 6.1. *Definition 6.2 directly implies that $K_0(\omega) \leq \widetilde{K}_0(\omega)$ because $K_\epsilon(\omega)$ is a non-decreasing function of ϵ. Furthermore, $K_0(\omega)$ can be strictly positive, for example, for a source with trivial system C where $\psi_x^A \psi_{x'}^A = 0$ holds for $x \neq x'$, we obtain $K_0(\omega) = S(X)$. This follows because Alice can measure her system and obtain the value of X and then copy this classical information to the register W.*

Lemma 6.1. *The rate $\widetilde{Q}(0)$ is lower bounded as:*

$$\begin{aligned}
\widetilde{Q}(0) &\geq \frac{1}{2}\left(S(A|B) + S(A|C)\right) - \frac{1}{2}\widetilde{K}_0 \\
&= \frac{1}{2}I(A:RXX'|B)_\omega - \frac{1}{2}\widetilde{K}_0,
\end{aligned}$$

where the above conditional mutual information is precisely the communication rate of QSR for the purified source

$$|\omega\rangle^{ACBRXX'} = \sum_x \sqrt{p(x)}\, |\psi_x\rangle^{ACBR} \otimes |x\rangle^X \otimes |x\rangle^{X'}. \qquad (6.8)$$

Moreover, if system C is trivial, then $\widetilde{Q}(0)$ is equal to this lower bound.

Definition 6.3 (Barnum *et al.* [65]). *An ensemble $\mathcal{E} = \{p(x), |\psi_x\rangle\langle\psi_x|^{ACBR}\}_{x \in \mathcal{X}}$ of pure states is called* reducible *if its states fall into two or more orthogonal subspaces. Otherwise the ensemble $\mathcal{E}$ is called* irreducible. *We apply the same terminology to the source state ω^{ACBRX}.*

Corollary 6.1. *For an irreducible source ω^{ACBRX}, $K_0 = \widetilde{K}_0 = 0$. Hence, the optimal asymptotically achievable per-copy-error rate and block-error rate are equal and*

$$Q_c^* = Q_b^* = \frac{1}{2}\left(S(A|C) + S(A|B)\right).$$

6.4 Converse

In this section, we first show some properties of the function $Q(\epsilon)$, which then we use to prove the converse for Theorem 6.1.

Lemma 6.2. *For $0 \leq \epsilon \leq 1$, $Q(\epsilon)$ is a monotonically non-increasing, convex function of ϵ. Consequently, for $0 < \epsilon < 1$ it is also continuous.*

Proof. The monotonicity directly follows from the definition. For the convexity, we verify Jensen's inequality, that is we start with maps $\mathcal{E}_1, \mathcal{D}_1$ eligible for error ϵ_1 with the output state $\xi_1^{\hat{A}\hat{C}\hat{B}RX}$, and $\mathcal{E}_2, \mathcal{D}_2$ eligible for error ϵ_2 with the output state $\xi_2^{\hat{A}\hat{C}\hat{B}RX}$, and $0 \leq p \leq 1$. By embedding into larger Hilbert spaces if necessary, we can w.l.o.g. assume that the maps act on the same systems for $i = 1, 2$. We define the following two maps:

$$\mathcal{E}(\rho) := p\mathcal{E}_1(\rho) \otimes |1\rangle\langle 1|^{Z'} + (1-p)\mathcal{E}_2(\rho) \otimes |2\rangle\langle 2|^{Z'},$$
$$\mathcal{D}(\rho) := \mathcal{D}_1(\langle 1|^{Z'} \rho |1\rangle^{Z'}) + \mathcal{D}_2(\langle 2|^{Z'} \rho |2\rangle^{Z'}).$$

They evidently realise the output state $\xi^{\hat{A}\hat{C}\hat{B}RX} = p\xi_1^{\hat{A}\hat{C}\hat{B}RX} + (1-p)\xi_2^{\hat{A}\hat{C}\hat{B}RX}$ with the following fidelity:

$$F(\omega^{ACBRX}, \xi^{\hat{A}\hat{C}\hat{B}RX})$$
$$= F(\omega^{ACBRX}, p\xi_1^{\hat{A}\hat{C}\hat{B}RX} + (1-p)\xi_2^{\hat{A}\hat{C}\hat{B}RX})$$
$$\geq pF(\omega^{ACBRX}, \xi_1^{\hat{A}\hat{C}\hat{B}RX}) + (1-p)F(\omega^{ACBRX}, \xi_2^{\hat{A}\hat{C}\hat{B}RX})$$
$$\geq 1 - (p\epsilon_1 + (1-p)\epsilon_2),$$

where the third line is due to simultaneous concavity of the fidelity in both arguments. The last line follows by the definitions of the states ξ_1 and ξ_2. Therefore, the maps $\mathcal{E}$ and $\mathcal{D}$ yield a fidelity of at least $1 - (p\epsilon_1 + (1-p)\epsilon_2) =: 1 - \epsilon$. Thus,

$$Q(\epsilon) \leq I(ZZ' : RXX'|B)_\xi$$
$$= pI(Z : RXX'|B)_{\xi_1} + (1-p)I(Z : R|B)_{\xi_2},$$

and taking the infimum over maps $\mathcal{E}_i, \mathcal{D}_i$ shows convexity.

The continuity statement follows from a mathematical folklore fact, stating that any real-valued function that is convex on an interval, is continuous on the interior of the interval. ∎

Proof of Theorem 6.1 (converse). We prove the converse for the per-copy fidelity criterion, therefore, the same converse bound holds for the block fidelity criterion as well. Consider a block length n code per-copy fidelity $1 - \epsilon$. The number of qubits, $\log|M|$, can be lower bounded as follows, with respect to the encoded state $(\mathcal{E} \otimes \mathrm{id}_{B_0 B^n R^n X^n X'^n})\omega^{A^n C^n B^n R^n X^n X'^n} \otimes \Phi_K^{A_0 B_0}$ of

99

the purified source:

$$
\begin{aligned}
2\log|M| &\geq 2S(M)\\
&\geq I(M:R^nX^nX'^n|B^nB_0)\\
&= I(\underbrace{MB_0}_{Z}:R^nX^nX'^n|B^n) - I(B_0:R^nX^nX'^n|B^n)\\
&= I(Z:R^nX^nX'^n|B^n)\\
&= \sum_{i=1}^{n} I(Z:R_iX_iX'_i|B^nR_{<i}X_{<i}X'_{<i})\\
&\qquad + \sum_{i=1}^{n} I(R_{<i}X_{<i}X'_{<i}B_{[n]\setminus i}:R_iX_iX'_i|B_i)\\
&= \sum_{i=1}^{n} I(ZR_{<i}X_{<i}X'_{<i}B_{[n]\setminus i}:R_iX_iX'_i|B_i)\\
&\geq \sum_{i=1}^{n} I(\underbrace{ZB_{[n]\setminus i}}_{Z_i}:R_iX_iX'_i|B_i),
\end{aligned}
$$

where in the first two inequalities we use standard entropy inequalities; the equation in the third line is due to the chain rule, and the second conditional information is 0 because B_0 is independent of B^nR^n; the fourth line introduces a new register Z, noting that the encoding together with the entangled state defines a CPTP map $\mathcal{E}_0 : A^n \to Z\hat{C}^n$, via $\mathcal{E}_0(\rho) = (\mathcal{E}\otimes \mathrm{id}_{B_0})(\rho \otimes \Phi_K^{A_0B_0})$; in the fifth we use the chain rule iteratively, and in the second term we introduce, each summand is 0 because for all i, $R_{<i}B_{[n]\setminus i}$ is independent of R_iB_i; in the sixth line we use again the chain rule for all i, and the last line is due to data processing.

Now, for the i-th copy of the source $\omega^{A_iC_iB_iR_iX_i}$, we define maps $\mathcal{E}_i : A_iC_i \to Z_i\hat{C}_i$ and $\mathcal{D}_i : B_iZ_i \to \hat{A}_i\hat{B}_i$, as follows:

$\mathcal{E}_i$: Alice tensors her system A_i with a dummy state $\omega^{\otimes[n]\setminus i}$ and with $\Phi_K^{A_0B_0}$ (note that all systems are in her possession). Then she applies $\mathcal{E} : A^nC^nA_0 \to M\hat{C}^n$, and sends $Z_i := MB_0B_{[n]\setminus i}$ to Bob, while keeping $\hat{A}_i\hat{C}_i$. All other systems, i.e. $\hat{A}_{[n]\setminus i}\hat{C}_{[n]\setminus i}R_{[n]\setminus i}X_{[n]\setminus i}$, are trashed.

$\mathcal{D}_i$: Bob applies $\mathcal{D}$ to $Z_iB_i = MB_0B^n$ and keeps $\hat{A}_i\hat{B}_i$, trashing the rest $\hat{A}_{[n]\setminus i}\hat{B}_{[n]\setminus i}$.

By definition, the output state

$$
\zeta^{\hat{A}_i\hat{C}_i\hat{B}_iR_iX_i} = (\mathcal{D}_i \otimes \mathrm{id}_{\hat{A}_i\hat{C}_iR_iX_i}) \circ (\mathcal{E}_i \otimes \mathrm{id}_{B_iR_iX_i})\omega^{A_iC_iB_iR_iX_i}
$$

equals $\xi^{\hat{A}_i \hat{C}_i \hat{B}_i R_i X_i}$ which has fidelity $1 - \epsilon_i$ with the source ω^{ACBRX}, and the fidelity for all copies satisfy $\frac{1}{n} \sum_i (1 - \epsilon_i) \geq 1 - \epsilon$. Thus, we obtain, with respect to the states $(\mathcal{E}_i \otimes \mathrm{id}_{B_i R_i X_i X_i'}) \omega^{A_i C_i B_i R_i X_i X_i'}$

$$
\begin{aligned}
\frac{1}{n} \log |M| &\geq \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} I(Z_i : R_i X_i X_i' | B_i) \\
&\geq \frac{1}{n} \sum_{i=1}^{n} Q(\epsilon_i) \geq Q\left(\frac{1}{n} \sum_{i=1}^{n} \epsilon_i \right) \geq Q(\epsilon),
\end{aligned}
$$

continuing from before, then by definition of $Q(\epsilon_i)$ since the pair $(\mathcal{E}_i, \mathcal{D}_i)$ results in fidelity $1 - \epsilon_i$, in the next inequality by convexity and finally by monotonicity of $Q(\epsilon)$ (Lemma 6.2). By the taking the limit of $\epsilon \to 0$ and $n \to \infty$, the claim follows. ∎

6.5   Discussion

We considered a variant of the quantum state redistribution task, where pure multipartite states from an ensemble are distributed between an encoder, a decoder and a reference system. We distinguish two figures of merit for the information processing, per-copy fidelity and block fidelity, and define the corresponding quantum communication rates depending on the fidelity criterion, when unlimited entanglement is available. For the per-copy fidelity criterion, we find that the optimal qubit rate of compression is equal to $\widetilde{Q}(0)$ from Definition 6.1, which is bounded from below by the rate of the conventional QSR task minus the limit of the single-letter non-negative function $\widetilde{K}_0$ from Definition 6.2:

$$
\widetilde{Q}(0) \geq \frac{1}{2} (S(A|B) + S(A|C)) - \frac{1}{2} \widetilde{K}_0 = \frac{1}{2} I(A : RXX'|B)_\omega - \frac{1}{2} \widetilde{K}_0,
$$

where the conditional mutual information is the rate of QSR for the purified source in Eq. (6.8). This lower bound is tight if system C is trivial (state merging scenario).

For the block fidelity criterion, we have found converse and achievability bounds:

$$
\widetilde{Q}(0) \leq Q_b \leq Q(0).
$$

The two bounds would match if we knew that the function $Q(\epsilon)$ were continuous at $\epsilon = 0$. However, we don not know this; for one thing, one cannot use compactness to show continuity because the output system W in Definition 6.2 is as priori unbounded.

101

For irreducible sources though, we show here $K_0 = \widetilde{K}_0 = 0$, which implies that the purified source model and the ensemble model lead to the same compression rate. For reducible sources the information that the encoder can obtain about the classical variable of the ensemble, i.e. system X, is effectively used as side information to achieve a smaller compression rate. Thus we reproduce the result of [32, Thm. III.3], which was proven only for irreducible product state ensembles.

There are other sources for which we know $K_0 = \widetilde{K}_0 = 0$ to hold. First, the "generic" sources in [49, Thm. 11], where it is shown that the function $\widetilde{I}_0 = 0$; this function is a special case of the function $\widetilde{K}_0$. Indeed, the source there is described by an ensemble $\{p(x), |\psi_x\rangle^{AR}|x\rangle^B\}$, which is always completely reducible, but generically the reduced states ψ_x^A have pairwise overlapping support, which is the condition under which vanishing $\widetilde{K}_0$ is shown. Secondly, the ensemble of four Bell states considered in [32, Thm. IV.1], $\{p(ij), |\Phi_{ij}\rangle^{AB}\}_{i,j=0,1}$, where the side information system C and the reference system R are trivial; for this source, the mutual information between Alice's system and the classical system X is zero, i.e. $I(A:X) = 0$. Thus, due to data processing inequality, we have $I(W:X) \le I(A:X) = 0$. Our main result reproduces the achievable rate $\frac{1}{2}H(p)$, and also the optimality, by very different, and somewhat more natural methods.

There are other special cases of the source model of this chapter that have been previously studied in the literature for which $K_0 > 0$ or at least $\widetilde{K}_0 > 0$. For instance in the source of [29], where Alice's system is classical with $A = X$ and system C is trivial, one can observe that $K_0 = S(X)$ holds. The rate we get is $Q^* = \frac{1}{2}S(X|B)$ under either error criterion, half of the quantity reported in [29] because of the free entanglement in our model, which allows for dense coding. Furthermore, the visible variant of Schumacher compression in [30, 61], where Alice's side information system is classical with $C = X$, the function has the value $K_0 = S(X)$, and the optimal rate is $Q^* = \frac{1}{2}S(A)$, again half of the optimal rate without entanglement, because we can use remote state preparation and dense coding. A third example is the ensemble $\{\frac{1}{3}, |\psi_i\rangle^A|\phi_i\rangle^B\}_{i=1}^3$ from [32, Sec. V.A], which is reducible, but where the reduced ensembles on systems A and B are both irreducible; it is shown there that the optimal compression rate is strictly smaller than $(S(A) + S(A|B))/2$.

Finally, recall that in our definition of the compression task we have assumed that the encoder and decoder share free entanglement. This was motivated so as to make a smoother connection to QSR. However, it is not known whether the pre-shared entanglement is always necessary to achieve the corresponding quantum rates. There are certainly cases where QSR does

not require prior entanglement, such as when Alice's side information C is trivial, which would carry over to our setting whenever $K_0 = \widetilde{K}_0 = 0$, for instance for an irreducible ensemble. More generally, in future work we plan to consider the trade-off between the quantum and entanglement rates.

Part II

Quantum Thermodynamics

Chapter 7

Resource theory of charges and entropy

In this chapter, we consider asymptotically many non-interacting systems with multiple conserved quantities or charges. We generalize the seminal results of Sparaciari, Oppenheim and Fritz [*Phys. Rev. A* 96:052112, 2017] to the case of multiple, in general non-commuting charges. To this aim we formulate a resource theory of thermodynamics of asymptotically many non-interacting systems with multiple conserved quantities or charges. To any quantum state, we associate a vector with entries of the expected charge values and entropy of that state. We call the set of all these vectors the phase diagram of the system, and show that it characterizes the equivalence classes of states under asymptotic unitary transformations that approximately conserve the charges. This chapter is based on the results from [53].

7.1 Resource theory of charges and entropy

Resource theory is a rigorous mathematical framework initially developed to characterize the role of entanglement in quantum information processing tasks. Later the framework was extended to characterize coherence, non-locality, asymmetry and many more, including quantum Shannon theory itself, see [35, 85–97]. The resource theory approach applies also to classical theories. In general, the resource theories have the following common features: (1) a well-defined set of resource-free states, and any states that do not belong to this set has a non-vanishing amount of resource; (2) a well-defined set of resource-free operations, also known as allowed operations, that cannot create or increases resource in a state. These allow one to quantify the resources present in the states or operations and characterize their roles in

the transformations between the states or the operations. In particular, it enables one to define and rigorously bound or even determine various resource measures; determine which states can be transformed to the others using allowed operation; how the property of states may be changed, and how these changes are bounded under the allowed operations, etc.

A system in our resource theory is a quantum system Q with a finite-dimensional Hilbert space (denoted Q, too, without danger of confusion), together with a Hamiltonian $H = A_1$ and other quantities ("charges") $A_2, \ldots, A_c$, all of which are Hermitian operators that do not necessarily commute with each other. We consider composition of n non-interacting systems, where the Hilbert space of the *composite* system Q^n is the tensor product $Q^{\otimes n} = Q_1 \otimes \cdots \otimes Q_n$ of the Hilbert spaces of the *individual* systems, and the j-th charge of the composite system is the sum of charges of individual systems as follows,

$$A_j^{(n)} = \sum_{i=1}^{n} \mathbb{1}^{\otimes(i-1)} \otimes A_j \otimes \mathbb{1}^{\otimes(n-i)}, \quad j = 1, 2, \ldots, c. \qquad (7.1)$$

For ease of notation, we will write throughout $A_j^{[Q_i]} = \mathbb{1}^{\otimes(i-1)} \otimes A_j \otimes \mathbb{1}^{\otimes(n-i)}$.

We wish to build a resource theory where the objects are states on a quantum system, which are transformed under thermodynamically meaningful operations. To any quantum state ρ is assigned the point $(\underline{a}, s) = (a_1, \ldots, a_c, s) = \left(\operatorname{Tr} \rho A_1, \ldots, \operatorname{Tr} \rho A_c, S(\rho) \right) \in \mathbb{R}^{c+1}$, which is an element in the *phase diagram* that has been originally introduced, for $c = 1$, as energy-entropy diagram in [98]; there it is shown, for a system where energy is the only conserved quantity, that the diagram is a convex set. In the case of commuting multiple conserved quantities, the charge-entropy diagram has been generalised and further investigated in [54]. Note that the set of all these vectors, denoted $\mathcal{P}^{(1)}$, is not in general convex (unless the quantities commute pairwise). An example is a qubit system with charges σ_x, σ_y and σ_z where charge values uniquely determine the state as a linear function of the $\operatorname{tr} \rho \sigma_i$, hence the entropy, while the von Neumann entropy itself is well-known to be strictly concave.

Moreover, the set of these points for a composite system with charges $A_1^{(n)}, \ldots, A_c^{(n)}$, which we denote $\mathcal{P}^{(n)}$ contains, but is not necessarily equal to $n\mathcal{P}^{(1)}$ (which however is true for commuting charges). Namely, consider the point $g = \left(\frac{1}{2} \operatorname{Tr} (\rho_1 + \rho_2) A_1, \ldots, \frac{1}{2} \operatorname{Tr} (\rho_1 + \rho_2) A_c, \frac{1}{2} S(\rho_1) + \frac{1}{2} S(\rho_2) \right)$, which does not necessarily belong to $\mathcal{P}^{(1)}$ but belongs to its convex hull; however, $2g \in \mathcal{P}^{(2)}$ due to the state $\rho_1 \otimes \rho_2$. Therefore, we consider the convex hull of
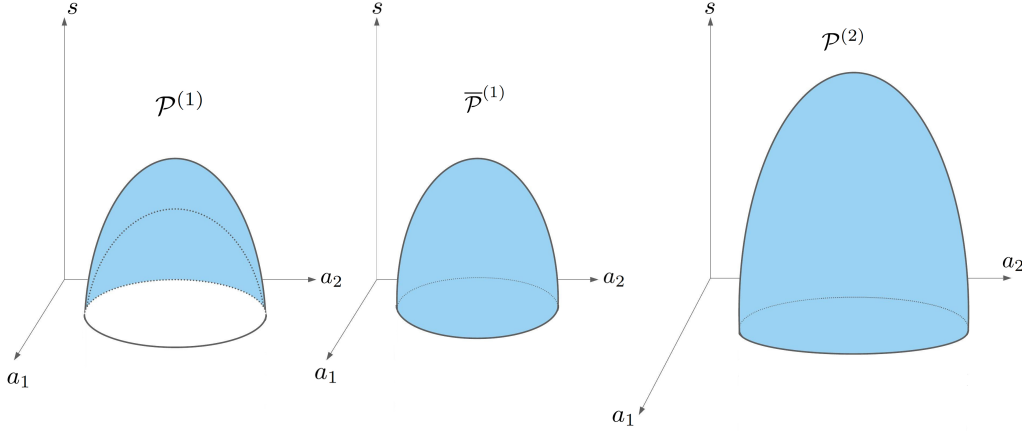
Figure 7.1: Schematic of the phase diagrams $\mathcal{P}^{(1)}$, $\mathcal{P}^{(2)}$ and $\overline{\mathcal{P}}$. As seen, $\mathcal{P}^{(1)}$ is not convex, and there is a hole inside the diagram.

the set $\mathcal{P}^{(1)}$ and call it the *phase diagram* of the system, denoted

$$\overline{\mathcal{P}} \equiv \overline{\mathcal{P}}^{(1)} := \left\{ \left(\sum_i p_i \operatorname{Tr} \rho_i A_1, \ldots, \sum_i p_i \operatorname{Tr} \rho_i A_c, \sum_i p_i S(\rho_i) \right) : 0 \le p_i \le 1, \ \sum_i p_i = 1 \right\}.$$
(7.2)

The interpretation is that the objects of our resource theory are ensembles of states $\{p_i, \rho_i\}$, rather than single states.

