The Data Compression Theorem for Ergodic Quantum Information Sources

Igor Bjelaković, Arleta Szkoła*

Technische Universität Berlin

Fakultät II - Mathematik und Naturwissenschaften

Institut für Mathematik MA 7-2

Straße des 17. Juni 136 10623 Berlin, Germany

January 21, 2003

Abstract

We extend the data compression theorem to the case of ergodic quantum information sources. Moreover, we provide an asymptotically optimal compression scheme which is based on the concept of high probability subspaces. The rate of this compression scheme is equal to the von Neumann entropy rate.

1 Introduction

In classical information theory the Shannon entropy rate h of discrete stochastic processes modelling *information sources* (IS) gives the average information carried by individual signals: Operationally it means that any ergodic information source can be compressed by means of block coding using asymptotically not more than h bits per signal in a way that there exist decompression alghorithms with asymptotically vanishing probability of error. Using an exponentially smaller number of bits the compression/decompression algorithms will fail to be asymptotically error-free.
In quantum information theory the corresponding quantity is the von Neumann entropy rate s. We show by construction (cf. Theorem 5.3 below) that for any ergodic *quantum information source* (QIS) there exists an asymptotically reliable compression scheme with rate R equal to the von Neumann entropy rate s. Here by rate we mean the asymptotic number of qubits used per signal to represent the QIS. Designing compression algorithms the goal is to achieve low rates. It turns out that s is the optimal rate in the sense that a rate $R \geq s$ is a necessary condition on asymptotical reliability of

*e-mail:{igor, szkola}@math.tu-berlin.de

compression schemes and even more the fidelity of any compression scheme with rate $R < s$ vanishes asymptotically. Of course, this result depends on the underlying fidelity notion for the compression/decompression operations on the QIS. There are different definitions of fidelity suited for different applications. In section 4 we will discuss some of them: the standard fidelity F between two quantum states (it can be seen as an extension of the overlap function of two pure states), the ensemble fidelity $\bar{F}$ and the entanglement fidelity F_e. Our result holds for $\bar{F}$ as well as for F_e .

The main tool to construct compression schemes achieving the optimal rate s are high probability subspaces. Compression maps that are essentially projections onto high probability subspaces provide a solution to the problem of the optimal data compression. A basic result concerning high probability subspaces is proved in [2]. It asserts the convergence of the minimal logarithmic dimension rate of these subspaces to the von Neumann entropy rate s in the case of ergodic QIS. This convergence was conjectured by Petz/Hiai in [8].

The concept of high probability subspaces is crucial in the work of Petz/Mosonyi [14], where they prove a coding theorem for the class of completely ergodic QIS. Using projections onto high probability subspaces they show that completely ergodic QIS can be compressed with any rate $R \geq s$ in such a way that the ensemble fidelity $\bar{F}$ is asymptotically equal to 1. On the other hand $\bar{F}$ cannot achieve 1 asymptotically if the rate satisfies $R < s$. The reason why they cannot conclude that in fact for $R < s$ the asymptotical fidelity $\bar{F}$ is equal to 0 is that they use the result of Hiai/Petz [8] which provides bounds on limit superior and limit inferior and not the limit of the minimal logarithmic dimension rate of the high probabilty subspaces. The result of Petz/Mosonyi represents an extension of the coding theorem formulated in [11] by Jozsa and Schumacher for the smaller class of independent identically distributed (i.i.d.) QIS and proved in [11] and [1]. An analogous result for i.i.d. QIS using the entanglement fidelity F_e as a criterion for the reliability of compression schemes is presented by Nielsen and Chuang in [13].

In [3] Datta/Suhov treat the case of certain weakly non-stationary quantum spin systems. They show that, under the condition of asymptotical reliability measured by the ensemble fidelity $\bar{F}$, the optimal rate for compression of information carried by Gibbs states of the considered interacting quantum spin systems is given by the von Neumann entropy rate.