We define the *zero-entropy diagram* and *max-entropy diagram*, respectively, as the sets

$$\overline{\mathcal{P}}_0^{(1)} = \{ (\underline{a}, 0) : \operatorname{Tr} \rho A_j = a_j \text{ for a state } \rho \},$$
$$\overline{\mathcal{P}}_{\max}^{(1)} = \left\{ \left(\underline{a}, S(\tau(\underline{a})) \right) : \operatorname{Tr} \rho A_j = a_j \text{ for a state } \rho \right\},$$

where $\tau(\underline{a})$ is the unique state maximising the entropy among all states with charge values $\operatorname{Tr} \rho A_j = a_j$ for all j, which is called generalized thermal state, or generalized Gibbs state, or also generalized grand canonical state [99]. Note that, as a linear image of the compact convex set of states, the zero-entropy diagram is compact and convex. We similarly define the set $\mathcal{P}^{(n)}$, the phase diagram $\overline{\mathcal{P}}^{(n)}$, zero-entropy diagram $\overline{\mathcal{P}}_0^{(n)}$ and max-entropy diagram $\overline{\mathcal{P}}_{\max}^{(n)}$ for the composition of n systems with charges $A_1^{(n)}, \ldots, A_c^{(n)}$.

Lemma 7.1. *For an individual and composite systems with charges A_j and $A_j^{(n)}$, respectively, we have:*

1. *$\overline{\mathcal{P}}^{(n)}$, for $n \ge 1$, is a compact and convex subset of $\mathbb{R}^{c+1}$.*

109

2. $\overline{\mathcal{P}}^{(n)}$, for $n \geq 1$, is the convex hull of the union $\overline{\mathcal{P}}_0^{(n)} \cup \overline{\mathcal{P}}_{\max}^{(n)}$, of the zero-entropy diagram and the max-entropy diagram.

3. $\overline{\mathcal{P}}^{(n)} = n\overline{\mathcal{P}}^{(1)}$ for all $n \geq 1$.

4. $\mathcal{P}^{(n)}$ is convex for all $n \geq 2$, and indeed $\mathcal{P}^{(n)} = \overline{\mathcal{P}}^{(n)} = n\overline{\mathcal{P}}^{(1)}$.

5. Every point of $\mathcal{P}^{(n)}$ is realised by a suitable tensor product state $\rho_1 \otimes \cdots \otimes \rho_n$, for all $n \geq d$.

6. All points $\big(\underline{a}, S(\tau(\underline{a}))\big) \in \overline{\mathcal{P}}_{\max}$ are extreme points of $\overline{\mathcal{P}}$.

Proof. 1. The phase diagram is convex by definition. Further, $\mathrm{Tr}\,\rho A_j^{(n)}$ and $S(\rho)$ are continuous functions defined on the set of quantum states which is a compact set; hence, the set $\mathcal{P}^{(n)}$ is also a compact set. The cxonvex hull of a finite-dimensional compact set is compact, so the phase diagram is a compact set.

2. Any point in the phase diagram according to the definition is a convex combination of the form

$$(a_1, \ldots, a_c, s) = \left(\sum_i p_i \mathrm{Tr}\,(\rho_i A_1), \ldots, \sum_i p_i \mathrm{Tr}\,(\rho_i A_c), \sum_i p_i S(\rho_i) \right).$$

The point $(a_1, \ldots, a_c, 0)$ belongs to $\overline{\mathcal{P}}_0^{(1)}$ because the state $\rho = \sum_i p_i \rho_i$ has charge values $a_1, \ldots, a_c$. Moreover, the state with charge values $a_1, \ldots, a_c$ of maximum entropy is the generalized thermal state $\tau(\underline{a})$, so we have

$$S(\tau(\underline{a})) \geq S(\rho) \geq \sum_i p_i S(\rho_i),$$

where the second inequality is due to concavity of the entropy. Therefore, any point $(\underline{a}, s)$ can be written as the convex combination of the points $(\underline{a}, 0)$ and $(\underline{a}, S(\tau(\underline{a}))$.

3. Due to item 2, it is enough to show that $\overline{\mathcal{P}}_0^{(n)} = n\overline{\mathcal{P}}_0^{(1)}$, and $\overline{\mathcal{P}}_{\max}^{(n)} = n\overline{\mathcal{P}}_{\max}^{(1)}$. The former follows from the definition. The latter is due to the fact that the thermal state for a composite system is the tensor power of the thermal state of the individual system.

4. Let $\tau(\underline{a}) = \sum_i p_i |i\rangle\langle i|$ be the diagonalization of the generalized thermal state. For $n \geq 2$, define $|v\rangle = \sum_i \sqrt{p_i} |i\rangle^{\otimes n}$. Obviously, the charge values of the states $\tau(\underline{a})^{\otimes n}$ and $|v\rangle\langle v|$ are the same, since they have the same reduced states on the individual systems; thus, there is a pure state for any point in the zero-entropy diagram of the composite system. Now, consider the state

110

$\lambda |v\rangle\langle v| + (1-\lambda)\tau(\underline{a})^{\otimes n}$, which has the same charge values as $\tau(\underline{a})^{\otimes n}$ and $|v\rangle\langle v|$. The entropy $S\big(\lambda |v\rangle\langle v| + (1-\lambda)\tau(\underline{a})^{\otimes n}\big)$ is a continuous function of λ; hence, for any value s between 0 and $S(\tau(\underline{a})^{\otimes n})$, there is a state with the given values and entropy s.

5. For $n \geq d$, it is elementary to see that any state ρ can be decomposed into a uniform convex combination of n pure states, i.e. $\rho = \frac{1}{n}\sum_{i=1}^{n} |\psi_i\rangle\langle\psi_i|$. Observe that the state $\psi^n = |\psi_1\rangle\langle\psi_1| \otimes \cdots \otimes |\psi_n\rangle\langle\psi_n|$ has the same charge values as the state $\rho^{\otimes n}$, but as it is pure it has entropy 0. Further, consider the thermal state τ with the same charge values as ρ, but the maximum entropy consistent with them. Now let $\rho_i := \lambda |\psi_i\rangle\langle\psi_i| + (1-\lambda)\tau$, and observe that $\rho_\lambda^n = \rho_1 \otimes \cdots \rho_n$ has the same charge values as ψ^n, ρ^n and $\tau^{\otimes n}$. Since the entropy $S(\rho_\lambda^n)$ is a continuous function of λ, thus interpolating smoothly between 0 and $nS(\tau)$, there is a tensor product state with the same given charge values and prescribed entropy s in the said interval.

6. This follows from the strict concavity of the von Neumann entropy $S(\rho)$ as a function of the state, which imparts the strict concavity on $\underline{a} \mapsto S(\tau(\underline{a}))$ ∎

The penultimate point of Lemma 7.1 motivates us to define a resource theory where the objects are sequences of states on composite systems of $n \to \infty$ parts. Inspired by [98], the allowed operations in this resource theory are those that respect basic principles of physics, namely entropy and charge conservation. We point out right here, that "physics" in the present context does not necessarily refer to the fundamental physical laws of nature, but to any rule that the system under consideration obeys. It is well-known that quantum operations that preserve entropy for all states are unitaries. The class of unitaries that conserve charges of a system are precisely those that commute with all charges of that system. However, it turns out that these constraints are too strong if imposed literally, when many charges are to be conserved, as it could easily happen that only trivial unitaries are allowed. Our way out is to consider the thermodynamic limit and at the same time relax the allowed operations to approximately entropy and charge conserving ones. As for the former, we couple the composite system to an ancillary system with corresponding Hilbert space $\mathcal{K}$ of dimension $2^{o(n)}$ where restricting the dimension of the ancilla ensures that the *average* entropy of an individual system, that is, entropy of the composite system per n does not change in the limit of large n. Moreover, as for charge conservation, we consider unitaries that preserve the average charges of an individual system, and we allow unitaries that are *almost* commuting with the total charges of the composite system and the ancilla. The precise definition goers as follows:

111

Definition 7.1. *A unitary operation U acting on a composite system coupled to an ancillary system with Hilbert spaces $\mathcal{H}^{\otimes n}$ and $\mathcal{K}$ of dimension $2^{o(n)}$, respectively, is called an* almost-commuting unitary *with the total charges of a composite system and an ancillary system if the operator norm of the normalised commutator for all total charges vanishes asymptotically for large n:*

$$\lim_{n\to\infty} \frac{1}{n}\left\|[U, A_j^{(n)} + A_j']\right\|_\infty =$$

$$\lim_{n\to\infty} \frac{1}{n}\left\|U(A_j^{(n)} + A_j') - (A_j^{(n)} + A_j')U\right\|_\infty = 0 \qquad j = 1, \ldots, c.$$

where $A_j^{(n)}$ and A_j' are respectively the charges of the composite system and the ancilla, such that $\left\|A_j'\right\|_\infty \le o(n)$.

We stress that the definition of almost-commuting unitaries automatically implies that the ancillary system has a relatively small dimension and charges with small operator norm compared to a composite system. The first step in the development of our resource theory is a precise characterisation of which transformations between sequences of product state are pssinble using almost commuting unitaries. To do so, we define *asymptotically equivalent* states as follows:

Definition 7.2. *Two sequences of product states $\rho^n = \rho_1 \otimes \cdots \otimes \rho_n$ and $\sigma^n = \sigma_1 \otimes \cdots \otimes \sigma_n$ of a composite system with charges $A_j^{(n)}$ for $j = 1, \ldots, c$, are called* asymptotically equivalent *if*

$$\lim_{n\to\infty} \frac{1}{n}|S(\rho^n) - S(\sigma^n)| = 0,$$

$$\lim_{n\to\infty} \frac{1}{n}\left|\mathrm{Tr}\,\rho^n A_j^{(n)} - \mathrm{Tr}\,\sigma^n A_j^{(n)}\right| = 0 \text{ for } j = 1, \ldots, c.$$

In other words, two sequences of product states are considered equivalent if their associated points in the normalised phase diagrams $\frac{1}{n}\mathcal{P}^{(n)}$ differ by a sequence converging to 0.

The *asymptotic equivalence theorem* of [98] characterizes feasible state transformations via *exactly* commuting unitaries where energy is the only conserved quantity of a system, showing that it is precisely given by asymptotic equivalence. We prove an extension of this theorem for systems with multiple conserved quantities, by allowing almost-commuting unitaries.

Theorem 7.1 (Asymptotic (approximate) Equivalence Theorem)**.** *Let $\rho^n = \rho_1 \otimes \cdots \otimes \rho_n$ and $\sigma^n = \sigma_1 \otimes \cdots \otimes \sigma_n$ be two sequences of product states of a composite system with charges $A_j^{(n)}$ for $j = 1, \ldots, c$. These two states are asymptotically equivalent if and only if there exist ancillary quantum systems with*

112

corresponding Hilbert space $\mathcal{K}$ of dimension $2^{o(n)}$ and an almost-commuting unitary U acting on $\mathcal{H}^{\otimes n} \otimes \mathcal{K}$ such that

$$\lim_{n \to \infty} \left\| U \rho^n \otimes \omega' U^\dagger - \sigma^n \otimes \omega \right\|_1 = 0,$$

where ω and ω' are states of the ancillary system, and charges of the ancillary system are trivial, $A'_j = 0$.

The *proof* of this theorem is given in Section 7.3, as it relies on a number of technical lemmas, among them a novel construction of approximately microcanonical subspaces (Section 7.2).

By grouping the Q-systems into blocks of k, we do not of course change the physics of our system, except that now in the asymptotic limit we only consider $n = k\nu$ copies of Q, but the state ρ^n is asymptotically equivalent to $\rho^{n+O(1)}$ via almost-commuting unitaries according to Definition 7.1 and Theorem 7.1. But now we consider Q^k with its charge observables $A_j^{(k)}$ as elementary systems, which have many more states than the k-fold product states we began with. Yet, Lemma 7.1 shows that the phase diagram for the k-copy system is simply the rescaled single-copy phase diagram, $\overline{\mathcal{P}}^{(k)} = k\overline{\mathcal{P}}^{(1)}$, and indeed for $k \geq d$, $\mathcal{P}^{(k)} = k\overline{\mathcal{P}}^{(1)}$. This means that we can extend the equivalence relation of asymptotic equivalence and the concomitant Asymptotic Equivalence Theorem (AET) 7.1 to any sequences of states that factor into product states of blocks Q^k, for any integer k, which freedom we shall exploit in our treatment of thermodynamics.

7.2 Approximate microcanonical (a.m.c.) subspace

In this section, we recall the definition of approximate microcanonical (a.m.c.) and give a new proof that it exists for certain explicitly given parameters. For charges A_j and average values v_j, a.m.c. is basically a *common* subspace for the spectral projectors of $A_j^{(n)}$ with corresponding values close to nv_j; that is, a subspace onto which a state projects with high probability if and only if it projects onto the spectral projectors of the charges with high probability. We show in Theorem 7.2 that for a large enough n such a subspace exits. An interesting property of an a.m.c. subspace is that any unitary acting on this subspace is an almost-commuting unitary with charges $A_j^{(n)}$.

Definition 7.3. *An* approximate microcanonical (a.m.c.) subspace*, or more precisely a* $(\epsilon, \eta, \eta', \delta, \delta')$*-approximate microcanonical subspace,* $\mathcal{M}$ *of* $\mathcal{H}^{\otimes n}$*, with projector* P*, for charges* A_j *and values* $v_j = \langle A_j \rangle$ *is one that consists, in a certain precise sense, of exactly the states with "very sharp" values of all the* $A_j^{(n)}$*. Mathematically, the following has to hold:*

1. *Every state* ω *with support contained in* $\mathcal{M}$ *satisfies* $\operatorname{tr} \omega \Pi_j^\eta \geq 1 - \delta$ *for all* j*.*

2. *Conversely, every state* ω *on* $\mathcal{H}^{\otimes n}$ *such that* $\operatorname{tr} \omega \Pi_j^{\eta'} \geq 1 - \delta'$ *for all* j*, satisfies* $\operatorname{tr} \omega P \geq 1 - \epsilon$*.*

Here, $\Pi_j^\eta := \left\{ nv_j - n\eta \Sigma(A_j) \leq A_j^{(n)} \leq nv_j + n\eta \Sigma(A_j) \right\}$ *is the spectral projector of* $A_j^{(n)}$ *of values close to* nv_j*, and* $\Sigma(A) = \lambda_{\max}(A) - \lambda_{\min}(A)$ *is the spectral diameter of the Hermitian* A*, i.e. the diameter of the smallest disc covering the spectrum of* A*.*

Remark 7.1. *It is shown in Theorem 3 of [100] that for every* $\epsilon > c\delta' > 0$*,* $\delta > 0$ *and* $\eta > \eta' > 0$*, and for all sufficiently large* n*, there exists a nontrivial* $(\epsilon, \eta, \eta', \delta, \delta')$*-a.m.c. subspace. However, there are two (related) reasons why one might be not completely satisfied with the argument in [100]: First, the proof uses a difficult result of Ogata [101] to reduce the non-commuting case to the seemingly easier of commuting observables; while this is conceptually nice, it makes it harder to perceive the nature of the constructed subspace. Secondly, despite the fact that the defining properties of an a.m.c. subspace are manifestly permutation symmetric (w.r.t. permutations of the* n *subsystems), the resulting construction does not have this property.*

Here we address both these concerns. Indeed, we shall show by essentially elementary means how to obtain an a.m.c. subspace that is by its definition permutation symmetric.

Theorem 7.2. *Under the previous assumptions, for every* $\epsilon > 2(n+1)^{3d^2} \delta' > 0$*,* $\eta > \eta' > 0$ *and* $\delta > 0$*, for all sufficiently large* n *there exists an approximate microcanonical subspace projector. In addition, the subspace can be chosen to be stable under permutations of the* n *systems:* $U^\pi \mathcal{M} = \mathcal{M}$*, or equivalently* $U^\pi P (U^\pi)^\dagger = P$*, for any permutation* $\pi \in S_n$ *and its unitary action* U^π*.*

More precisely, given $\eta > \eta' > 0$ *and* $\epsilon > 0$*, there exists a* $\alpha > 0$ *such that there is a non-trivial* $(\epsilon, \eta, \eta', \delta, \delta')$*-a.m.c. subspace with*

$$\delta = (c+3)(5n)^{5d^2} e^{-\alpha n} \ \text{ and}$$

$$\delta' = \frac{\epsilon}{2(n+1)^{3d^2}} - (c+3)(5n)^{2d^2} e^{-\alpha n}.$$

Furthermore, we may choose $\alpha = \frac{(\eta - \eta')^2}{8c^2(d+1)^2}$*.*

114

Proof. For $s > 0$, partition the state space $\mathcal{S}(\mathcal{H})$ on $\mathcal{H}$ into

$$\mathcal{C}_s(\underline{v}) = \big\{\sigma : \forall j \mid \mathrm{tr}\,\sigma A_j - v_j| \le s\Sigma(A_j)\big\},$$
$$\mathcal{F}_s(\underline{v}) = \big\{\sigma : \exists j \mid \mathrm{tr}\,\sigma A_j - v_j| > s\Sigma(A_j)\big\} = \mathcal{S}(\mathcal{H}) \smallsetminus \mathcal{C}_s(\underline{v}),$$

which are the sets of states with A_j-expectation values "close" to and "far" from $\underline{v}$. Note that if $\rho \in \mathcal{C}_s(\underline{v})$ and $\sigma \in \mathcal{F}_t(\underline{v})$, $0 < s < t$, then $\|\rho - \sigma\|_1 \ge t - s$.

Choosing the precise values of $s > \eta'$ and $t < \eta$ later, we pick a universal distinguisher $(P, P^\perp)$ between $\mathcal{C}_s(\underline{v})^{\otimes n}$ and $\mathcal{F}_t(\underline{v})^{\otimes n}$, according to Lemma 7.2 below:

$$\forall \rho \in \mathcal{C}_s(\underline{v}) \;\; \mathrm{tr}\,\rho^{\otimes n} P^\perp \le (c+2)(5n)^{2d^2} e^{-\zeta n}, \tag{7.3}$$

$$\forall \sigma \in \mathcal{F}_t(\underline{v}) \;\; \mathrm{tr}\,\sigma^{\otimes n} P \le (c+2)(5n)^{2d^2} e^{-\zeta n}, \tag{7.4}$$

with $\zeta = \frac{(t-s)^2}{2c^2(2d^2+1)}$. Our a.m.c. subspace will be $\mathcal{M} := \mathrm{supp}\,P$; by Lemma 7.2, P and likewise $\mathcal{M}$ are permutation symmetric.

It remains to check the properties of the definition. First, let ω be supported on $\mathcal{M}$. Since we are interested in $\mathrm{tr}\,\omega\Pi_j^\eta$, we may without loss of generality assume that ω is permutation symmetric. Thus, by the "constrained de Finetti reduction" (aka "Postselection Lemma") [102, Lemma 18],

$$\omega \le (n+1)^{3d^2} \int \mathrm{d}\sigma\, \sigma^{\otimes n} F(\omega, \sigma^{\otimes n})^2, \tag{7.5}$$

with a certain universal probability measure $\mathrm{d}\sigma$ on $\mathcal{S}(\mathcal{H})$, and the fidelity $F(\rho, \sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1$ between states. We need the monotonicity of the fidelity under cptp maps, which we apply to the test $(P, P^\perp)$:

$$F(\omega, \sigma^{\otimes n})^2 \le F\big((\mathrm{tr}\,\sigma^{\otimes n}P, 1 - \mathrm{tr}\,\sigma^{\otimes n}P), (1,0)\big)^2 \le \mathrm{tr}\,\sigma^{\otimes n}P,$$

which holds because $\mathrm{tr}\,\omega P = 1$. Thus,

$$\mathrm{tr}\,\omega(\Pi_j^\eta)^\perp \le (n+1)^{3d^2} \int \mathrm{d}\sigma\,\big(\mathrm{tr}\,\sigma^{\otimes n}(\Pi_j^\eta)^\perp\big)(\mathrm{tr}\,\sigma^{\otimes n}P). \tag{7.6}$$

Now we split the integral on the right hand side of Eq. (7.6) into two parts, $\sigma \in \mathcal{C}_t(\underline{v})$ and $\sigma \notin \mathcal{F}_t(\underline{v})$: If $\sigma \in \mathcal{F}_t(\underline{v})$, then by Eq. (7.4) we have

$$\mathrm{tr}\,\sigma^{\otimes n}P \le (c+2)(5n)^{2d^2} e^{-\zeta n}.$$

On the other hand, if $\sigma \in \mathcal{C}_t(\underline{v})$, then because of $t < \eta$ we have

$$\mathrm{tr}\,\sigma^{\otimes n}(\Pi_j^\eta)^\perp \le 2e^{-2(\eta-t)^2 n},$$

which follows from Hoeffding's inequality [103]: Indeed, let Z_ℓ be the i.i.d. random variables obtained by the measurement of A_j on the state σ. They take values in the interval $[\lambda_{\min}(A_j), \lambda_{\max}(A_j)]$, their expectation values satisfy $\mathbb{E}Z_j = \operatorname{tr}\sigma A_j \in [v_j \pm t\Sigma(A_j)]$, while

$$\operatorname{tr}\sigma^{\otimes n}(\Pi_j^\eta)^\perp = \Pr\left\{\frac{1}{n}\sum_\ell Z_\ell \notin [v_j \pm \eta\Sigma(A_j)]\right\}$$

$$\leq \Pr\left\{\frac{1}{n}\sum_\ell Z_\ell \notin [\operatorname{tr}\sigma A_j \pm (\eta - t)\Sigma(A_j)]\right\},$$

so Hoeffding's inequality applies. All taken together, we have

$$\operatorname{tr}\omega(\Pi_j^\eta)^\perp \leq (n+1)^{3d^2}\left((c+2)(5n)^{2d^2}e^{-\zeta n} + 2e^{-2(\eta-t)^2 n}\right)$$

$$\leq (c+3)(5n)^{5d^2}e^{-2(\eta-t)^2 n},$$

because we can choose t such that

$$\eta - t = \frac{t-s}{2c\sqrt{2d^2+1}} \geq \frac{t-s}{4cd}. \tag{7.7}$$

Secondly, let ω be such that $\operatorname{tr}\omega\Pi_j^\eta \geq 1 - \delta'$; as we are interested in $\operatorname{tr}\omega P$, we may again assume without loss of generality that ω is permutation symmetric, and invoke the constrained de Finetti reduction [102, Lemma 18], Eq. (7.5). From that we get, much as before,

$$\operatorname{tr}\omega P^\perp \leq (n+1)^{3d^2}\int \mathrm{d}\sigma\,(\operatorname{tr}\sigma^{\otimes n}P^\perp)F(\omega,\sigma^{\otimes n})^2,$$

and we split the integral on the right hand side into two parts, depending on $\sigma \in \mathcal{F}_s(\underline{v})$ or $\sigma \in \mathcal{C}_s(\underline{v})$: In the latter case, $\operatorname{tr}\sigma^{\otimes n}P^\perp \leq (c+2)(5n)^{2d^2}e^{-\zeta n}$, by Eq. (7.3). In the former case, there exists a j such that $\operatorname{tr}\sigma A_j = w_j \notin [v_j \pm s\Sigma(A_j)]$, and so

$$F(\omega,\sigma^{\otimes n})^2 \leq F\left((1-\delta',\delta'),(\operatorname{tr}\sigma^{\otimes n}\Pi_j^{\eta'}, 1-\operatorname{tr}\sigma^{\otimes N}\Pi_j^{\eta'})\right)$$

$$\leq \left(\sqrt{\delta'} + \sqrt{\operatorname{tr}\sigma^{\otimes n}\Pi_j^{\eta'}}\right)^2$$

$$\leq 2\delta' + 2\operatorname{tr}\sigma^{\otimes n}\Pi_j^{\eta'}$$

$$\leq 2\delta' + 4e^{-2(s-\eta')^2 n},$$

the last line again by Hoeffding's inequality; indeed, with the previous notation,

$$\operatorname{tr}\sigma^{\otimes n}\Pi_j^{\eta'} = \Pr\left\{\frac{1}{n}\sum_\ell Z_\ell \in [v_j \pm \eta'\Sigma(A_j)]\right\}$$

$$\leq \Pr\left\{\frac{1}{n}\sum_\ell Z_\ell \notin [w_j \pm (s-\eta')\Sigma(A_j)]\right\}.$$

116

All taken together, we get

$$\operatorname{tr}\omega P^\perp \le (n+1)^{3d^2}\left((c+2)(5n)^{2d^2}e^{-\zeta n} + 4e^{-2(s-\eta')^2 n} + 2\delta'\right)$$
$$\le (n+1)^{3d^2}(c+3)(5n)^{2d^2}e^{-2(s-\eta')^2 n} + 2(n+1)^{3d^2}\delta',$$

because we can choose s such that

$$s - \eta' = \frac{t-s}{2c\sqrt{2d^2+1}} \ge \frac{t-s}{4cd}. \tag{7.8}$$

From eqs. (7.7) and (7.8) we get by summation

$$\eta - \eta' = t - s + \frac{t-s}{c\sqrt{2d^2+1}} \le (t-s)\left(1 + \frac{1}{cd}\right),$$

from which we obtain

$$s - \eta' = \eta - t \ge \frac{\eta - \eta'}{4c(d+1)},$$

concluding the proof.   ∎

Lemma 7.2. *For all $0 < s < t$ there exists $\zeta > 0$, such that for all n there exists a permutation symmetric projector P on $\mathcal{H}^{\otimes n}$ with the properties*

$$\forall \rho \in \mathcal{C}_s(\underline{v}) \quad \operatorname{tr}\rho^{\otimes n} P^\perp \le (c+2)(5n)^{2d^2}e^{-\zeta n}, \tag{7.9}$$

$$\forall \sigma \in \mathcal{F}_t(\underline{v}) \quad \operatorname{tr}\sigma^{\otimes n} P \le (c+2)(5n)^{2d^2}e^{-\zeta n}. \tag{7.10}$$

The constant ζ may be chosen as $\zeta = \frac{(t-s)^2}{2c^2(2d^2+1)}$.

Proof. We start by showing that there is a POVM $(M, \mathbb{1} - M)$ with

$$\forall \rho \in \mathcal{C}_s(\underline{v}) \quad \operatorname{tr}\rho^{\otimes n}(\mathbb{1} - M) \le ce^{-\frac{(t-s)^2}{2c^2}n}, \tag{7.11}$$

$$\forall \sigma \in \mathcal{F}_t(\underline{v}) \qquad \operatorname{tr}\sigma^{\otimes n} M \le e^{-\frac{(t-s)^2}{2c^2}n}. \tag{7.12}$$

Namely, for each $\ell = 0, \ldots, n$ choose $j_\ell \in \{1, \ldots, c\}$ uniformly at random and measure A_{j_ℓ} on the ℓ-th system. Denote the outcome by the random variable $Z_\ell^{j_\ell}$ and let $Z_\ell^j = 0$ for $j \ne j_\ell$. Thus, for all j, the random variables Z_ℓ^j are i.i.d. with mean $\mathbb{E}Z_\ell^j = \frac{1}{c}\operatorname{tr}\rho A_j$, if the measured state is $\rho^{\otimes n}$.

Outcome M corresponds to the event

$$\forall j \quad \frac{1}{n}\sum_\ell Z_\ell^j \in \frac{1}{c}\left[v_j \pm \frac{s+t}{2}\Sigma(A_j)\right];$$

117

outcome $\mathbb{1} - M$ corresponds to the complementary event

$$\exists j \; \frac{1}{n} \sum_\ell Z_\ell^j \notin \frac{1}{c} \left[v_j \pm \frac{s+t}{2} \Sigma(A_j) \right].$$

We can use Hoeffding's inequality to bound the traces in question. For $\rho \in \mathcal{C}_s(\underline{v})$, we have $|\mathbb{E} Z_\ell^j - v_j| \le \frac{s}{c} \Sigma(A_j)$ for all j, and so:

$$
\begin{aligned}
\operatorname{tr} \rho^{\otimes n} (\mathbb{1} - M) &= \Pr \left\{ \exists j \; \frac{1}{n} \sum_\ell Z_\ell^j \notin \frac{1}{c} \left[v_j \pm \frac{s+t}{2} \Sigma(A_j) \right] \right\} \\
&\le \sum_{j=1}^c \Pr \left\{ \frac{1}{n} \sum_\ell Z_\ell^j \notin \frac{1}{c} \left[v_j \pm \frac{s+t}{2} \Sigma(A_j) \right] \right\} \\
&\le \sum_{j=1}^c \Pr \left\{ \frac{1}{n} \sum_\ell Z_\ell^j \notin \frac{1}{c} \left[v_j \pm \frac{s+t}{2} \Sigma(A_j) \right] \right\} \\
&\le \sum_{j=1}^c \Pr \left\{ \left| \frac{1}{n} \sum_\ell Z_\ell^j - \mathbb{E} Z_1^j \right| > \frac{t-s}{2c} \Sigma(A_j) \right\} \\
&\le c \, e^{-\frac{(t-s)^2}{2c^2} n}.
\end{aligned}
$$

For $\sigma \in \mathcal{F}_t(\underline{v})$, there exists a j such that $|\mathbb{E} Z_\ell^j - v_j| > \frac{t}{c} \Sigma(A_j)$. Thus,

$$
\begin{aligned}
\operatorname{tr} \sigma^{\otimes n} M &\le \Pr \left\{ \frac{1}{n} \sum_\ell Z_\ell^j \in \frac{1}{c} \left[v_j \pm \frac{s+t}{2} \Sigma(A_j) \right] \right\} \\
&\le \Pr \left\{ \left| \frac{1}{n} \sum_\ell Z_\ell^j - \mathbb{E} Z_1^j \right| > \frac{t-s}{2c} \Sigma(A_j) \right\} \\
&\le e^{-\frac{(t-s)^2}{2c^2} n}.
\end{aligned}
$$

This POVM is, by construction, permutation symmetric, but M is not a projector. To fix this, choose λ-nets $\mathcal{N}_C^\lambda$ in $\mathcal{C}_s(\underline{v})$ and $\mathcal{N}_F^\lambda$ in $\mathcal{F}_t(\underline{v})$, with $\lambda = e^{-\zeta n}$, with $\zeta = \frac{(t-s)^2}{2c^2(2d^2+1)}$. This means that every state $\rho \in \mathcal{C}_s(\underline{v})$ is no farther than λ in trace distance from a $\rho' \in \mathcal{N}_C^\lambda$, and likewise for $\mathcal{F}_t(\underline{v})$. By [104, Lemma III.6] (or rather, a minor variation of its proof), we can find such nets with $|\mathcal{N}_C^\lambda|, |\mathcal{N}_F^\lambda| \le \left(\frac{5n}{\lambda} \right)^{2d^2}$ elements. Form the two states

$$\Gamma := \frac{1}{|\mathcal{N}_C^\lambda|} \sum_{\rho \in \mathcal{N}_C^\lambda} \rho^{\otimes n},$$

$$\Phi := \frac{1}{|\mathcal{N}_F^\lambda|} \sum_{\sigma \in \mathcal{N}_F^\lambda} \sigma^{\otimes n},$$

and let
$$P := \{\Gamma - \Phi \geq 0\}$$

be the Helstrom projector which optimally distinguishes Γ from Φ. But we know already a POVM that distinguishes the two states, hence $(P, P^\perp = \mathbb{1} - P)$ cannot be worse:

$$\operatorname{tr} \Gamma P^\perp + \operatorname{tr} \Phi P \leq \operatorname{tr} \Gamma(\mathbb{1} - M) + \operatorname{tr} \Phi M \leq (c+1)e^{-\frac{(t-s)^2}{2c^2}n},$$

thus for all $\rho \in \mathcal{N}_C^\lambda$ and $\sigma \in \mathcal{N}_F^\lambda$,

$$\operatorname{tr} \rho^{\otimes n} P^\perp, \; \operatorname{tr} \sigma^{\otimes n} P \leq (c+1)\left(\frac{5n}{\lambda}\right)^{2d^2} e^{-\frac{(t-s)^2}{2c^2}n}.$$

So, by the λ-net property, we find for all $\rho \in \mathcal{C}_s(\underline{v})$ and $\sigma \in \mathcal{F}_t(\underline{v})$,

$$\operatorname{tr} \rho^{\otimes n} P^\perp, \; \operatorname{tr} \sigma^{\otimes n} P \leq \lambda + (c+1)\left(\frac{5n}{\lambda}\right)^{2d^2} e^{-\frac{(t-s)^2}{2c^2}n} \leq (c+2)(5n)^{2d^2}e^{-\zeta n},$$

by our choice of λ. ∎

Corollary 7.1. *For charges A_j, values $v_j = \langle A_j \rangle$ and $n > 0$, Theorem 7.2 implies that there is an a.m.c. subspace $\mathcal{M}$ of $\mathcal{H}^{\otimes n}$ for any $\eta' > 0$ with the following parameters:*

$$\eta = 2\eta',$$
$$\delta' = \frac{c+3}{2}(5n)^{2d^2} e^{-\frac{n\eta'^2}{8c^2(d+1)^2}},$$
$$\delta = (c+3)(5n)^{2d^2} e^{-\frac{n\eta'^2}{8c^2(d+1)^2}},$$
$$\epsilon = 2(c+3)(n+1)^{3d^2}(5n)^{2d^2} e^{-\frac{n\eta'^2}{8c^2(d+1)^2}}.$$

Moreover, let $\rho^n = \rho_1 \otimes \cdots \otimes \rho_n$ be a state with $\frac{1}{n}\left|\operatorname{Tr}\left(\rho^n A_j^{(n)}\right) - v_j\right| \leq \frac{1}{2}\eta'\Sigma(A_j)$. Then, ρ^n projects onto a.m.c. subspace with probability ϵ:

$$\operatorname{Tr}\left(\rho^n P\right) \geq 1 - \epsilon.$$

Proof. For simplicity of notation we drop the subscript j from A_j, v_j and $\Pi_j^{\eta'}$, so let $\sum_{l=1}^{d} E_l |l\rangle\langle l|$ be the spectral decomposition of A. Define independent random variables X_i for $i = 1, \ldots, n$ taking values in the set $\{E_1, \ldots, E_d\}$ with probabilities $p_i(E_l) = \operatorname{Tr}\left(\rho_i |l\rangle\langle l|\right)$. Furthermore, define random variable $\overline{X} = \frac{X_1 + \ldots + X_n}{n}$ which has the following expectation value

$$\mathbb{E}(\overline{X}) = \frac{1}{n}\operatorname{Tr}\left(\rho^n A^{(n)}\right).$$

Therefore, we obtain

$$
\begin{aligned}
1 - \mathrm{Tr}\left(\rho^n \Pi^{\eta'}\right) \\
&= \sum_{\substack{l_1,\ldots,l_n: \\ |E_{l_1}+\ldots+E_{l_n}-nv| \geq n\eta'\Sigma(A)}} \langle l_1|\,\rho_1\,|l_1\rangle \ldots \langle l_n|\,\rho_n\,|l_n\rangle \\
&= \mathrm{Pr}\left(\left|\overline{X}-v\right| \geq \eta'\Sigma(A)\right) \\
&= \mathrm{Pr}\left(\overline{X}-\mathbb{E}(\overline{X}) \geq \eta'\Sigma(A)+v-\mathbb{E}(\overline{X}) \bigcup \overline{X}-\mathbb{E}(\overline{X}) \leq -\eta'\Sigma(A)+v-\mathbb{E}(\overline{X})\right) \\
&\leq \exp\left(-\frac{2n(\eta'\Sigma(A)+v-\mathbb{E}(\overline{X}))^2}{(\Sigma(A))^2}\right) + \exp\left(-\frac{2n(\eta'\Sigma(A)-v+\mathbb{E}(\overline{X}))^2}{(\Sigma(A))^2}\right) \\
&\leq 2\exp\left(-\frac{n\eta'^2}{2}\right) \\
&\leq \delta',
\end{aligned}
$$

where the second line follows because random the variables $X_1,\ldots,X_n$ are independent and as a result $\mathrm{Pr}\{\overline{X}=\frac{E_{l_1}+\ldots+E_{l_n}}{n}\} = \langle l_1|\,\rho_1\,|l_1\rangle \cdots \langle l_n|\,\rho_n\,|l_n\rangle$. The fourth line is due to Hoeffding's inequality (Lemma A.5). The fifth line is due to assumption $\left|\mathbb{E}(\overline{X})-v\right| \leq \frac{1}{2}\eta'\Sigma(A)$.

Thus, by the definition of a.m.c. subspace $\mathrm{Tr}\left(\rho^n P\right) \geq 1-\epsilon$. $\blacksquare$

7.3   Proof of the AET Theorem 7.1

Here, we first prove the following lemma where we will use points 3 and 4 to prove the main theorem. Corollary 7.1 implies that assuming $\frac{1}{n}\mathrm{Tr}\,\rho^n A_j^{(n)} \approx \frac{1}{n}\mathrm{Tr}\,\sigma^n A_j^{(n)} \approx v_j$ the states ρ^n and σ^n project onto the a.m.c. subspace with high probability. Hence, in Lemma 7.3, we show that one can find states $\widetilde{\rho}$ and $\widetilde{\sigma}$ with support inside the a.m.c. subspace which are very close to the original states in trace norm, that is, $\widetilde{\rho} \approx \rho^n$ and $\widetilde{\sigma} \approx \sigma^n$, and there are unitaries V_1 and V_2 that factorizes these states to the tensor product of maximally mixed states τ and τ' and some other state of very small dimension:

$$
V_1 \widetilde{\rho} V_1^\dagger = \tau \otimes \omega \quad \text{and} \quad V_2 \widetilde{\sigma} V_2^\dagger = \tau' \otimes \omega'.
$$

Further, assuming that the states ρ^n and σ^n have very close entropy rates, i.e. $\frac{1}{n}S(\rho^n) \approx \frac{1}{n}S(\sigma^n)$, one can find states τ and τ' with the same dimension that is $\tau=\tau'$. Thus, we observe that two states $\widetilde{\rho}\otimes\omega'$ and $\widetilde{\sigma}\otimes\omega$ have exactly the same spectrum, so there is unitary acting on the a.m.c. subspace and the ancillary system taking one state to another. Based on the properties of the a.m.c. subspace, we show that this unitary is an almost-commuting unitary with the charges $A_j^{(n)}$.

Lemma 7.3. *Let subspace $\mathcal{M}$ of $\mathcal{H}^{\otimes n}$ with projector P be a high probability subspace for state $\rho^n = \rho_1 \otimes \cdots \otimes \rho_n$, i.e. $\mathrm{Tr}\,(\rho^n P) \geq 1 - \epsilon$. Then, for sufficiently large n there is a subspace $\widetilde{\mathcal{M}} \subseteq \mathcal{M}$ with projector $\widetilde{P}$ and state $\widetilde{\rho}$ with support inside $\widetilde{\mathcal{M}}$ such that the following holds:*

1. *$\mathrm{Tr}\,\left(\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n \widetilde{P}\right) \geq 1 - 2\sqrt{\epsilon} - \frac{1}{O(\alpha)}$.*

2. *$2^{-\sum_{i=1}^n S(\rho_i) - 2\alpha\sqrt{n}} \widetilde{P} \leq \widetilde{P}\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n \widetilde{P} \leq 2^{-\sum_{i=1}^n S(\rho_i) + \alpha\sqrt{n}} \widetilde{P}$.*

3. *There is a unitary U such that $U\widetilde{\rho}U^\dagger = \tau \otimes \omega$ where τ is a maximally mixed state of dimension $2^{\sum_{i=1}^n S(\rho_i) - O(\alpha\sqrt{n})}$, and ω is a state of dimension $2^{O(\alpha\sqrt{n})}$.*

4. *$\left\|\widetilde{\rho} - \rho^n\right\|_1 \leq 2\sqrt{\epsilon} + \frac{1}{O(\alpha)} + 2\sqrt{2\sqrt{\epsilon} + \frac{1}{O(\alpha)}}$.*

Proof. 1. Let $E \geq 0$ and $F \geq 0$ be two positive operators such that $E + F = P\Pi_{\alpha,\rho^n}^n P$ where all eigenvalues of F are smaller than $2^{-\alpha\sqrt{n}}$, and define $\widetilde{P}$ to be the projection onto the support of E. In other words, $\widetilde{P}$ is the projection onto the support of $P\Pi_{\alpha,\rho^n}^n P$ with corresponding eigenvalues greater $2^{-\alpha\sqrt{n}}$. Then, we obtain

$$
\begin{aligned}
&\mathrm{Tr}\,\left(\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n \widetilde{P}\right) \\
&\quad \geq \mathrm{Tr}\,\left(\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n E\right) \\
&\quad = \mathrm{Tr}\,\left(\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n P\Pi_{\alpha,\rho^n}^n P\right) - \mathrm{Tr}\,\left(\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n F\right) \\
&\quad \geq \mathrm{Tr}\,\left(\rho^n P\Pi_{\alpha,\rho^n}^n P\right) - \left\|\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n - \rho^n\right\|_1 - \mathrm{Tr}\,\left(\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n F\right) \\
&\quad \geq \mathrm{Tr}\,\left(\rho^n \Pi_{\alpha,\rho^n}^n\right) - \left\|P\rho^n P - \rho^n\right\|_1 - \left\|\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n - \rho^n\right\|_1 - \mathrm{Tr}\,\left(\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n F\right) \\
&\quad \geq \mathrm{Tr}\,\left(\rho^n \Pi_{\alpha,\rho^n}^n\right) - \left\|P\rho^n P - \rho^n\right\|_1 - \left\|\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n - \rho^n\right\|_1 - 2^{-\alpha\sqrt{n}} \\
&\quad \geq 1 - \frac{\beta}{\alpha^2} - 2\sqrt{\epsilon} - 2\frac{\sqrt{\beta}}{\alpha} - 2^{-\alpha\sqrt{n}},
\end{aligned}
$$

where the first line follows from the fact that $\widetilde{P} \geq E$. The third, forth and fifth lines are due to Hölder inequality. The last line follows from Lemma A.14 and gentle operator lemma A.13.

2. By the fact that in the typical subspace the eigenvalues of ρ^n are bounded (Lemma A.14), we obtain

$$
\begin{aligned}
\widetilde{P}\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n \widetilde{P} &\leq 2^{-\sum_{i=1}^n S(\rho_i) + \alpha\sqrt{n}} \widetilde{P}\Pi_{\alpha,\rho^n}^n \widetilde{P} \\
&\leq 2^{-\sum_{i=1}^n S(\rho_i) + \alpha\sqrt{n}} \widetilde{P}.
\end{aligned}
$$

121

For the lower bound notice that

$$\widetilde{P}\Pi^n_{\alpha,\rho^n}\rho^n\Pi^n_{\alpha,\rho^n}\widetilde{P} \geq 2^{-\sum_{i=1}^n S(\rho_i)-\alpha\sqrt{n}}\widetilde{P}\Pi^n_{\alpha,\rho^n}\widetilde{P}$$
$$= 2^{-\sum_{i=1}^n S(\rho_i)-\alpha\sqrt{n}}\widetilde{P}P\Pi^n_{\alpha,\rho^n}P\widetilde{P}$$
$$\geq 2^{-\sum_{i=1}^n S(\rho_i)-2\alpha\sqrt{n}}\widetilde{P},$$

where the equality holds because $\widetilde{P} \subseteq \mathcal{M}$, therefore $\widetilde{P}P = \widetilde{P}$. The last inequality follows because $\widetilde{P}$ is the projection onto support of $P\Pi^n_{\alpha,\rho^n}P$ with eigenvalues greater $2^{-\alpha\sqrt{n}}$.

3. Consider the unnormalized state $\widetilde{P}\Pi^n_{\alpha,\rho^n}\rho^n\Pi^n_{\alpha,\rho^n}\widetilde{P}$ with support inside $\widetilde{\mathcal{M}}$. From from point 2, we know that all the eigenvalues of this state belongs to the interval $[2^{-\sum_{i=1}^n S(\rho_i)-2\alpha\sqrt{n}}, 2^{-\sum_{i=1}^n S(\rho_i)+\alpha\sqrt{n}}]$ which we denote it by $[p_{\min}, p_{\max}]$. We divide this interval to $b = 2^{\lfloor 5\alpha\sqrt{n}\rfloor}$ many intervals (bins) with equal length of $\Delta p = \frac{p_{\max}-p_{\min}}{b}$. Now, we *trim* the eigenvalues of this unnormalized state in three steps as follows.

(a) Each eigenvalue belongs to a bin which is an interval $[p_k, p_{k+1})$ for some $0 \leq k \leq b-1$ with $p_k = p_{\min} + \Delta p \times k$. For example, eigenvalue λ_l is equal to $p_k + q_l$ for some k such that $0 \leq q_l < \Delta p$. We throw away q_l part of each eigenvalue λ_l. The sum of these parts over all eigenvalues is very small

$$\sum_{l=1}^{|\widetilde{\mathcal{M}}|} q_l \leq \Delta p|\widetilde{\mathcal{M}}| \leq 2^{-2\alpha\sqrt{n}+1},$$

where the dimension of the subspace $\widetilde{\mathcal{M}}$ is bounded as $|\widetilde{\mathcal{M}}| \leq 2^{\sum_{i=1}^n S(\rho_i)+2\alpha\sqrt{n}}$ which follows from point 2 of the lemma.

(b) We throw away the bins which contain less than $2^{\sum_{i=1}^n S(\rho_i)-10\alpha\sqrt{n}}$ many eigenvalues. The sum of all the eigenvalues that are thrown away is bounded by

$$2^{\sum_{i=1}^n S(\rho_i)-10\alpha\sqrt{n}} \times 2^{5\alpha\sqrt{n}} \times 2^{-\sum_{i=1}^n S(\rho_i)+\alpha\sqrt{n}} \leq 2^{-4\alpha\sqrt{n}},$$

in the left member, the first number is the number of eigenvalues in the bin; the second is the number of bins, and the third is the maximum eigenvalue.

(c) If a bin, e.g. kth bin, is not thrown away in the previous step, it contains M_k many eigenvalues with the same value with

$$2^{\sum_{i=1}^n S(\rho_i)-10\alpha\sqrt{n}} \leq M_k \leq 2^{\sum_{i=1}^n S(\rho_i)+2\alpha\sqrt{n}}. \tag{7.13}$$

122

Let

$$L = 2^{\lfloor \sum_{i=1}^n S(\rho_i) - 10\alpha\sqrt{n} \rfloor} \tag{7.14}$$

and for the kth bin, let m_k be an integer number such that

$$m_k L \leq M_k \leq (m_k + 1)L. \tag{7.15}$$

Then, m_k is bounded as follows

$$m_k \leq 2^{12\alpha\sqrt{n}}. \tag{7.16}$$

From the kth bin, we keep $m_k L$ number of eigenvalues and throw away the rest where there are $M_k - m_k L \leq L$ many of them; the sum of the eigenvalues that are thrown away in this step is bounded by

$$\sum_{k=0}^{b-1} p_k(M_k - m_k L) \leq L \sum_{k=0}^{b-1} p_k \leq 2^{-4\alpha\sqrt{n}}.$$

Therefore, for sufficiently large n the sum of the eigenvalues thrown away in the last three steps is bounded by

$$2^{-2\alpha\sqrt{n}+1} + 2^{-4\alpha\sqrt{n}} + 2^{-4\alpha\sqrt{n}} \leq 2^{-\alpha\sqrt{n}} \tag{7.17}$$

The kept eigenvalues of all bins form an L-fold degenerate unnormalized state of dimension $\sum_{k=0}^{b-1} m_k L$ because each eigenvalue has at least degeneracy of the order of L. Thus, up to unitary $U^\dagger$, it can be factorized into the tensor product of a maximally mixed state τ and unnormalized state ω' of dimensions L and $\sum_{k=0}^{b-1} m_k$, respectively. From (7.16), the dimension of ω' is bounded by

$$\sum_{k=0}^{b-1} m_k \leq 2^{12\alpha\sqrt{n}} \times 2^{5\alpha\sqrt{n}} = 2^{17\alpha\sqrt{n}}.$$

Then, let $\omega = \frac{\omega'}{\mathrm{Tr}(\omega')}$ and define

$$\widetilde{\rho} = U\tau \otimes \omega U^\dagger.$$

4. From points 3 and 1, we obtain

$$\mathrm{Tr}(\omega') = \mathrm{Tr}(\tau \otimes \omega') \tag{7.18}$$

$$\geq \mathrm{Tr}(\widetilde{P}\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n \widetilde{P}) - 2^{-\alpha\sqrt{n}} \tag{7.19}$$

$$\geq 1 - 2\sqrt{\epsilon} - 2\frac{\sqrt{\beta}}{\alpha} - \frac{\beta}{\alpha^2} - 2^{-\alpha\sqrt{n}+1}. \tag{7.20}$$

123

Thereby, we get the following

$$
\begin{aligned}
\left\|\widetilde{\rho} - \rho^n\right\|_1 &\leq \left\|\widetilde{\rho} - U\tau \otimes \omega' U^\dagger\right\|_1 + \left\|U\tau \otimes \omega' U^\dagger - \widetilde{P}\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n \widetilde{P}\right\|_1 \\
&\quad + \left\|\widetilde{P}\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n \widetilde{P} - \rho^n\right\|_1 \\
&\leq 1 - \operatorname{Tr}\left(\omega'\right) + \left\|U\tau \otimes \omega' U^\dagger - \widetilde{P}\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n \widetilde{P}\right\|_1 \\
&\quad + \left\|\widetilde{P}\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n \widetilde{P} - \rho^n\right\|_1 \\
&\leq 1 - \operatorname{Tr}\left(\omega'\right) + 2^{-\alpha\sqrt{n}} + \left\|\widetilde{P}\Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n \widetilde{P} - \rho^n\right\|_1 \\
&\leq 1 - \operatorname{Tr}\left(\omega'\right) + 2^{-\alpha\sqrt{n}} + 2\sqrt{2\sqrt{\epsilon} + 2\frac{\sqrt{\beta}}{\alpha} + \frac{\beta}{\alpha^2} + 2^{-\alpha\sqrt{n}}} \\
&= 2\sqrt{\epsilon} + 2\frac{\sqrt{\beta}}{\alpha} + \frac{\beta}{\alpha^2} + 2^{-\alpha\sqrt{n}+1} + 2\sqrt{2\sqrt{\epsilon} + 2\frac{\sqrt{\beta}}{\alpha} + \frac{\beta}{\alpha^2} + 2^{-\alpha\sqrt{n}}},
\end{aligned}
$$

where the first line is due to triangle inequality. The second, third and fourth lines are due to Eqs. (7.18) and (7.17), and Lemma A.13, respectively. ∎

Proof of Theorem 7.1. We first prove the *if* part. If there is an almost-commuting unitary U and an ancillary system with the desired properties stated in the theorem, then we obtain

$$
\begin{aligned}
\frac{1}{n}\left|S(\rho^n) - S(\sigma^n)\right| &\leq \frac{1}{n}\left|S(\rho^n \otimes \omega') - S(\sigma^n \otimes \omega)\right| + \frac{1}{n}\left|S(\omega') - S(\omega)\right| \\
&\leq \frac{1}{n}\left|S(\rho^n \otimes \omega') - S(\sigma^n \otimes \omega)\right| + \frac{2}{n}\log 2^{o(n)} \\
&= \frac{1}{n}\left|S(U(\rho^n \otimes \omega')U^\dagger) - S(\sigma^n \otimes \omega)\right| + o(1) \\
&\leq \frac{1}{n}o(1)\log\left(d^n \times 2^{o(n)}\right) + \frac{1}{n}h\left(o(1)\right) + o(1) \\
&= o(1),
\end{aligned}
$$

where the first line follows from additivity of the von Neumann entropy and triangle inequality. The second line is due to the fact that von Neumann entropy of a state is upper bounded by the logarithm of the dimension. The penultimate line follows from continuity of von Neumann entropy [72, 73] where $h(x) = -x\log x - (1-x)\log(1-x)$ is the binary entropy function.

Moreover, we obtain

$$\frac{1}{n}\left|\operatorname{Tr}\left(\rho^n A_j^{(n)}\right) - \operatorname{Tr}\left(\sigma^n A_j^{(n)}\right)\right|$$

$$= \frac{1}{n}\left|\operatorname{Tr}\left(\rho^n \otimes \omega'(A_j^{(n)} + A_j')\right) - \operatorname{Tr}\left(\sigma^n \otimes \omega(A_j^{(n)} + A_j')\right)\right|$$

$$\leq \frac{1}{n}\left|\operatorname{Tr}\left(\rho^n \otimes \omega'(A_j^{(n)} + A_j')\right) - \operatorname{Tr}\left(U\rho^n \otimes \omega'U^\dagger(A_j^{(n)} + A_j')\right)\right|$$

$$\quad + \frac{1}{n}\left|\operatorname{Tr}\left(U\rho^n \otimes \omega'U^\dagger(A_j^{(n)} + A_j')\right) - \operatorname{Tr}\left(\sigma^n \otimes \omega(A_j^{(n)} + A_j')\right)\right|$$

$$= \frac{1}{n}\left|\operatorname{Tr}\left(\rho^n \otimes \omega'\left(A_j^{(n)} + A_j' - U^\dagger(A_j^{(n)} + A_j')U\right)\right)\right| \tag{7.21}$$

$$\quad + \frac{1}{n}\left|\operatorname{Tr}\left(\left(U\rho^n \otimes \omega'U^\dagger - \sigma^n \otimes \omega\right)\left(A_j^{(n)} + A_j'\right)\right)\right|$$

$$\leq \frac{1}{n}\operatorname{Tr}\left(\rho^n \otimes \omega'\right)\left\|U(A_j^{(n)} + A_j')U^\dagger - (A_j^{(n)} + A_j')\right\|_\infty \tag{7.22}$$

$$\quad + \frac{1}{n}\left\|U\rho^n \otimes \omega'U^\dagger - \sigma^n \otimes \omega\right\|_1\left\|A_j^{(n)} + A_j'\right\|_\infty$$

$$= o(1), \tag{7.23}$$

the second line follows because $A_j' = 0$ for all j. The third and fifth lines are due to triangle inequality and Hölder's inequality, respectively.

Now, we prove the *only if* part. Assume for the sates ρ^n and σ^n the following holds:

$$\frac{1}{n}|S(\rho^n) - S(\sigma^n)| \leq \gamma_n$$

$$\frac{1}{n}\left|\operatorname{Tr}\left(A_j^{(n)}\rho^n\right) - \operatorname{Tr}\left(A_j^{(n)}\sigma^n\right)\right| \leq \gamma_n' \qquad j = 1,\ldots,c,$$

for vanishing γ_n and γ_n'. According to Theorem 7.2, for charges A_j, values $v_j = \frac{1}{n}\operatorname{Tr}\left(\rho^n A_j^{(n)}\right)$, $\eta' > 0$ and any $n > 0$, there is an a.m.c. subspace $\mathcal{M}$ of $\mathcal{H}^{\otimes n}$ with projector P and the following parameters:

$$\eta = 2\eta',$$

$$\delta' = \frac{c+3}{2}(5n)^{2d^2}e^{-\frac{n\eta'^2}{8c^2(d+1)^2}},$$

$$\delta = (c+3)(5n)^{2d^2}e^{-\frac{n\eta'^2}{8c^2(d+1)^2}},$$

$$\epsilon = 2(c+3)(n+1)^{3d^2}(5n)^{2d^2}e^{-\frac{n\eta'^2}{8c^2(d+1)^2}}.$$

Choose η' as the following such that δ, δ' and ϵ vanish for large n:

$$
\eta' = \begin{cases} \frac{\sqrt{8}c(d+1)}{n^{\frac{1}{4}}\Sigma(A)_{\min}} & \text{if} \quad \gamma'_n \le \frac{1}{n^{\frac{1}{4}}} \\ \frac{\sqrt{8}c(d+1)\gamma'_n}{\Sigma(A)_{\min}} & \text{if} \quad \gamma'_n > \frac{1}{n^{\frac{1}{4}}} \end{cases}
$$

where $\Sigma(A)_{\min}$ is the minimum spectral diameter among all spectral diameters of charges $\Sigma(A_j)$. Since $\frac{1}{n}\mathrm{Tr}\left(\rho^n A_j^{(n)}\right) = v_j$ and $\left|\frac{1}{n}\mathrm{Tr}\left(\sigma^n A_j^{(n)}\right) - v_j\right| \le \frac{1}{2}\eta'\Sigma(A_j)$, Corollary 7.1 implies that states ρ^n and σ^n project onto this a.m.c. subspace with probability ϵ:

$$
\mathrm{Tr}\left(\rho^n P\right) \ge 1 - \epsilon,
$$
$$
\mathrm{Tr}\left(\sigma^n P\right) \ge 1 - \epsilon.
$$

Moreover, consider the typical projectors Π^n_{α,ρ^n} and Π^n_{α,σ^n} of states ρ^n and σ^n, respectively, with $\alpha = n^{\frac{1}{3}}$. Then point 3 and 4 of Lemma 7.3 implies that there are states $\widetilde{\rho}$ and $\widetilde{\sigma}$ with support inside the a.m.c. subspace $\mathcal{M}$ and unitaries V_1 and V_2 such that

$$
\begin{aligned}
&\|\widetilde{\rho} - \rho^n\|_1 \le o(1), \\
&\|\widetilde{\sigma} - \sigma^n\|_1 \le o(1), \\
&V_1\widetilde{\rho}V_1^\dagger = \tau \otimes \omega, \\
&V_2\widetilde{\sigma}V_2^\dagger = \tau' \otimes \omega', \quad\quad\quad\quad\quad (7.24)
\end{aligned}
$$

where τ and τ' are maximally mixed states; since $|S(\rho^n) - S(\sigma^n)| \le n\gamma_n$, one may choose the dimension of them in Eq. (7.14) to be exactly the same as $L = 2^{\lfloor \sum_{i=1}^n S(\rho_i) - 10z \rfloor}$ with $z = \max\{\alpha\sqrt{n}, n\gamma_n\}$, hence, we obtain $\tau = \tau'$. Then, ω and ω' are states with support inside Hilbert space $\mathcal{K}$ of dimension $2^{o(z)} = 2^{o(n)}$. Then, it is immediate to see that the states $\widetilde{\rho} \otimes \omega'$ and $\widetilde{\sigma} \otimes \omega$ on Hilbert space $\mathcal{M}_t = \mathcal{M} \otimes \mathcal{K}$ have exactly the same spectrum; thus, there is a unitary $\widetilde{U}$ on subspace $\mathcal{M}_t$ such that

$$
\widetilde{U}\widetilde{\rho} \otimes \omega'\widetilde{U}^\dagger = \widetilde{\sigma} \otimes \omega. \quad\quad\quad\quad\quad (7.25)
$$

We extend the unitary $\widetilde{U}$ to $U = \widetilde{U} \oplus \mathbb{1}_{\mathcal{M}_t^\perp}$ acting on $\mathcal{H}^{\otimes n} \otimes \mathcal{K}$ and obtain

$$
\begin{aligned}
&\left\|U\rho^n \otimes \omega'U^\dagger - \sigma^n \otimes \omega\right\|_1 \\
&\quad \le \left\|U\rho^n \otimes \omega'U^\dagger - U\widetilde{\rho} \otimes \omega'U^\dagger\right\|_1 + \left\|\sigma^n \otimes \omega - \widetilde{\sigma} \otimes \omega\right\|_1 + \left\|U\widetilde{\rho} \otimes \omega'U^\dagger - \widetilde{\sigma} \otimes \omega\right\|_1 \\
&\quad = \left\|U\rho^n \otimes \omega'U^\dagger - U\widetilde{\rho} \otimes \omega'U^\dagger\right\|_1 + \left\|\sigma^n \otimes \omega - \widetilde{\sigma} \otimes \omega\right\|_1 \\
&\quad \le o(1),
\end{aligned}
$$

126

where the second and last lines are due to Eqs. (7.25) and (7.24), respectively.

As mentioned before, $\mathcal{M}_t = \mathcal{M} \otimes \mathcal{K}$ is a subspace of $\mathcal{H}^{\otimes n} \otimes \mathcal{K}$ with projector $P_t = P \otimes \mathbb{1}_{\mathcal{K}}$ where P is the corresponding projector of a.m.c. subspace. We define total charges $A_j^t = A_j^{(n)} + A_j'$ and let $A_j' = 0$ for all j and show that every unitary of the form $U = U_{\mathcal{M}_t} \oplus \mathbb{1}_{\mathcal{M}_t^\perp}$ asymptotically commutes with all total charges:

$$
\begin{aligned}
\left\| U A_j^t U^\dagger - A_j^t \right\|_\infty &= \left\| (P_t + P_t^\perp)(U A_j^t U^\dagger - A_j^t)(P_t + P_t^\perp) \right\|_\infty \\
&\le \left\| P_t(U A_j^t U^\dagger - A_j^t)P_t \right\|_\infty + \left\| P_t^\perp(U A_j^t U^\dagger - A_j^t)P_t \right\|_\infty \\
&\quad + \left\| P_t(U A_j^t U^\dagger - A_j^t)P_t^\perp \right\|_\infty + \left\| P_t^\perp(U A_j^t U^\dagger - A_j^t)P_t^\perp \right\|_\infty \\
&= \left\| P_t(U A_j^t U^\dagger - A_j^t)P_t \right\|_\infty + 2\left\| P_t^\perp(U A_j^t U^\dagger - A_j^t)P_t \right\|_\infty \\
&\le 3\left\| (U A_j^t U^\dagger - A_j^t)P_t \right\|_\infty \\
&= 3\left\| (U A_j^t U^\dagger - n v_j \mathbb{1} + n v_j \mathbb{1} - A_j^t)P_t \right\|_\infty \\
&\le 3\left\| (U A_j^t U^\dagger - n v_j \mathbb{1})P_t \right\|_\infty + 3\left\| (A_j^t - n v_j \mathbb{1})P_t \right\|_\infty \\
&= 6\left\| (A_j^t - n v_j \mathbb{1})P_t \right\|_\infty \\
&= 6 \max_{|v\rangle \in \mathcal{M}_t} \left\| (A_j^t - n v_j \mathbb{1}) |v\rangle \right\|_2 \\
&= 6 \max_{|v\rangle \in \mathcal{M}_t} \left\| (A_j^t - n v_j \mathbb{1})(\Pi_j^\eta \otimes \mathbb{1}_{\mathcal{K}} + \mathbb{1} - \Pi_j^\eta \otimes \mathbb{1}_{\mathcal{K}}) |v\rangle \right\|_2 \\
&\le 6 \max_{|v\rangle \in \mathcal{M}_t} \left\| (A_j^t - n v_j \mathbb{1})\Pi_j^\eta \otimes \mathbb{1} |v\rangle \right\|_2 \\
&\quad + 6 \max_{|v\rangle \in \mathcal{M}_t} \left\| (A_j^t - n v_j \mathbb{1})(\mathbb{1} - \Pi_j^\eta \otimes \mathbb{1}) |v\rangle \right\|_2 \\
&\le 6n \Sigma(A_j)\eta + 6 \max_{|v\rangle \in \mathcal{M}_t} \left\| (A_j^t - n v_j I)(\mathbb{1} - \Pi_j^\eta \otimes \mathbb{1}) |v\rangle \right\|_2,
\end{aligned}
$$

where the first line is due to the fact that $P_t + P_t^\perp = \mathbb{1}_{\mathcal{H}^{\otimes n}} \otimes \mathbb{1}_{\mathcal{K}}$. The forth line follows because $U A_j^t U^\dagger - A_j^t$ is a Hermitian operator with zero eigenvalues in the subspace $P_t^\perp$. The fifth line is due to Lemma A.4. The twelfth line is due to the definition of the a.m.c. subspace. Now, bound the second term in the above:

$$
\begin{aligned}
6 \max_{|v\rangle \in \mathcal{M}_t} &\left\| (A_j^t - n v_j I)(\mathbb{1} - \Pi_j^\eta \otimes \mathbb{1}) |v\rangle \right\|_2 \\
&\le 6 \max_{|v\rangle \in \mathcal{M}_t} \left\| A_j^t - n v_j \mathbb{1} \right\|_\infty \left\| (\mathbb{1} - \Pi_j^\eta \otimes \mathbb{1}) |v\rangle \right\|_2 \\
&= 6\left\| A_j^t - n v_j \mathbb{1} \right\|_\infty \max_{|v\rangle \in \mathcal{M}_t} \sqrt{\mathrm{Tr}\left((\mathbb{1} - \Pi_j^\eta \otimes \mathbb{1}) |v\rangle\langle v| \right)} \\
&= 6n\left\| A_j - v_j \mathbb{1} \right\|_\infty \max_{v \in \mathcal{M}} \sqrt{\mathrm{Tr}\left((\mathbb{1} - \Pi_j^\eta) v \right)} \\
&\le 6n\left\| A_j - v_j \mathbb{1} \right\|_\infty \sqrt{\delta},
\end{aligned}
$$

127

the first line is due to Lemma A.4. The last line is by definition of the a.m.c. subspace. Thus, for vanishing δ and η we obtain:

$$\frac{1}{n}\left\| U A_j^t U^\dagger - A_j^t \right\|_\infty \le o(1),$$

concluding the proof. ∎

7.4 Discussion

We have considered an asymptotic resource theory with states of tensor product structure as the objects and allowed operations which are thermodynamically meaningful, namely operations which preserve the entropy and and charges of a system asymptotically. The allowed operations classify the objects into asymptotically equivalent objects that are interconvertible under allowed operations. The basic result on which our theory is built is that the objects are interconvertible via allowed operations if and only if they have the same average entropy and average charge values in the asymptotic limit.

The existence of the allowed operations between the objects of the same class is based on two pillars: First, for objects with the same average entropy there are states with sublinear dimension which can be coupled to the objects to make their spectrum asymptotically identical. Second, objects with the same average charge values project onto a common subspace of the charges of the system which has the property that any unitary acting on this subspace is an almost-commuting unitary with the corresponding charges. Therefore, the spectrum of the objects of the same class can be modified using small ancillary systems and then they are interconvertible via unitaries that asymptotically preserve the charges of the system. The notion of a common subspace for different charges, which are Hermitian operators, is introduced in [100] as approximate microcanonical (a.m.c.) subspace. In this chapter, for given charges and parameters, we show the existence of an a.m.c. which is by construction a permutation-symmetry subspace, which is not guaranteed by the construction in [100].

128

Chapter 8

Asymptotic thermodynamics of multiple conserved quantities

As a thermodynamic theory, or even as a resource theory in general, transformations by almost-commuting unitaries, which we developed in the previous chapter, do not appear to be the most fruitful: they are reversible and induce an equivalence relation among the sequences of product states. In particular, every point $(\underline{a}, s)$ of the phase diagram $\overline{\mathcal{P}}^{(1)}$ defines an equivalence class, namely of all state sequences with charges and entropy converging to $\underline{a}$ and s, respectively.

To make the theory more interesting, and more resembling of ordinary thermodyanmics, including irreversibility as expressed in its first and second laws, we now specialise to a setting considered in many previous papers in the resource theory of thermodynamics, both with with single or multiple conserved quantities. Specifically, we consider an asymptotic analogue of the setting proposed in [105] concerning the interaction of thermal baths with a quantum system and batteries, where it was shown that the second law constrains the combination of extractable charge quantities. In [105], explicit protocols for state transformations to saturate the second law are presented, that store each of several commuting charges in its corresponding battery. However, for the case of non-commuting charges, one battery, or a so-called reference frame, stores all different types of charges [100, 106]. Only recently it was shown that reference frames for non-commuting charges can be constructed, at least under certain conditions, which store the different charge types in physically separated subsystems [107]. Moreover, the size of the bath required to perform the transformations is not addressed in these works, as only the limit of asymptotically large bath was considered. We will address these questions in a similar setting but in the asymptotic regime, where Theorem 7.1 provides the necessary and sufficient condition for physically

possible state transformations. In this new setting, the *asymptotic* second law constrains the combination of extractable charges; we provide explicit protocols for realising transformations satisfying the second law, where each battery can store its corresponding type of work in the general case of non-commuting charges. Furthermore, we determine the minimum number of thermal baths of a given type that is required to perform a transformation.

8.1   System model, batteries and the first law

We consider a system being in contact with a bath and suitable batteries, with a total Hilbert space $Q = S \otimes B \otimes W_1 \otimes \cdots \otimes W_c$, consisting of many non-interacting subsystems; namely, the work system, the thermal bath and c battery systems with Hilbert spaces S, B and W_j for $j = 1, \ldots, c$, respectively. We call the j-th battery system the j-type battery as it is designed to absorb j-type work. The work system and the thermal bath have respectively the charges A_{S_j} and A_{B_j} for all j, but j-type battery has only one nontrivial charge A_{W_j}, and all its other charges are zero because it is meant to store only the j-th charge. The total charge is the sum of the charges of the subsystems $A_j = A_{S_j} + A_{B_j} + A_{W_j}$ for all j. Furthermore, for a charge A, let $\Sigma(A) = \lambda_{\max}(A) - \lambda_{\min}(A)$ denote the spectral diameter, where $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ are the largest and smallest eigenvalues of the charge A, respectively. We assume that the total spectral diameter of the work system and the thermal bath is bounded by the spectral diameter of the battery, that is $\Sigma(A_{S_j}) + \Sigma(A_{B_j}) \leq \Sigma(A_{W_j})$ for all j; this assumption ensures that the batteries can absorb or release charges for transformations.

As we discussed in the previous section, the generalized thermal state $\tau(\underline{a})$ is the state that maximizes the entropy subject to the constraint that the charges A_j have the values a_j. This state is equal to $\frac{1}{Z} e^{-\sum_{j=1}^{c} \beta_j A_j}$ for real numbers β_j called inverse temperatures and chemical potentials; each of them is a smooth function of charge values $a_1, \ldots, a_c$, and $Z = \operatorname{Tr} e^{-\sum_{j=1}^{c} \beta_j A_j}$ is the generalized partition function. Therefore, the generalized thermal state can be equivalently denoted $\tau(\underline{\beta})$ as a function of the inverse temperatures, associated uniquely with the charge values $\underline{a}$. We assume that the thermal bath is initially in a generalized thermal state $\tau_b(\underline{\beta})$, for globally fixed $\underline{\beta}$. This is because in [100] it was argued that these are precisely the *completely passive* states, from which no energy can be extracted into a suitable battery storing energy but no other conserved quantity, by means of almost-commuting unitaries, even when unlimited copies of the state are available. We assume that the work system with state ρ_s and the thermal bath are initially uncorrelated, and furthermore that the battery systems can acquire only pure

states.

Therefore, the initial state of an *individual* global system Q is assumed to be of the following form,

$$\rho_{SBW_1\dots W_c} = \rho_S \otimes \tau(\underline{\beta})_B \otimes |w_1\rangle\langle w_1|_{W_1} \otimes \cdots \otimes |w_c\rangle\langle w_c|_{W_c}, \qquad (8.1)$$

and the final states we consider are of the form

$$\sigma_{SBW_1\dots W_c} = \sigma_{SB} \otimes |w_1'\rangle\langle w_1'|_{W_1} \otimes \cdots \otimes |w_c'\rangle\langle w_c'|_{W_c}, \qquad (8.2)$$

where ρ_S and σ_{SB} are states of the system and system-plus-bath, respectively, and w_j and w_j' label pure states of the j-type battery before and after the transformation. The notation is meant to convey the expectation value of the j-type work, i.e. $w_j^{(\prime)}$ is a real number and $\mathrm{Tr}\, |w_j^{(\prime)}\rangle\langle w_j^{(\prime)}|A_{W_j} = w_j^{(\prime)}$.

The established resource theory of thermodynamics treats the batteries and the bath as 'enablers' of transformations of the system S, and we will show first and second laws that express the essential constraints that any such transformation has to obey. We start with the batteries. With the notations $\underline{W} = W_1\dots W_c$, $|\underline{w}\rangle = |w_1\rangle\cdots|w_c\rangle$, and $|\underline{w}'\rangle = |w_1'\rangle\cdots|w_c'\rangle$, let us look at a sequence $\rho^n = \rho_{S^n} = \rho_{S_1} \otimes \cdots \otimes \rho_{S_n}$ of initial system states, and a sequence $|\underline{w}\rangle\langle\underline{w}|^n = |\underline{w}_1\rangle\langle\underline{w}_1|_{\underline{W}_1} \otimes \cdots \otimes |\underline{w}_n\rangle\langle\underline{w}_n|_{\underline{W}_n}$ of initial battery states, recalling that the baths are initially all in the same thermal state, $\tau_{B^n} = \tau(\underline{\beta})^{\otimes n}$; furthermore a sequence of target states $\sigma^n = \sigma_{S^n B^n} = \sigma_{S_1 B_1} \otimes \cdots \otimes \sigma_{S_n B_n}$ of the system and bath, and a sequence $|\underline{w}'\rangle\langle\underline{w}'|^n = |\underline{w}_1'\rangle\langle\underline{w}_1'|_{\underline{W}_1} \otimes \cdots \otimes |\underline{w}_n'\rangle\langle\underline{w}_n'|_{\underline{W}_n}$ of target states of the batteries.

Definition 8.1. *A sequence of states ρ^n on any system Q^n is called* regular *if its charge and entropy rates converge, i.e. if*

$$a_j = \lim_{n\to\infty} \frac{1}{n}\mathrm{Tr}\, \rho^n A_j^{(n)}, \ \ j = 1,\dots,c, \ \ and$$

$$s = \lim_{n\to\infty} \frac{1}{n} S(\rho^n)$$

exist. To indicate the dependence on the state sequence, we write $a_j(\{\rho^n\})$ and $s(\{\rho^n\})$.

According to the AET and the other results of the previous section, every point $(\underline{a}, s)$ in the phase diagram $\overline{\mathcal{P}}^{(1)}$ labels an equivalence class of regular sequences of product states under transformations by almost-commuting unitaries.

In the rest of the chapter we will essentially focus on regular sequences, so that we can simply identify them, up to asymptotic equivalence, with a

point in the phase diagram. However, it should be noted that at the expense of clumsier expressions, most of our expositions can be extended to arbitrary sequences of product states or block-product states.

Now, for regular sequences ρ_{S^n} of initial states of the system and final states of the system plus bath, $\sigma_{S^n B^n}$, as well as regular sequences of initial and final battery states, $|\underline{w}\rangle\langle\underline{w}|^n$ and $|\underline{w}'\rangle\langle\underline{w}'|^n$, respectively, define the asymptotic rate of j-th charge change of the j-type battery as

$$\Delta A_{W_j} := a_j(\{|w_j'\rangle\langle w_j'|^n\}) - a_j(\{|w_j\rangle\langle w_j|^n\}) = \lim_{n\to\infty} \frac{1}{n} \mathrm{Tr}\,(|w_j'\rangle\langle w_j'|^n - |w_j\rangle\langle w_j|^n) A_{W_j}^{(n)}. \tag{8.3}$$

Where there is no danger of confusion, we denote this number also as W_j, the j-type work extracted (if $W_j < 0$, this means that the work $-W_j$ is done on system S and bath B).

Similarly, we define the asymptotic rate of j-th charge change of the work system and the bath as

$$\Delta A_{S_j} := a_j(\{\sigma_{S^n}\}) - a_j(\{\rho_{S^n}\}) = \lim_{n\to\infty} \frac{1}{n} \mathrm{Tr}\,(\sigma_{S^n} - \rho_{S^n}) A_{S_j}^{(n)},$$

$$\Delta A_{B_j} := a_j(\{\sigma_{B^n}\}) - a_j(\{\tau(\underline{\beta})_{B^n}\}) = \lim_{n\to\infty} \frac{1}{n} \mathrm{Tr}\,(\sigma_{B^n} - \tau(\underline{\beta})_B^{\otimes n}) A_{B_j}^{(n)},$$

Theorem 8.1 (First Law). *Under the above notations, if the regular sequences $\rho_{S^n B^n \underline{W}^n} = \rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes n} \otimes |\underline{w}\rangle\langle\underline{w}|^n$ and $\sigma_{S^n B^n \underline{W}^n} = \sigma_{S^n B^n} \otimes |\underline{w}'\rangle\langle\underline{w}'|^n$ are equivalent under almost-commuting unitaries, then*

$$s(\{\sigma_{S^n B^n}\}) = s(\{\rho_{S^n}\}) + S(\tau(\underline{\beta})) \text{ and}$$
$$W_j = -\Delta A_{S_j} - \Delta A_{B_j} \text{ for all } j = 1, \dots, c.$$

Conversely, given regular sequences ρ_{S^n} and $\sigma_{S^n B^n}$ of product states such that
$$s(\{\sigma_{S^n B^n}\}) = s(\{\rho_{S^n}\}) + S(\tau(\underline{\beta})),$$
and assuming that the spectral radius of the battery observables W_{A_j} is large enough (see the discussion at the start of this section), then there exist regular sequences of product states of the j-type battery, $|w_j\rangle\langle w_j|^n$ and $|w_j'\rangle\langle w_j'|^n$, for all $j = 1, \dots, c$, such that

$$\rho_{S^n B^n \underline{W}^n} = \rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes n} \otimes |\underline{w}\rangle\langle\underline{w}|^n \text{ and} \tag{8.4}$$
$$\sigma_{S^n B^n \underline{W}^n} = \sigma_{S^n B^n} \otimes |\underline{w}'\rangle\langle\underline{w}'|^n \tag{8.5}$$

can be transformed into each other by almost-commuting unitaries.

Proof. The first part is by definition, since the almost-commuting unitaries asymptotically preserve the entropy rate and the work rate of all charges.

In the other direction, all we have to do is find states $|w_j\rangle\langle w_j|$ and $|w_j'\rangle\langle w_j'|$ of the j-type battery W_j, such that $W_j = \Delta A_{W_j} = -\Delta A_{S_j} - \Delta A_{B_j}$, for all $j = 1, \ldots, c$. This is clearly possible if the spectral radius of W_{A_j} is large enough. With this, the states in Eqs. (8.4) and (8.5) have the same asymptotic entropy and charge rates. Hence, the claim follows from the AET, Theorem 7.1. ∎

Remark 8.1. *The second part of Theorem 8.1 says that for regular product state sequences, as long as the initial and final states of the work system and the thermal bath have asymptotically the same entropy, they can be transformed one into the another because there are always batteries that can absorb or release the necessary charge difference. Furthermore, we can even fix the initial (or final) state of the batteries and design the matching final (initial) battery state, assuming that the charge expectation value of the initial (final) state is far enough from the edge of the spectrum of A_{W_j}.*

For any such states, we say that there is a work transformation *taking one to the other, denoted $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes n} \to \sigma_{S^n B^n}$. This transformation is always feasible, implicitly assuming the presence of suitable batteries for all j-type works to balance to books.*

Remark 8.2. *As a consequence of the previous remark, we now change our point of view what a transformation is. Of our complicated S-B-$\underline{W}$ compound, we only focus on SB and its state, and treat the batteries as implicit. Since we insist that batteries need to remain in a pure state, which thus factors off and does not contribute to the entropy, and due to the above first law Theorem 8.1, we can indeed understand everything that is going on by looking at how $\rho_{S^n B^n}$ transforms into $\sigma_{S^n B^n}$.*

Note that in this context, it is in a certain sense enough that the initial state ρ_{S^n} is a regular sequence of product states and that the target states $\sigma_{S^n B^n}$ are a regular sequence. This is because the first part of the First Law Theorem 8.1 only requires regularity, and since the target state defines a unique point $(\underline{a}', s')$ in the phase diagram, we can find a sequence of product states $\widetilde{\sigma}_{S^n B^n}$ in its equivalence class, and use the second part of Theorem 8.1 to realise the work transformation $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes n} \to \widetilde{\sigma}_{S^n B^n}$.

8.2 The second law

If the first law in our framework arises from focusing on the system-plus-bath compound SB, while making the batteries implicit, the second law comes

about from trying to understand the action on the work system S alone. Following [100, 105], the second law constrains the different combinations of commuting conserved quantities that can be extracted from the work system. We show here that in the asymptotic regime, the second law similarly bounds the extractable work rate via the rate of free entropy of the system.

The *free entropy* for a system with state ρ, charges A_j and inverse temperatures β_j is defined in [105] as

$$\widetilde{F}(\rho) = \sum_{j=1}^{c} \beta_j \operatorname{Tr} \rho A_j - S(\rho). \tag{8.6}$$

It is shown in [105] that the generalized thermal state $\tau(\underline{\beta})$ is the state that minimizes the free entropy for fixed β_j.

For any work transformation $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes n} \to \sigma_{S^n B^n}$ between regular sequences of states, we define the asymptotic rate of free entropy change for the work system and the thermal bath respectively as follows:

$$\begin{aligned}
\Delta \widetilde{F}_S &:= \lim_{n \to \infty} \frac{1}{n} \left(\widetilde{F}(\sigma_{S^n}) - \widetilde{F}(\rho_{S^n}) \right), \\
\Delta \widetilde{F}_B &:= \lim_{n \to \infty} \frac{1}{n} \left(\widetilde{F}(\sigma_{B^n}) - n \widetilde{F}(\tau_B) \right),
\end{aligned} \tag{8.7}$$

where the free entropy is with respect to the charges of the work system and the thermal bath with fixed inverse temperatures β_j.

Theorem 8.2 (Second Law). *For any work transformation $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes n} \to \sigma_{S^n B^n}$ between regular sequences of states, the j-type works W_j that are extracted (and they are necessarily $W_j = -\Delta A_{S_j} - \Delta A_{B_j}$ according to the first law) are constrained by the rate of free entropy change of the system:*

$$\sum_{j=1}^{c} \beta_j W_j \le -\Delta \widetilde{F}_S.$$

Conversely, for arbitrary regular sequences of product states, ρ_{S^n} and σ_{S^n}, and any real numbers W_j with $\sum_{j=1}^{c} \beta_j W_j < -\Delta \widetilde{F}_S$, there exists a bath system B and a regular sequence of product states $\sigma_{S^n B^n}$ with $\operatorname{Tr}_{B^n} \sigma_{S^n B^n} = \sigma_{S^n}$, such that there is a work transformation $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes n} \to \sigma_{S^n B^n}$ with accompanying extraction of j-type work at rate W_j.

Proof. We start with the first statement of the theorem. Consider the global system transformation $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes n} \to \sigma_{S^n B^n}$ by almost-commuting unitaries. We use the definition of work (8.3) and free entropy (8.6), as well as the first law, Theorem 8.1, to get

$$\begin{aligned}
\sum_{j} \beta_j W_j &= -\sum_{j} \beta_j \left(\Delta A_{S_j} + \Delta A_{B_j} \right) \\
&= -\Delta \widetilde{F}_S - \Delta \widetilde{F}_B - \Delta s_S - \Delta s_B.
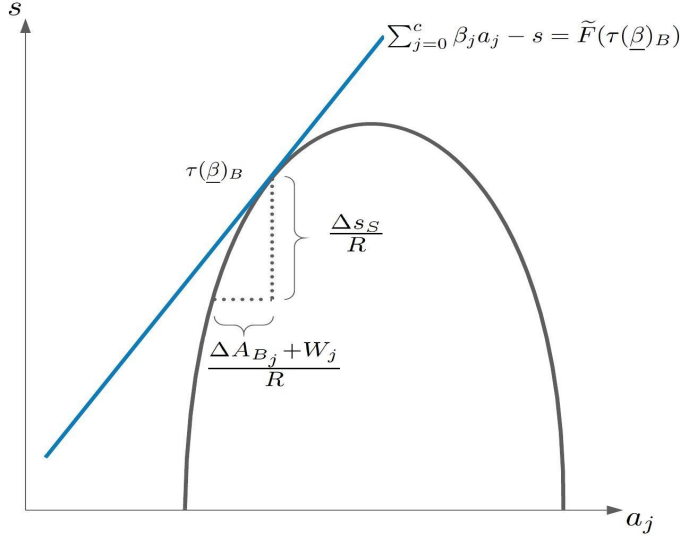\end{aligned} \tag{8.8}$$

134

Figure 8.1: State change of the bath for a given work transformation under works W_j extracted, viewed in the phase diagram of the bath $\overline{\mathcal{P}}_B$. The blue line represents the tangent hyper-plane at the corresponding point of the generalized thermal state $\tau(\underline{\beta})_B$.

The second line is due to the definition in Eq. (8.7). Now observe that

$$
\begin{aligned}
\Delta s_S + \Delta s_B &= \lim_{n \to \infty} \frac{1}{n} \big(S(\sigma_{S^n}) - S(\rho_{S^n}) \big) + \frac{1}{n} \big(S(\sigma_{B^n}) - nS(\tau(\underline{\beta})_B) \big) \\
&\geq \lim_{n \to \infty} \frac{1}{n} \big(S(\sigma_{SB^n}) - S(\rho_{S^n}) - S(\tau(\underline{\beta})_B^{\otimes n}) \big) = 0,
\end{aligned}
\tag{8.9}
$$

where the inequality is due to sub-additivity of von Neumann entropy, and the final equation due to asymptotic entropy conservation. Further, the generalized thermal state $\tau(\underline{\beta})_B$ has the minimum free entropy [105], hence $\Delta \widetilde{F}_B \geq 0$.

For the second statement of the theorem, the achievability part of the second law, we aim to show that there is a work transformation $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes n} \to \sigma_{S^n} \otimes \xi_{B^n}$, with a suitable regular sequences of product states, and works $W_1, \ldots, W_c$ are extracted. This will be guaranteed, by the first law, Theorem 8.1, and the AET, Theorem 7.1, if

$$
\begin{aligned}
s(\{\xi_{B^n}\}) &= S(\tau(\underline{\beta})_B) - \Delta s_S, \\
a_j(\{\xi_{B^n}\}) &= \operatorname{Tr} \tau(\underline{\beta})_B A_{B_j} - \Delta A_{S_j} - W_j \quad \text{for all } j = 1, \ldots, c.
\end{aligned}
\tag{8.10}
$$

135

The left hand side here defines a point $(\underline{a}, s)$ in the charges-entropy space of the bath, and our task is to show that it lies in the phase diagram, for which purpose we have to define the bath characteristics suitably. On the right hand side, $\left(\operatorname{Tr}\tau(\underline{\beta})_B A_{B_1}, \ldots, \operatorname{Tr}\tau(\underline{\beta})_B A_{B_c}, S(\tau(\underline{\beta})_B)\right)$ is the point corresponding to the initial state of the bath, which due to its thermal nature is situated on the upper boundary of the region. At that point, the region has a unique tangent hyperplane, which has the equation $\sum_j \beta_j a_j - s = \widetilde{F}(\tau(\underline{\beta})_B)$, and the phase diagram is contained in the half space $\sum_j \beta_j a_j - s \geq \widetilde{F}(\tau(\underline{\beta})_B)$, corresponding to the fact that their free entropy is larger than that of the thermal state. In fact, due to the strict concavity of the entropy, and hence of the upper boundary of the phase diagram, the phase diagram, with the exception of the thermal point $\left(\operatorname{Tr}\tau(\underline{\beta})_B\underline{A_B}, S(\tau(\underline{\beta})_B)\right)$ is contained in $\sum_j \beta_j a_j - s > \widetilde{F}(\tau(\underline{\beta})_B)$.

One of many ways to construct a suitable bath B is as many $(R \gg 1)$ non-interacting copies of a system b: $B = b^R$ and charges $A_{B_j} = A_{b_j}^{(R)}$, so that the GGS of B is $\tau(\underline{\beta})_B = \tau(\underline{\beta})_b^{\otimes R}$. We claim that for large enough R, the left hand side of Eq. (8.10) defines a point in the phase diagram of B. Indeed, we can express the conditions in terms of b, assuming that we aim for a regular sequence of product states $\xi_{b^{nR}}$:

$$
\begin{aligned}
s(\{\xi_{b^{nR}}\}) &= S(\tau(\underline{\beta})_b) - \frac{1}{R}\Delta s_S, \\
a_j(\{\xi_{b^{nR}}\}) &= \operatorname{Tr}\tau(\underline{\beta})_b A_{b_j} - \frac{1}{R}(\Delta A_{S_j} + W_j) \quad \text{for all } j = 1, \ldots, c.
\end{aligned}
\tag{8.11}
$$

For all sufficiently large R, these points $(\underline{a}, s)$ are arbitrarily close to where the bath starts off, at $(\underline{a}_{\underline{\beta}}, s_{\underline{\beta}}) = \left(\operatorname{Tr}\tau(\underline{\beta})_b A_{b_1}, \ldots, \operatorname{Tr}\tau(\underline{\beta})_b A_{b_c}, S(\tau(\underline{\beta})_b)\right)$, while they always remains in the open half plane $\sum_j \beta_j a_j - s > \widetilde{F}(\tau(\underline{\beta})_b)$. Indeed, they all lie on a straight line pointing from $(\underline{a}_{\underline{\beta}}, s_{\underline{\beta}})$ into the interior of that half plane. Hence, for sufficiently large R, $(\underline{a}, s) \in \overline{\mathcal{P}}$, the phase diagram of b, and by point 5 of Lemma 7.1 there does indeed exist a regular sequence of product states corresponding to it. ∎

8.3 Finiteness of the bath: tighter constraints and negative entropy

In the previous two subsections we have elucidated the traditional statements of the first and second law of thermodynamics, as emerging in our resource theory. In particular, the second law is tight, if sufficiently large baths are allowed to be used.

Here, we specifically look at the the second statement (achievability) of the second law in the presence of an explicitly given, finite bath B. It will turn out that typically, equality in the second law cannot be attained, only up to a certain loss that is due to the finiteness of the bath. We also discover a purely quantum effect whereby the system and the bath remain entangled after effecting a certain state transformation, allowing quantum engines to perform tasks impossible classically (i.e. with separable correlations). The question we want to address is the following refinement of the one answered in the previous subsection:

> Given regular sequences ρ_{S^n} and σ_{S^n} of product states, and numbers W_j, are there extensions $\sigma_{S^n B^n}$ of σ_{S^n} forming a regular sequence of product states, such that the work transformation $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes n} \to \sigma_{S^n B^n}$ is feasible, with accompanying extraction of j-type work at rate W_j?

To answer it, we need the following *extended phase diagram*. For a give state σ_S of the system S, and a bath B, define the the following set:

$$\mathcal{P}_{|\sigma_S}^{(1)} := \left\{ \left(\operatorname{Tr} \xi_B A_1^{(B)}, \ldots, \operatorname{Tr} \xi_B A_c^{(B)}, S(B|S)_\xi \right) : \xi_{SB} \text{ state with } \operatorname{Tr}_B \xi_{SB} = \sigma_S \right\}, \tag{8.12}$$

furthermore its n-copy version

$$\mathcal{P}_{|\sigma_{S^n}}^{(n)} := \tag{8.13}$$
$$\left\{ \left(\operatorname{Tr} \xi_{B^n} A_1^{(B^n)}, \ldots, \operatorname{Tr} \xi_{B^n} A_c^{(B^n)}, S(B^n|S^n)_\xi \right) : \xi_{S^n B^n} \text{ state with } \operatorname{Tr}_{B^n} \xi_{S^n B^n} = \sigma_S^{\otimes n} \right\}. \tag{8.14}$$

Finally, define the *conditional entropy phase diagram* as

$$\overline{\mathcal{P}}_{|s} := \overline{\mathcal{P}}_{|s}^{(1)} := \tag{8.15}$$
$$\left\{ \left(\underline{a}, s' \right) : a_j = \operatorname{Tr} \xi_B A_j^{(B)}, \ -\min\{s, S(\tau(\underline{a}))\} \le s' \le S(\tau(\underline{a})) \text{ for a state } \xi_B \right\}, \tag{8.16}$$

and likewise its n-copy version $\overline{\mathcal{P}}_{|ns}^{(n)}$, for a number s (intended to be an entropy or entropy rate). The relation between the sets, and the name of the latter, are explained in the following lemma.

Lemma 8.1. *With the previous notation, we have:*

1. *For all k, $\mathcal{P}_{|\sigma_{S^k}}^{(k)} \subset \overline{\mathcal{P}}_{|S(\sigma_{S^k})}^{(k)}$, and the latter is a closed convex set.*
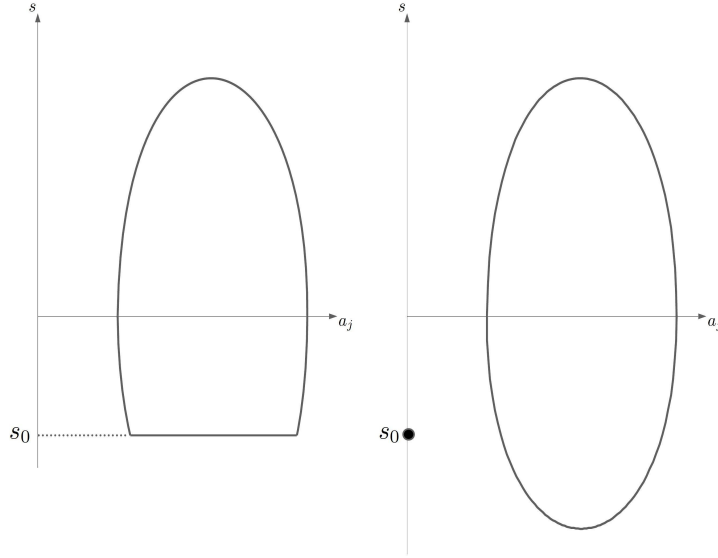
137

Figure 8.2: Schematic of the extended phase diagram $\overline{\mathcal{P}}_{|s_0}$. Depending on the value of s_0, the diagram may acquire either of the above shapes.

2. For all k, $\overline{\mathcal{P}}_{|ks}^{(k)} = k\overline{\mathcal{P}}_{|s}^{(1)}$.

3. For a regular sequence $\{\sigma_{S^k}\}$ of product states with entropy rate $s = s(\{\sigma_{S^k}\})$, every point in $\overline{\mathcal{P}}_{|s}$ is arbitrarily well approximated by points in $\frac{1}{k}\mathcal{P}_{|\sigma_{S^k}}^{(k)}$ for all sufficiently large k. I.e., $\overline{\mathcal{P}}_{|s} = \lim_{k \to \infty} \frac{1}{k}\mathcal{P}_{|S(\sigma_{S^k})}^{(k)}$.

Proof. 1. We only have to convince ourselves that for a state $\xi_{S^k B^k}$ with $\operatorname{Tr}_{B^k} \xi_{S^k B^k} = \sigma_{S^k}$,

$$-\min\{S(\sigma_{S^k}), kS(\tau(\underline{a}))\} \le S(B^k|S^k)_\xi \le kS(\tau(\underline{a})),$$

where $\underline{a} = (a_1, \ldots, a_c)$ with $a_i = \frac{1}{k}\operatorname{Tr}\xi_{B^k}A_i^{(B^k)}$. The upper bound follows from subadditivity, since $S(B^k|S^k)_\xi \le S(B^k)_\xi \le kS(\tau(\underline{a}))$. The lower bound consists of two inequalities: first, by purifying ξ to a state $|\phi\rangle \in S^k B^k R$ and strong subadditivity, $S(B^k|S^k)_\xi \ge S(B^k|S^kR)_\phi = -S(B^k)_\xi \ge -kS(\tau(\underline{a}))$. Secondly, $S(B^k|S^k)_\xi \ge -S(S^k)_\xi = -S(\sigma_{S^k})$.

2. Follows easily from the definition.

3. It is enough to show that the points of the minimum entropy diagram

$$\overline{\mathcal{P}}_{\min|s} := \left\{\left(\underline{a}, -\min\{s, S(\tau(\underline{a}))\}\right) : \operatorname{Tr}\xi_B A_j^{(B)} = a_j \text{ for a state } \xi_B\right\}$$

138

can be approximated as claimed by an admissible k-copy state $\xi_{S^k B^k}$. This is because the maximum entropy diagram $\overline{\mathcal{P}}_{\max}^{(k)}$ is realized by states $\vartheta_{S^k B^k} :=$ $\sigma_{S^k} \otimes \tau(\underline{a})_B^{\otimes k}$, and by interpolating the states, i.e. $\lambda \xi + (1-\lambda)\vartheta$ for $0 \leq \lambda \leq 1$, we can realize the same charge values $\underline{a}$ with entropies in the whole interval $[S(B^k|S^k)_\xi; kS(\tau(\underline{a}))]$.

The approximation of $\overline{\mathcal{P}}_{\min|s}$ can be proved invoking results from quantum Shannon theory. For this, consider a purification $|\Psi\rangle \in S^k B^k R^k$ of the state $\sigma_{S^k} \otimes \tau(\underline{a})_B^{\otimes k}$, which can be chosen in such a way as to be a product state itself: $|\Psi\rangle = |\Psi_1\rangle_{S_1 B_1 R_1} \otimes \cdots \otimes |\Psi_k\rangle_{S_k B_k R_k}$. Now let D be an integer with $\log D \geq |S(\sigma_{S^k}) - kS(\tau(\underline{a}))| + \epsilon k$, for any small $\epsilon > 0$, and consider a Haar random projector P on R^k of rank D. Then, for sufficiently large k, with high probability the state

$$\xi(P)_{S^k B^k} := \frac{1}{\operatorname{tr} \Psi_{R^k} P} \operatorname{tr}_{R^k} |\Psi\rangle\langle\Psi| (\mathbb{1}_{S^k B^k} \otimes P_{R^k})$$

satisfies

$$\frac{1}{2} \|\xi(P)_{S^k} - \sigma_{S^k}\|_1 \leq \epsilon,$$

$$\frac{1}{2} \|\xi(P)_{B^k} - \tau(\underline{a})_B^{\otimes k}\|_1 \leq \epsilon.$$

This follows from the general one-shot decoupling theorem [108, Thm. 3.1], applied simultaneously to $\Psi_{S^k R^k}$ and $\Psi_{B^k R^k}$, in conjunction with the asymptotic equipartition property of the conditional min-entropy [109].

Fix using Uhlmann's theorem (on a purification of σ_{S^k} and a purification of $\xi(P)_{S^k B^k}$) and the well-known relations between fidelity and trace norm, we can find $\xi_{S^k B^k}$ with $\xi_{S^k} = \sigma_{S^k}$ and $\frac{1}{2} \|\xi(P)_{S^k B^k} - \xi_{S^k B^k}\|_1 \leq \sqrt{\epsilon(2-\epsilon)}$, thus

$$\frac{1}{2} \|\xi_{B^k} - \tau(\underline{a})_B^{\otimes k}\|_1 \leq \epsilon + \sqrt{\epsilon(2-\epsilon)}.$$

From the latter bound it follows that $\frac{1}{k} \operatorname{tr} \xi_{B^k} A_j^{(B^k)} \approx a_j$. It remains to bound the conditional entropy $S(B^k|S^k)_\xi$:

$$\begin{aligned}
\frac{1}{k} S(B^k|S^k)_\xi &= \frac{1}{k} S(\xi_{S^k B^k}) - \frac{1}{k} S(\xi_{S^k}) \\
&\approx \frac{1}{k} S(\xi(P)_{S^k B^k}) - \frac{1}{k} S(\sigma_{S^k}) \\
&\leq \frac{1}{k} \log D - \frac{1}{k} S(\sigma_{S^k}) \\
&\leq \frac{1}{k} |S(\sigma_{S^k}) - kS(\tau(\underline{a}))| - \frac{1}{k} S(\sigma_{S^k}) + \epsilon \\
&= -\frac{1}{k} \min \{S(\sigma_{S^k}), kS(\tau(\underline{a}))\} + \epsilon,
\end{aligned}$$

139

where in the second line we have used the Fannes inequality on the continuity of the entropy, in the third line that $\xi(P)_{S^k B^k}$ has rank at most D, and in the fourth line the upper bound on D by construction. This concludes the proof. $\blacksquare$

Coming back to our question, if a work transformation $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes n} \to \sigma_{S^n B^n}$ is feasible for regular sequences on the left hand side, by the first law this implies that

$$s(\{\sigma_{S^n B^n}\}) = s(\{\rho_{S^n}\}) + S(\tau(\underline{\beta})) \text{ and}$$
$$W_j = -\Delta A_{S_j} - \Delta A_{B_j}$$
$$= a_j(\{\rho_{S^n}\}) - a_j(\{\sigma_{S^n}\}) + a_j(\{\tau(\underline{\beta})_{B^n}\}) - a_j(\{\sigma_{B^n}\}).$$

When σ_{S^n} and the W_j are given, this constrains the possible states $\sigma_{S^n B^n}$ as follows: for each n,

$$\frac{1}{n} S(B^n | S^n)_\sigma \approx S(\tau(\underline{\beta})) - \Delta s_S,$$
$$\frac{1}{n} \text{Tr} \, \sigma_{B^n} A_{B_j}^{(n)} \approx \text{Tr} \, \tau(\underline{\beta})_B A_{B_j} - \Delta A_{S_j} - W_j, \quad \text{for all } j = 1, \dots, c.$$

Since by Lemma 8.1 the left hand sides converge to the components of a point in $\overline{\mathcal{P}}_{|s(\{\sigma_{S^n}\})}$, meaning that a necessary condition for the feasibility of the work transformation in question is that

$$(\underline{a}, t) \in \overline{\mathcal{P}}_{|s(\{\sigma_{S^n}\})}, \text{ with } a_j := \text{Tr} \, \tau(\underline{\beta})_B A_{B_j} - \Delta A_{S_j} - W_j,$$
$$t := S(\tau(\underline{\beta})) - \Delta s_S. \tag{8.17}$$

Again by Lemma 8.1, this is equivalent to all a_j to be contained in the set of joint quantum expectations of the observables A_{B_j}, and

$$-\min\{s(\{\sigma_{S^n}\}), S(\tau(\underline{a}))\} \le t \le S(\tau(\underline{a})).$$

The following theorem shows that this is also sufficient, when we allow blockings of the asymptotically many systems.

Theorem 8.3 (Second Law with fixed bath)**.** *For arbitrary regular sequences ρ_{S^n} and σ_{S^n} of product states, a given bath B, and any real numbers W_j, if there exists a regular sequence of block product states $\sigma_{S^n B^n}$ with $\text{Tr}_{B^n} \sigma_{S^n B^n} = \sigma_{S^n}$, such that there is a work transformation $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes n} \to \sigma_{S^n B^n}$ with accompanying extraction of j-type work at rate W_j, then Eq. (8.17) defines a point $(\underline{a}, t) \in \overline{\mathcal{P}}_{|s(\{\sigma_{S^n}\})}$.*
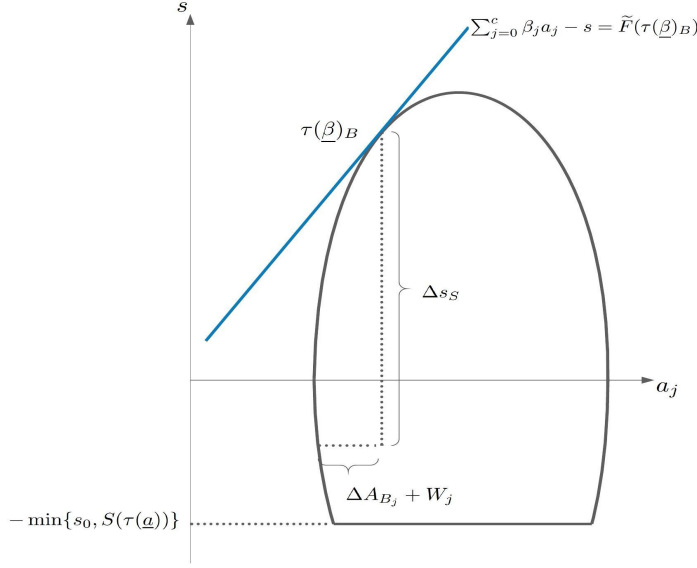
140

Figure 8.3: State change of the bath for a given work transformation under works W_j extracted, viewed in the extended phase diagram of the bath. The blue line represents the tangent hyper-plane at the corresponding point of the generalized thermal state $\tau(\underline{\beta})_B$.

Conversely, assuming additionally that $\sigma_{S^n} = \sigma_S^{\otimes n}$ is an i.i.d. state, if Eq. (8.17) defines a point $(\underline{a}, t) \in \overline{\mathcal{P}}^0_{|S(\sigma_S)}$ in the interior of the extended phase diagram, then for every $\epsilon > 0$ there is a work transformation $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes n} \to \sigma_{S^n B^n}$ with block product states $\sigma_{S^n B^n}$ such that $\mathrm{Tr}_{B^n} \sigma_{S^n B^n} = \sigma_{S^n}$, and with accompanying extraction of j-type work at rate $W_j \pm \epsilon$.

Proof. We have already argued the necessity of the condition. It remains to show its sufficiency. Using Lemma 8.1, this is not hard: Namely, by its point 3, for sufficiently large k, $(\underline{a}, t) \in \overline{\mathcal{P}}_{|s}$ is ϵ-approximated by $\frac{1}{k}\mathcal{P}^{(k)}_{|\sigma_S^{\otimes k}}$, i.e. there exists a $\sigma_{S^k B^k}$ with $\mathrm{tr}_{B^k} \sigma_{S^k B^k} = \sigma_S^{\otimes k}$ with $\frac{1}{k}S(B^k|S^k)_\sigma \le t - \epsilon$ and $\frac{1}{k}\mathrm{tr}\,\sigma_{B^k} A_j^{(B^k)} \approx a_j$ for all $j = 1, \ldots, c$. By mixing σ with a small fraction of $\left(\tau(\underline{a})_B \otimes \sigma_S\right)^{\otimes k}$, we can in fact assume that $\frac{1}{k}S(B^k|S^k)_\sigma = t$ while preserving $\frac{1}{k}\mathrm{tr}\,\sigma_{B^k} A_j^{(B^k)} \approx a_j$. Now our target block product states will be $\sigma_{S^n B^n} := \left(\sigma_{S^k B^k}\right)^{\otimes \frac{n}{k}}$ for n a multiple of k. By construction, this sequence has the same entropy rate as the initial regular sequence of product states $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes n}$, so by the first law, Theorem 8.1, and the AET, Theorem 7.1, there is indeed

a corresponding work transformation with j-type work extracted equal to $W_j \pm \epsilon$. ∎

Remark 8.3. *One might object that tensor power target states are not general enough in Theorem 8.3, as we had observed in Section 7.1 that such states do not generate the full phase diagram $\overline{P}$ of the system S. However, by considering blocks of ℓ systems S^ℓ, we can apply the theorem to block tensor power target states $\sigma_{S^n} = \left(\sigma_1 \otimes \cdots \otimes \sigma_\ell\right)^{\otimes \frac{n}{\ell}}$, and these latter are in fact a rich enough class to exhaust the entire phase diagram $\overline{P}$, when $\ell \geq \dim S$ (point 5 of Lemma 7.1).*

More generally, we can allow as target uniformly regular *sequences of product states σ_{S^n}, by which we mean the following strengthening of the condition in Definition 8.1. Denoting $B_{N+1}^{N+n} \coloneqq B_{N+1} \ldots B_{N+n}$, we require that for all $\epsilon > 0$ and all sufficiently large n,*

$$\left| a_j - \frac{1}{n} \operatorname{Tr} \sigma_{B_{N+1}^{N+n}} A_j^{(n)} \right| < \epsilon, \ j = 1, \ldots, c, \ and$$

$$\left| s - \frac{1}{n} S(\sigma_{B_{N+1}^{N+n}}) \right| < \epsilon,$$

uniformly in N.

8.4 Tradeoff between thermal bath rate and work extraction

Here we consider a different take on the question of the work deficit due to finiteness of the bath. Namely, we still consider a given fixed finite bath system B, but now as which state transformations and associated generalized works are possible when for each copy of the subsystem S, $R \geq 0$ copies of B are present. It is clear what that means when R is an integer, but below we shall give a meaning to this rate as a real number. Our first result is that "large enough bath" in Theorem 8.2 can be taken to mean B^R for the given bath B and sufficiently large rate R.

To give meaning to a rational rate $R = \frac{\ell}{k}$, group the systems of S^n, $n = \nu k$, into blocks of k, which we denote $\widetilde{S} = S^k$, and consider $\rho_{S^n} \equiv \rho_{\widetilde{S}^\nu}$ as a ν-party state, and likewise $\sigma_{S^n} \equiv \sigma_{\widetilde{S}^\nu}$. For each $\widetilde{S} = S^k$ we assume ℓ copies of the thermal bath, $\tau(\underline{\beta})_B^{\otimes \ell} = \tau(\underline{\beta})_{\widetilde{B}}$, with $\widetilde{B} = B^\ell$. If $\{\rho_{S^n}\}$ and $\{\sigma_{S^n}\}$ are regular sequences of product states, then evidently so are $\{\rho_{\widetilde{S}^\nu}\}$ and $\{\sigma_{\widetilde{S}^\nu}\}$.

Now, for the given sequences $\{\rho_{S^n}\}$ and $\{\sigma_{S^n}\}$ of initial and final states, respectively, as well as works $W_1, \ldots, W_c$ satisfying $\sum_j \beta_j W_j = -\Delta \widetilde{F}_S - \delta$, $\delta \geq 0$, we can ask what is the infimum over all rates $R = \frac{\ell}{k}$ such that there is a work transformation

$$\rho_{S^n} \otimes \tau(\underline{\beta})_{B^{nR}} \equiv \rho_{\widetilde{S}^\nu} \otimes \tau(\underline{\beta})_{\widetilde{B}}^{\otimes \nu \ell} \to \sigma_{\widetilde{S}^\nu \widetilde{B}^{\nu \ell}} \equiv \sigma_{S^n B^{nR}},$$

where as before the final state is intended to satisfy $\mathrm{Tr}_{\widetilde{B}^{\nu \ell}} \sigma_{\widetilde{S}^\nu \widetilde{B}^{\nu \ell}} = \sigma_{\widetilde{S}^\nu}$.

We observe that if $S(\rho_{S^n}) = S(\sigma_{S^n})$ and $\sum_j \beta_j W_j = -\Delta \widetilde{F}_S$, then the work transformation is possible without using any thermal bath, which follows from Eq. (8.8). That is, the thermal bath is not necessary for extracting work if the entropy of the work system does not change. Conversely, the role of the thermal bath is precisely to facilitate changes of entropy in the work system.

To answer the above question after the minimum bath rate R^*, we first show the following lemma.

Lemma 8.2. *Consider regular sequences of product states, ρ_{S^n} and σ_{S^n}, and real numbers W_j, and assume that for large enough rate R there is a work transformation $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes nR} \to \sigma_{S^n B^{nR}}$, with σ_{S^n} as the reduced final state on the work system, and works $W_1, \ldots, W_c$ are extracted. Then there is another work transformation $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes nR} \to \sigma_{S^n} \otimes \xi_{B^{nR}}$, in which the final state of the work system and the thermal bath are uncorrelated, $\xi_{B^{nR}}$ is a regular sequence of product states, and the same works $W_1, \ldots, W_c$ are extracted.*

Proof. Assuming that $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes nR} \to \sigma_{S^n B^{nR}}$ is a work transformation, the second law implies that $\sum_j \beta_j W_j = -\Delta \widetilde{F}_s - \delta$ for some $\delta \geq 0$, and we obtain

$$
\begin{aligned}
s(\{\sigma_{B^{nR}}\}) &= S(\tau(\underline{\beta})_B) - \frac{1}{R}\Delta s_S + \frac{\delta'}{R}, \\
a_j(\{\sigma_{B^{nR}}\}) &= \mathrm{Tr}\,\tau(\underline{\beta})_B A_{B_j} - \frac{1}{R}(\Delta A_{S_j} + W_j) \quad \text{for all } j = 1, \ldots, c.
\end{aligned}
\tag{8.18}
$$

for $0 \leq \delta' \leq \delta$ where the first equality is due to the fact that $\Delta \widetilde{F}_B + \Delta s_S + \Delta s_B = \delta$ as seen in Eq. (8.8) and positivity of the entropy rate change from Eq. (8.9). The second equality follows from the first law, Theorem 8.1, and the AET, Theorem 7.1. If R is large enough, due to the convexity of the phase diagram of the thermal bath $\overline{\mathcal{P}}_B^{(1)}$, the following coordinates belong to the phase diagram as well

$$
\begin{aligned}
s(\{\xi_{B^{nR}}\}) &= S(\tau(\underline{\beta})_B) - \frac{1}{R}\Delta s_S, \\
a_j(\{\xi_{B^{nR}}\}) &= \mathrm{Tr}\,\tau(\underline{\beta})_B A_{B_j} - \frac{1}{R}(\Delta A_{S_j} + W_j) \quad \text{for all } j = 1, \ldots, c.
\end{aligned}
\tag{8.19}
$$

143

Therefore, due to point 3 and 5 of Lemma 7.1, there is a state tensor product state ξ_{B^nR} with coordinate of Eq. (8.19) on $\overline{\mathcal{P}}_B^{(1)}$, so the first law, Theorem 8.1 implies that the desired transformation exists, and works $W_1, \ldots, W_c$ are extracted. ∎

Theorem 8.4. *For regular sequences of product states, ρ_{S^n} and σ_{S^n}, and real numbers W_j satisfying $\sum_j \beta_j W_j = -\Delta\widetilde{F}_s - \delta$, let R^* be the infimum of rates such that there is a work transformation $\rho_{S^n} \otimes \tau(\underline{\beta})_B^{\otimes nR} \to \sigma_{S^n} \otimes \xi_{B^nR}$ under which works $W_1, \ldots, W_c$ are extracted, and ξ_{B^nR} is a regular sequence of product states.*

Then, this minimum R^* is achieved for a state ξ_{B^nR} on the boundary of the phase diagram $\overline{\mathcal{P}}_B$ of the thermal bath.

For $\delta \ll 1$, the minimum rate is

$$R \approx \frac{1}{2\delta} \sum_{ij} \frac{\partial\beta_j}{\partial a_i}(-\Delta A_{S_i} - W_i)(-\Delta A_{S_j} - W_j), \tag{8.20}$$

where $\Delta A_{S_j} = a(\{\sigma_{S^n}\}) - a(\{\rho_{S^n}\})$.

Proof. The final state of the thermal bath ξ_{B^nR} is a tensor product state, so the first law, Theorem 8.1, and the AET, Theorem 7.1 imply that

$$
\begin{aligned}
s(\{\xi_{B^nR}\}) &= S(\tau(\underline{\beta})_B) - \frac{1}{R}\Delta s_S, \\
a_j(\{\xi_{B^nR}\}) &= \operatorname{Tr}\tau(\underline{\beta})_B A_{B_j} - \frac{1}{R}(\Delta A_{S_j} + W_j) \quad \text{for all } j = 1, \ldots, c,
\end{aligned}
\tag{8.21}
$$

where $\Delta s_S = s(\{\sigma_{S^n}\}) - s(\{\rho_{S^n}\})$. Due to point 3 of Lemma 7.1, the above coordinates belong to $\overline{\mathcal{P}}_B^{(1)}$. For $R = R_0$ assume that the above coordinates belong to the point $(\underline{a}_0, s_0)$ on the boundary of the phase diagram $\overline{\mathcal{P}}_B^{(1)}$. Then, for $R > R_0$ the point of Eq. (8.21) is the convex combination of the points $(\underline{a}_0, s_0)$ and the corresponding point of the state $\tau(\underline{\beta})_B$, so it belongs to the phase diagram due to its convexity. Therefore, all points with $R > R_0$ are inside the diagram.

To approximate the minimum R, define the function $S(\underline{a}) := S(\tau(\underline{a})_B))$ for $\underline{a} = (a_1, \ldots, a_c)$. Consider the Taylor approximation of this function at the corresponding point of the thermal state $S\big(\tau(\underline{a}_0)_B\big)$ of the bath:

$$S(\underline{a}) \approx S(\underline{a}_0) + \sum_j \beta_j(a_j - a_{j0}) + \frac{1}{2}\sum_{ij} \frac{\partial\beta_j}{\partial a_i}(a_j - a_{j0})(a_i - a_{i0}),$$

144

where we use the well-know relation $\frac{\partial S}{\partial a_i} = \beta_i$. From Eq. (8.21), we obtain

$$S(\underline{a}) - S(\underline{a}_0) = -\frac{\Delta s_S}{R},$$
$$a_j - a_{j0} = \frac{1}{R}(-\Delta A_{S_j} - W_j),$$

and by substituting these values in the Taylor approximation, the theorem follows. ∎

Remark 8.4. *For a single charge, $c = 1$, which we traditionally interpret as the internal energy E of a system, Eq. (8.20) takes on the very simple form*

$$R \approx \frac{1}{2\delta}\frac{\partial \beta}{\partial E}(\Delta E + W)^2,$$

and where we can use the usual thermodynamic definitions to rewrite $\frac{\partial \beta}{\partial E} = \frac{\partial \frac{1}{T}}{\partial E} = -\frac{1}{T^2}\frac{1}{C}$, with the heat capacity $C = \frac{\partial E}{\partial T}$, all taken at the corresponding Gibbs equilibrium states. This gives a clear operational interpretation of the heat capacity in terms of the rate of the bath to approach the second law tightly.

For larger numbers of charges, the matrix $\left[\frac{\partial \beta_j}{\partial a_i}\right]_{ij} = \left[\frac{\partial^2 S}{\partial a_i \partial a_j}\right]_{ij}$ is actually the Hessian of the entropy $S(\tau(\underline{a}))$ with respect to the charges, and the r.h.s. side of Eq. (8.20) is $\frac{1}{2\delta}$ times the corresponding quadratic form evaluated on the vector $(-\Delta A_{S_1} - W_1, \ldots, -\Delta A_{S_c} - W_c)$. In the same vein as the single-parameter discussion before, the Hessian matrix can be read as being composed of generalized heat capacities, which likewise receive their operational interpretation in terms of the required rate of the bath.

8.5   Discussion

The traditional framework of thermodynamics assumes a system containing an asymptotically large number of particles interacts with an even larger bath. So that all the thermodynamic quantities of interest, e.g., energy, entropy, etc., can be expressed in terms of average or mean values. Also, the notion of temperature there remains meaningful as any exchange of energy hardly drives the bath away from equilibrium as it is considerably large. The quantum thermodynamics attempts to go beyond this assumption. For instance, the system that interacts with a large bath may have a fewer number of quantum particles. In this case, the average quantities are not sufficient

145

to characterize the system as there may be large quantum fluctuations that cannot be ignored. To address this issue, the resource theory of quantum thermodynamics is developed and it shows that the classical laws are not sufficient to characterize the thermodynamic transformations. One rather needs many second laws associated with many one-shot free energies (based on Renyi α-relative entropies) [110, 111]. However, this formalism is still not enough to study the situation where a quantum system interacts with a bath and they are of comparable size. Clearly, the very notion of temperature is questionable as the bath may get driven out of equilibrium after an interaction with the system. To address this, a resource theory is developed based on information conservation [54, 98] and it is only applicable to the regime where asymptotically large number system-bath composites are considered. This in turn also allows one to consider the system and bath on the same footing.

Here we have developed a resource theoretic formalism applicable to a more general scenario where a system with multiple conserved quantities (i.e., charges) interacts with a bath, and the system and bath may be of comparable size. These charges may not commute with each other, as allowed by quantum mechanics. The non-commutative nature implies that any (unitary) evolution cannot strictly conserve all these changes simultaneously. We overcome this problem by considering the notion of approximate micro-canonical ensembles, initially developed in [100]. This is an essential requirement and forms the basis of the (approximate) first law for thermodynamics with non-commuting charges. With this, we have developed a resource theory for work and heat for thermodynamics with non-commuting charges. We introduce the charge-entropy diagram that conceptually captures all the essential aspects of thermodynamics and an equivalence theorem to show the thermodynamic equivalence between quantum states sharing the same point on the charge-entropy diagram. Then we have derived the second law with the help of the diagram to characterize the state transformations and to quantify the thermodynamics resources such as works corresponding to different charges. We have also considered the situation where the bath is finite and quantified the rate of state transformations. Interestingly the rate of transformation has been shown to have a direct link with the generalized heat-capacity of the bath. All these then extended to the cases where the systems have (quantum) correlation with the bath. There the charge-entropy diagram has been expressed in terms of conditional-entropy of the bath which may get negative in presence of entanglement and, using that, the second law has been derived.

Appendix A

Miscellaneous definitions and facts

In this Appendix, we list a number of useful definitions and facts that we often refer to in various chapters.

For an operator X, the *trace norm*, the *Hilbert-Schmidt norm* and the *operator norm* are defined respectively in terms of $|X| = \sqrt{X^\dagger X}$:

$$\|X\|_1 = \operatorname{Tr}|X|,$$
$$\|X\|_2 = \sqrt{\operatorname{Tr}|X|^2},$$
$$\|X\|_\infty = \lambda_{\max}(|X|),$$

where $\lambda_{\max}(X)$ is the largest eigenvalue of X.

Lemma A.1 (Cf. [112]). *For any operator X,*

$$\|X\|_1 \le \sqrt{d}\|X\|_2 \le d\|X\|_\infty, \qquad\qquad (A.1)$$

where d equals the rank of X. ∎

Lemma A.2 (Cf. [112]). *For any self-adjoint operator X,*

$$\|X\|_1 = \max_{-\mathbb{1} \le Q \le \mathbb{1}} \operatorname{Tr}(QX).$$ ∎

Lemma A.3 (Cf. [112]). *For any self-adjoint operator X and any operator T,*

$$\left\|TXT^\dagger\right\|_1 \le \|T\|_\infty^2 \|X\|_1.$$ ∎

Lemma A.4 (Cf. Bhatia [112]). *For operators A, B and C and for any norm $p \in [1, \infty]$ the following holds*

$$\|ABC\|_p \le \|A\|_\infty \|B\|_p \|C\|_\infty.$$

Lemma A.5 (Hoeffding's inequality, Cf. [103])**.** *Let $X_1, X_2, \ldots, X_n$ be independent random variables with $a_i \leq X_i \leq b_i$. Define the empirical mean of these variables as $\overline{X} = \frac{X_1 + \ldots + X_n}{n}$, then for any $t > 0$*

$$\Pr\left\{\overline{X} - \mathbb{E}(\overline{X}) \geq t\right\} \leq \exp\left(-\frac{2n^2 t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right),$$

$$\Pr\left\{\overline{X} - \mathbb{E}(\overline{X}) \leq -t\right\} \leq \exp\left(-\frac{2n^2 t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

The fidelity of two states is defined as

$$F(\rho, \sigma) = \operatorname{Tr}\sqrt{\sigma^{\frac{1}{2}}\rho\sigma^{\frac{1}{2}}}.$$

When one of the arguments is pure, then

$$F(\rho, |\psi\rangle\langle\psi|) = \sqrt{\operatorname{Tr}\left(\rho\,|\psi\rangle\langle\psi|\right)} = \sqrt{\langle\psi|\,\rho\,|\psi\rangle}.$$

Lemma A.6. *The fidelity is related to the trace norm as follows [57]:*

$$1 - F(\rho, \sigma) \leq \frac{1}{2}\|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho, \sigma)^2} =: P(\rho, \sigma),$$

where $P(\rho, \sigma)$ is the so-called purified distance, or Bhattacharya distance, between quantum states. ∎

Lemma A.7 (Pinsker's inequality, cf. [19])**.** *The trace norm and relative entropy are related by*

$$\|\rho - \sigma\|_1 \leq \sqrt{2\ln 2\, S(\rho\|\sigma)}.$$

∎

Lemma A.8 (Uhlmann [83])**.** *Let ρ^A and σ^A be two quantum states with fidelity $F(\rho^A, \sigma^A)$. Let ρ^{AB} and σ^{AC} be purifications of these two states, then there exists an isometry $V : B \to C$ such that*

$$F\left((\mathbb{1}_A \otimes V^{B \to C})\rho^{AB}(\mathbb{1}_A \otimes V^{B \to C})^\dagger, \sigma^{AC}\right) = F(\rho^A, \sigma^A).$$ ∎

A consequence of this, due to [97, Lemma 2.2], is as follows.

Lemma A.9. *Let ρ^A and σ^A be two quantum states with trace distance $\frac{1}{2}\|\rho^A - \sigma^A\|_1 \leq \epsilon$, and let ρ^{AB} and σ^{AC} be purifications of these two states. Then there exists an isometry $V : B \to C$ such that*

$$\left\|(\mathbb{1}_A \otimes V^{B \to C})\rho^{AB}(\mathbb{1}_A \otimes V^{B \to C})^\dagger - \sigma^{AC}\right\|_1 \leq \sqrt{\epsilon(2 - \epsilon)}.$$ ∎

Lemma A.10 (Fannes [72]; Audenaert [73]). *Let ρ and σ be two states on Hilbert space A with trace distance $\frac{1}{2}\|\rho - \sigma\|_1 \le \epsilon$, then*

$$|S(\rho) - S(\sigma)| \le \epsilon \log |A| + h(\epsilon),$$

where $h(\epsilon) = -\epsilon \log \epsilon - (1-\epsilon)\log(1-\epsilon)$ is the binary entropy.

There is also an extension of the Fannes inequality for the conditional entropy; this lemma is very useful especially when the dimension of the system conditioned on is unbounded.

Lemma A.11 (Alicki-Fannes [76]; Winter [77]). *Let ρ and σ be two states on a bipartite Hilbert space $A \otimes B$ with trace distance $\frac{1}{2}\|\rho - \sigma\|_1 \le \epsilon$, then*

$$|S(A|B)_\rho - S(A|B)_\sigma| \le 2\epsilon \log |A| + (1+\epsilon)h\left(\frac{\epsilon}{1+\epsilon}\right).$$

∎

Lemma A.12. *Let ρ be a state with full support on the Hilbert space A, i.e. it has positive minimum eigenvalue $\lambda_{\min}$, and let $|\psi\rangle^{AR}$ be a purification of ρ on the Hilbert space $A \otimes R$. Then any purification of another state σ on A is of the form*

$$(\mathbb{1}_A \otimes T)|\psi\rangle^{AR},$$

where T is an operator acting on system R with $\|T\|_\infty \le \frac{1}{\sqrt{\lambda_{\min}}}$.

Proof. Let $\rho = \sum_i \lambda_i |e_i\rangle\langle e_i|$ and $\sigma = \sum_j \mu_j |f_j\rangle\langle f_j|$ be spectral decompositions of the states. The purification of ρ is $|\psi\rangle^{AR} = \sum_i \sqrt{\lambda_i}|e_i\rangle|i\rangle$. Define $|\phi\rangle^{AR} = \sum_j \sqrt{\mu_j}|f_j\rangle|j\rangle$. Any purification of the state σ is of the form $(\mathbb{1}_A \otimes V)|\phi\rangle^{AR}$ where V is an isometry acting on system R. Write the eigenbasis $\{|f_j\rangle\}$ as linear combination of eigenbasis $\{|e_j\rangle\}$, that is, $|f_j\rangle = \sum_i \alpha_{ij}|e_i\rangle$. Then, we have $|\phi\rangle^{AR} = \sum_{i,j} \sqrt{\mu_j}\alpha_{ij}|e_i\rangle|j\rangle$. Define the operator $P = \sum_{jk} p_{jk}|j\rangle\langle k|$ where $p_{jk} = \alpha_{kj}\sqrt{\frac{\mu_j}{\lambda_k}}$. It is immediate to see that

$$|\phi\rangle^{AR} = (\mathbb{1}_A \otimes P)|\psi\rangle^{AR}.$$

Thus, we have $(\mathbb{1}_A \otimes V)|\phi\rangle^{AR} = (\mathbb{1}_A \otimes VP)|\psi\rangle^{AR}$. Defining $T = VP$, we then

149

have

$$\lambda_{\max}(T^\dagger T) = \lambda_{\max}(P^\dagger P)$$

$$\leq \mathrm{Tr}\left(P^\dagger P\right)$$

$$= \sum_{j,k} |p_{jk}|^2$$

$$= \sum_{j,k} \frac{|\alpha_{kj}|^2 \mu_j}{\lambda_k}$$

$$\leq \frac{1}{\lambda_{\min}},$$

where the last inequality follows from the orthonormality of the basis $\{\,|f_j\rangle\,\}$.

∎

Lemma A.13 (Gentle Operator Lemma [113–115])**.** *If a quantum state ρ with diagonalization $\rho = \sum_j p_j \pi_j$ projects onto operator Λ with probability $1-\epsilon$, which is bounded as $0 \leq \Lambda \leq I$, i.e. $\mathrm{Tr}(\rho\Lambda) \geq 1 - \epsilon$ then*

$$\sum_j p_j \left\| \pi_j - \sqrt{\Lambda} \pi_j \sqrt{\Lambda} \right\|_1 \leq 2\sqrt{\epsilon}.$$

Definition A.1. *Let $\rho_1, \ldots, \rho_n$ be quantum states on a d-dimensional Hilbert space $\mathcal{H}$ with diagonalizations $\rho_i = \sum_j p_{ij} \pi_{ij}$ and one-dimensional projectors π_{ij}. For $\alpha > 0$ and $\rho^n = \rho_1 \otimes \cdots \otimes \rho_n$ define the set of entropy typical sequences as*

$$\mathcal{T}_{\alpha,\rho^n}^n = \left\{ j^n = j_1 j_2 \ldots j_n : \left| \sum_{i=1}^n -\log p_{ij_i} - S(\rho_i) \right| \leq \alpha\sqrt{n} \right\}.$$

Define the entropy typical projector of ρ^n with constant α as

$$\Pi_{\alpha,\rho^n}^n = \sum_{j^n \in \mathcal{T}_{\alpha,\rho^n}^n} \pi_{1j_1} \otimes \cdots \otimes \pi_{nj_n}.$$

Lemma A.14. *(Cf. [79]) There is a constant $0 < \beta \leq \max\left\{ (\log 3)^2, (\log d)^2 \right\}$ such that the entropy typical projector has the following properties for any $\alpha > 0$, $n > 0$ and arbitrary state $\rho^n = \rho_1 \otimes \cdots \otimes \rho_n$:*

$$\mathrm{Tr}\left(\rho^n \Pi_{\alpha,\rho^n}^n\right) \geq 1 - \frac{\beta}{\alpha^2},$$

$$2^{-\sum_{i=1}^n S(\rho_i) - \alpha\sqrt{n}} \Pi_{\alpha,\rho^n}^n \leq \Pi_{\alpha,\rho^n}^n \rho^n \Pi_{\alpha,\rho^n}^n \leq 2^{-\sum_{i=1}^n S(\rho_i) + \alpha\sqrt{n}} \Pi_{\alpha,\rho^n}^n, \quad and$$

$$\left(1 - \frac{\beta}{\alpha^2}\right) 2^{\sum_{i=1}^n S(\rho_i) - \alpha\sqrt{n}} \leq \mathrm{Tr}\left(\Pi_{\alpha,\rho^n}^n\right) \leq 2^{\sum_{i=1}^n S(\rho_i) + \alpha\sqrt{n}}.$$

150

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. Journal*, vol. 27, no. 4, pp. 623–656, Oct 1948.

[2] A. S. Holevo, "Bounds for the quantity of information transmitted by a quantum communication channel," *Probl. Inf. Transm.*, vol. 9, no. 3, pp. 3–11, 1973.

[3] J. P. Gordon, "Noise at optical frequencies; information theory," in *Proc. Int. School Phys. Enrico Fermi, Course XXXI*, New York, 1964, pp. 156–181.

[4] L. B. Levitin, "On quantum measure of information," in *Proc. IV All-Union Conference on Information Transmission and Coding Theory*, Tashkent, Uzbekistan, 1969, pp. 111–115.

[5] G. D. Jr. Forney, Master's Thesis, MIT, Boston, 1963 (unpublished).

[6] R. L. Stratonovich, "The transmission rate for certain quantum communications channels," *Probl. Inf. Transm.*, vol. 2, no. 1, pp. 35–44, 1966.

[7] A. S. Holevo, "On capacity of a quantum communication channel," *Probl. Inf. Transm.*, vol. 15, no. 4, pp. 247–253, 1979.

[8] B. Schumacher, "Quantum coding," *Phys. Rev. A*, vol. 51, no. 4, pp. 2738–2747, Apr 1995.

[9] P. W. Shor, "Scheme for reducing decoherence in quantum computer memory," *Phys. Rev. A*, vol. 52, pp. R2493–R2496, Oct 1995.

[10] B. Schumacher and M. A. Nielsen, "Quantum data processing and error correction," *Phys. Rev. A*, vol. 54, pp. 2629–2635, Oct 1996.

[11] B. Schumacher, "Sending entanglement through noisy quantum channels," *Phys. Rev. A*, vol. 54, pp. 2614–2628, Oct 1996.

[12] E. Knill and R. Laflamme, "Theory of quantum error-correcting codes," *Phys. Rev. A*, vol. 55, pp. 900–911, Feb 1997.

[13] H. Barnum, M. A. Nielsen, and B. W. Schumacher, "Information transmission through a noisy quantum channel," *Phys. Rev. A*, vol. 57, no. 6, pp. 4153–4175, June 1998.

[14] B. Schumacher and M. D. Westmoreland, "Quantum privacy and quantum coherence," *Phys. Rev. Lett.*, vol. 80, pp. 5695–5697, Jun 1998.

[15] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted classical capacity of noisy quantum channels," *Phys. Rev. Lett.*, vol. 83, pp. 3081–3084, Oct 1999.

[16] H.-K. Lo and S. Popescu, "Classical communication cost of entanglement manipulation: Is entanglement an interconvertible resource?" *Phys. Rev. Lett.*, vol. 83, pp. 1459–1462, Aug 1999.

[17] H. Barnum, E. Knill, and M. A. Nielsen, "On quantum fidelities and channel capacities," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1317–1329, July 2000.

[18] A. S. Holevo, "On entanglement-assisted classical capacity," *J. Math. Phys.*, vol. 43, no. 9, pp. 4326–4333, Sep. 2002.

[19] B. Schumacher and M. D. Westmoreland, "Approximate quantum error correction," *Quantum Inf. Process.*, vol. 1, no. 1-2, pp. 5–12, Apr 2002.

[20] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse shannon theorem," *IEEE Trans. Inf. Theory*, vol. 48, no. 10, pp. 2637–2655, Oct 2002.

[21] I. Devetak and A. Winter, "Distillation of secret key and entanglement from quantum states," *Proc. R. Soc. A.*, vol. 461, no. 2053, pp. 207–235, Jan. 2005.

[22] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 44–55, Jan 2005.

[23] M. Horodecki, J. Oppenheim, and A. Winter, "Quantum state merging and negative information," *Commun. Math. Phys.*, vol. 269, no. 1, pp. 107–136, Jan 2007.

[24] ——, "Partial quantum information," *Nature*, vol. 436, no. 7051, p. 673–676, Aug 2005.

[25] I. Devetak and J. T. Yard, "Exact cost of redistributing multipartite quantum states," *Phys. Rev. Lett.*, vol. 100, p. 230501, Jun 2008.

[26] J. T. Yard and I. Devetak, "Optimal Quantum Source Coding with Quantum Side Information at the Encoder and Decoder," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5339–5351, Nov 2009.

[27] M. Hayashi, "Optimal visible compression rate for mixed states is determined by entanglement of purification," *Phys. Rev. A*, vol. 73, p. 060301, June 2006.

[28] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, "The quantum reverse shannon theorem and resource tradeoffs for simulating quantum channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2926–2959, May 2014.

[29] I. Devetak and A. Winter, "Classical data compression with quantum side information," *Phys. Rev. A*, vol. 68, p. 042301, Oct 2003.

[30] A. Winter, "Coding theorems of quantum information theory," PhD Thesis, Universität Bielefeld, Department of Mathematics, Germany, July 1999, arXiv:quant-ph/9907077.

[31] M. Koashi and N. Imoto, "Compressibility of quantum mixed-state signals," *Phys. Rev. Lett.*, vol. 87, no. 1, p. 017902, July 2001.

[32] C. Ahn, A. C. Doherty, P. Hayden, and A. Winter, "On the distributed compression of quantum information," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4349–4357, Oct 2006.

[33] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, "Resource Theory of Quantum States Out of Thermal Equilibrium," *Physical Review Letters*, vol. 111, p. 250404, Dec 2013.

[34] M. Weilenmann, L. Kraemer, P. Faist, and R. Renner, "Axiomatic relation between thermodynamic and information-theoretic entropies," *Phys. Rev. Lett.*, vol. 117, p. 260601, Dec 2016.

[35] A. Winter and D. Yang, "Operational resource theory of coherence," *Phys. Rev. Lett.*, vol. 116, p. 120404, Mar 2016.

[36] E. T. Jaynes, "Information theory and statistical mechanics II," *Phys. Rev.*, vol. 108, pp. 171–190, Oct 1957.

[37] ——, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 106, pp. 620–630, May 1957.

[38] L. Brillouin, *Science and Information Theory.* Academic Press, New York, 1962.

[39] E. T. Jaynes, *Papers on Probability, Statistics and Statistical Physics.* Dordrecht, The Netherlands, 1982.

[40] F. Binder, L. Correa, C. Gogolin, J. Anders, and G. Adesso, *Thermodynamics in the Quantum Regime: Fundamental Aspects and New Directions-Fundamental Aspects and New Directions.* Springer, 2018.

[41] P. Hayden, R. Jozsa, D. Petz, and A. Winter, "Structure of states which satisfy strong subadditivity of quantum entropy with equality," *Commun. Math. Phys.*, vol. 246, no. 2, pp. 359–374, Apr 2004.

[42] M. Koashi and N. Imoto, "Operations that do not disturb partially known quantum states," *Phys. Rev. A*, vol. 66, no. 2, p. 022318, Aug 2002.

[43] D. S. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, July 1973.

[44] J. Oppenheim, "State redistribution as merging: introducing the coherent relay," *preprint (2008)*, arXiv[quant-ph]:0805.1065.

[45] Z. B. Khanian and A. Winter, "General mixed state quantum data compression with and without entanglement assistance," *preprint (2019)*, arXiv[quant-ph]:1912.08506.

[46] ——, "General mixed state quantum data compression with and without entanglement assistance," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Los Angeles, CA, USA, June 2020, pp. 1852–1857.

[47] ——, "Entanglement-assisted quantum data compression," *preprint (2019)*, arXiv[quant-ph]:1901.06346.

[48] ——, "Entanglement-assisted quantum data compression," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, July 2019, pp. 1147–1151.

[49] ——, "Distributed compression of correlated classical-quantum sources or: The price of ignorance," *IEEE Trans. Inf. Theory*, vol. 66, no. 9, pp. 5620–5633, Sep 2020, arXiv[quant-ph]:1811.09177.

[50] ——, "Distributed compression of correlated classical-quantum sources," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, July 2019, pp. 1152–1156.

[51] ——, "Rate distortion perspective of quantum state redistribution," *in preparation*, 2020.

[52] ——, "Quantum state redistribution for ensemble sources," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Los Angeles, CA, USA, June 2020, pp. 1858–1863.

[53] Z. B. Khanian, M. Nath Bera, A. Riera, M. Lewenstein, , and A. Winter, "Resource theory of heat and work and everything else," *in preparation*, 2020.

[54] M. Nath Bera, A. Riera, M. Lewenstein, Z. B. Khanian, and A. Winter, "Thermodynamics as a Consequence of Information Conservation," *Quantum*, vol. 3, p. 121, Jul. 2018, arXiv[quant-ph]:1707.01750v3.

[55] W. F. Stinespring, "Positive Functions on C^*-Algebras," *Proc. Amer. Math. Society*, vol. 6, no. 2, pp. 211–216, Apr 1955.

[56] R. Jozsa, "Fidelity for mixed quantum states," *J. Mod. Optics*, vol. 41, no. 12, pp. 2315–2323, Dec 1994.

[57] C. A. Fuchs and J. v. de Graaf, "Cryptographic distinguishability measures for quantum-mechanical states," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1216–1227, May 1999.

[58] H. Araki and E. H. Lieb, "Entropy inequalities," *Commun. Math. Phys.*, vol. 18, pp. 160–170, June 1970.

[59] E. H. Lieb and M. B. Ruskai, "Proof of the strong subadditivity of quantum-mechanical entropy," *J. Math. Phys.*, vol. 14, no. 12, p. 1938–1941, Dec 1973.

[60] R. Jozsa and B. Schumacher, "A new proof of the quantum noiseless coding theorem," *J. Mod. Optics*, vol. 41, no. 12, pp. 2343–2349, Apr 1994.

[61] H. Barnum, C. A. Fuchs, R. Jozsa, and B. Schumacher, "General fidelity limit for quantum channels," *Phys. Rev. A*, vol. 54, no. 6, pp. 4707–4711, Dec 1996.

[62] M. Horodecki, "Limits for compression of quantum information carried by ensembles of mixed states," *Phys. Rev. A*, vol. 57, no. 6, pp. 3364–3369, May 1998.

[63] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, "On quantum coding for ensembles of mixed states," *J. Phys. A: Math. Gen.*, vol. 34, no. 35, pp. 6767–6785, Aug 2001.

[64] M. Horodecki, "Optimal compression for mixed signal states," *Phys. Rev. A*, vol. 61, p. 052309, Apr 2000.

[65] H. Barnum, P. Hayden, R. Jozsa, and A. Winter, "On the reversible extraction of classical information from a quantum source," *Proc. Royal Soc. London Ser. A*, vol. 457, no. 2012, pp. 2019–2039, Aug. 2001.

[66] C. H. Bennett, P. Hayden, D. W. Leung, P. W. Shor, and A. Winter, "Remote preparation of quantum states," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 56–74, Jan 2005.

[67] P. Hayden, R. Jozsa, and A. Winter, "Trading quantum for classical resources in quantum data compression," *J. Math. Phys.*, vol. 43, no. 9, pp. 4404–4444, Sept 2002.

[68] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter, "The mother of all protocols: restructuring quantum information's family tree," *Proc. Royal Soc. London Ser. A*, vol. 465, pp. 2537–2563, May 2009.

[69] I. Savov, "Distributed compression and squashed entanglement," Master's Thesis, McGill University, Montreal, Quebec, 2007.

[70] D. Avis, P. Hayden, and I. Savov, "Distributed compression and multiparty squashed entanglement," *J. Phys. A: Math. Theor.*, vol. 41, no. 11, p. 115301, Mar 2008.

[71] C. H. Bennett and S. J. Wiesner, "Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states," *Phys. Rev. Lett.*, vol. 69, no. 20, pp. 2881–2884, Nov 1992.

[72] M. Fannes, "A continuity property of the entropy density for spin lattice systems," *Commun. Math. Phys.*, vol. 21, no. 4, pp. 291–294, Dec 1973.

[73] K. M. R. Audenaert, "A sharp continuity estimate for the von neumann entropy," *J. Phys. A: Math. Theor.*, vol. 40, no. 28, pp. 8127–8136, June 2007.

[74] R. T. Rockafeller, *Convex Analysis.* Princeton University Press, 1970.

[75] M. Ohya and D. Petz, *Quantum Entropy and Its Use.* Springer Verlag, Berlin Heidelberg, 1993 (2nd edition 2004).

[76] R. Alicki and M. Fannes, "Continuity of quantum conditional information," *J. Phys. A: Math. Gen.*, vol. 37, no. 5, pp. L55–L57, Jan 2004.

[77] A. Winter, "Tight uniform continuity bounds for quantum entropies: Conditional entropy, relative entropy distance and energy constraints," *Commun. Math. Phys.*, vol. 347, no. 1, pp. 291–313, Oct 2016.

[78] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels," *Phys. Rev. Lett.*, vol. 70, no. 13, pp. 1895–1899, Mar 1993.

[79] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Cambridge Univ. Press, 2nd ed. 2011.

[80] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," *preprint (2020)*, Apr 2020, arXiv[physics]:0004057.

[81] S. Salek, D. Cadamuro, P. Kammerlander, and K. Wiesner, "Quantum rate-distortion coding of relevant information," *IEEE Trans. Inf. Theory*, vol. 65, pp. 2603–2613, Apr 2019, arXiv[quant-ph]:1704.02903.

[82] N. Datta, C. Hirche, and A. Winter, "Convexity and operational interpretation of the quantum information bottleneck function," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, July 2019, pp. 1157–1161, arXiv[quant-ph]:1810.03644.

[83] A. Uhlmann, "The transition probability in the state space of a *-algebra," vol. 9, no. 2, pp. 273–279, Oct 1976.

[84] I. Devetak, A. W. Harrow, and A. Winter, "A family of quantum protocols," *Phys. Rev. Lett.*, vol. 93, no. 23, p. 230504, Dec 2004.

[85] T. Baumgratz, M. Cramer, and M. B. Plenio, "Quantifying coherence," *Physical Review Letters*, vol. 113, p. 140401, Sep 2014.

[86] E. Chitambar and G. Gour, "Critical examination of incoherent operations and a physically consistent resource theory of quantum coherence," *Physical Review Letters*, vol. 117, p. 030401, Jul 2016.

[87] I. Marvian and R. W. Spekkens, "How to quantify coherence: Distinguishing speakable and unspeakable notions," *Physical Review A*, vol. 94, p. 052324, Nov 2016.

[88] J. I. de Vicente and A. Streltsov, "Genuine quantum coherence," *Journal of Physics A: Mathematical and Theoretical*, vol. 50, no. 4, p. 045301, dec 2016.

[89] I. Marvian, R. W. Spekkens, and P. Zanardi, "Quantum speed limits, coherence, and asymmetry," *Physical Review A*, vol. 93, p. 052331, May 2016.

[90] A. Streltsov, G. Adesso, and M. B. Plenio, "Colloquium: Quantum coherence as a resource," *Reviews in Modern Physics*, vol. 89, p. 041003, Oct 2017.

[91] A. Streltsov, S. Rana, M. N. Bera, and M. Lewenstein, "Towards resource theory of coherence in distributed scenarios," *Physical Review X*, vol. 7, p. 011024, Mar 2017.

[92] G. Gour and A. Winter, "How to quantify a dynamical quantum resource," *Physical Review Letters*, vol. 123, p. 150401, Oct 2019.

[93] P. Contreras-Tejada, C. Palazuelos, and J. I. de Vicente, "Resource theory of entanglement with a unique multipartite maximally entangled state," *Physical Review Letters*, vol. 122, p. 120503, Mar 2019.

[94] F. Shahandeh, *Quantum Correlations*. Springer Theses, 2019.

[95] J. I. de Vicente, "On nonlocality as a resource theory and nonlocality measures," *Journal of Physics A: Mathematical and Theoretical*, vol. 47, no. 42, p. 424017, oct 2014.

[96] C. Duarte and B. Amaral, "Resource theory of contextuality for arbitrary prepare-and-measure experiments," *Journal of Mathematical Physics*, vol. 59, no. 6, p. 062202, 2018.

[97] I. Devetak, A. W. Harrow, and A. Winter, "A resource framework for quantum shannon theory," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4587–4618, Oct 2008.

[98] C. Sparaciari, J. Oppenheim, and T. Fritz, "A Resource Theory for Work and Heat," *Physical Review A*, vol. 96, p. 052112, Nov. 2017, arXiv[quant-ph]:1607.01302.

[99] Y.-K. Liu, "The Complexity of the Consistency and N-Representability Problems for Quantum States," Ph.D. dissertation, Department of Computer Science, University of California, San Diego, Dec 2007.

[100] N. Yunger Halpern, P. Faist, J. Oppenheim, and A. Winter, "Microcanonical and resource-theoretic derivations of the thermal state of a quantum system with noncommuting charges," *Nature Communications*, vol. 7, p. 12051, Jul. 2016, arXiv[quant-ph]:1512.01189.

[101] Y. Ogata, "Approximating macroscopic observables in quantum spin systems with commuting matrices," *Journal of Functional Analysis*, vol. 264, pp. 2005–2033, 2013.

[102] R. Duan, S. Severini, and A. Winter, "On Zero-Error Communication via Quantum Channels in the Presence of Noiseless Feedback," *IEEE Transactions on Information Theory*, vol. 62, pp. 5260–5277, Sep 2016.

[103] A. Dembo and O. Zeitouni, *Large Deviations: Techniques and Applications.* Springer Verlag, 1998.

[104] P. Hayden, D. W. Leung, and A. Winter, "Aspects of generic entanglement," *Communications in Mathematical Physics*, vol. 265, pp. 95–117, Jul 2006.

[105] Y. Guryanova, S. Popescu, A. J. Short, R. Silva, and P. Skrzypczyk, "Thermodynamics of quantum systems with multiple conserved quantities," *Nature Communications*, vol. 7, p. 12049, Jul. 2016, arXiv[quant-ph]:1512.01190.

[106] S. Popescu, A. B. Sainz, A. J. Short, and A. Winter, "Quantum Reference Frames and Their Applications to Thermodynamics," *Philosophical Transactions of the Royal Society A*, vol. 376, p. 20180111, 2018, arXiv[quant-phg]:1804.03730.

[107] S. Popescu, A. Sainz, A. J. Short, and A. Winter, "Reference frames which separately store non-commuting conserved quantities," 2019, arXiv[quant-ph]:1908.02713.

[108] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner, "One-Shot Decoupling," *Communications in Mathematical Physics*, vol. 328, no. 1, pp. 251–284, Mar 2014, arXiv[quant-ph]:1012.6044.

[109] M. Tomamichel, "A Framework for Non-Asymptotic Quantum Information Theory," Ph.D. dissertation, Department of Physics, ETH Zurich, 2012, arXiv[quant-ph]:1203.2142.

[110] M. Horodecki and J. Oppenheim, "Fundamental limitations for quantum and nanoscale thermodynamics," *Nature Communications*, vol. 4, p. 2059, Jun. 2013, arXiv[quant-ph]:1111.3834.

[111] F. G. S. L. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, "The second laws of quantum thermodynamics," *Proceedings of the National Academy of Sciences*, vol. 112, pp. 3275–3279, 2015.

[112] R. Bhatia, *Matrix Analysis*. New York: Springer-Verlag, 1997.

[113] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Transactions on Information Theory*, vol. 45, pp. 2481–2485, Nov 1999.

[114] T. Ogawa and H. Nagaoka, "Making Good Codes for Classical-Quantum Channel Coding via Quantum Hypothesis Testing," *IEEE Transactions on Information Theory*, vol. 53, pp. 2261–2266, June 2007.

[115] M. M. Wilde, *Quantum Information Theory*. Cambridge University Press, 2013.