2 Quantum Information Sources

Classical discrete IS are stochastic processes, i.e. sequences of random variables $\{X_i\}_{i\in\mathbb{Z}}$ with a joint distribution P, each random variable X_i taking values from a set A called alphabet. We will consider only the case of IS over finite alphabets. A possible realisation of an (discrete) IS over an alphabet A is a physical process producing at discrete times physical systems with identical state spaces equal to A and the individual states of the systems being random variables taking values from A according to a probability rule P determined by the given IS. Alternatively a classical IS can be viewed as a classical spin chain possibly coupled to an external enviroment. An IS is the first stage in the process of information transmission or storage, it provides information which is sent via a classical channel.

Equivalent to the stochastic process model is the Kolmogorov representation of IS (cf. [16]). It describes an IS as a dynamical system $(A^\infty, \mathfrak{A}^\infty, \nu, T)$ on a doubly infinite product space $A^\infty := \ldots \times A \times A \times \ldots$, where $\mathfrak{A}^\infty$ is the σ-field generated by cylinder sets, T is the shift on A^∞ and ν a probability measure on $(A^\infty, \mathfrak{A}^\infty)$ uniquely determined by the probability distribution P of the stochastic process.

A discrete QIS can be viewed as a quantum spin chain possibly coupled to an external environment where one focuses on the information carried by the quantum state. Alternatively, we can think of a QIS as a device that sends quantum physical systems of a fixed type, prepared in a joint generally mixed and entangled state. In both cases, a QIS provides input for quantum channels.

Before we present a mathematical model for QIS, which corresponds to the Kolmogorov representation of classical IS, we introduce quasilocal C^*-algebras as the non-commutative counterpart of the doubly infinite product space A^∞. This standard mathematical formalism is introduced in detail e.g. in [15]. One starts with the group $\mathbb{Z}$. To each $z \in \mathbb{Z}$ there is associated a C^*-algebra $\mathcal{A}_z$. Each $\mathcal{A}_z$ is isomorphic to a fixed finite dimensional unital C^*-algebra $\mathcal{A}$, which is in general non-commutative. The C^*-algebra $\mathcal{A}$ corresponds to an algebra of observables of a quantum system and the isomorphism between the $\mathcal{A}_z$ reflects the assumption that the source is emitting quantum systems of a fixed type.

In this paper we will be mainly concerned with the case $\mathcal{A} = \mathcal{B}(\mathcal{H})$, the linear operators on the finite dimensional Hilbert space $\mathcal{H}$. For a finite subset $\Lambda \subset \mathbb{Z}$ the algebra $\mathcal{A}_\Lambda$ is given by $\mathcal{A}_\Lambda := \bigotimes_{z\in\Lambda} \mathcal{A}_z$. The quasilocal C^*-algebra $\mathcal{A}^\infty$ is defined as the operator norm closure of the local *-algebra $\mathcal{A}_{\mathrm{loc}} := \bigcup_{\Lambda\subset\mathbb{Z}} \mathcal{A}_\Lambda$.

A state on the quasilocal algebra is given by a normed positive functional Ψ, i.e. $\Psi(\mathbf{1}) = 1$ and $\Psi(A) \geq 0$ for all $A \in \mathcal{A}^\infty$ with $A \geq 0$. There is one-to-one correspondence between the state Ψ and a consistent family of states $\{\Psi^{(\Lambda)}\}_{\Lambda\subset\mathbb{Z}}$, where each $\Psi^{(\Lambda)}$ is the restriction of Ψ to the finite dimensional subalgebra $\mathcal{A}_\Lambda$ of $\mathcal{A}^\infty$ and consistency means that $\Psi^{(\Lambda)} = \Psi^{(\Lambda')} \upharpoonright \mathcal{A}_\Lambda$ for $\Lambda \subset \Lambda'$. For most purposes it suffices to deal only with integer intervals $\Lambda = \{z_1, z_1 + 1, \ldots, z_2\}$ with integers $z_1 \leq z_2$. For each $\Psi^{(\Lambda)}$ there exists a unique density operator $\rho^{(\Lambda)} \in \mathcal{A}_\Lambda$, such that $\Psi^{(\Lambda)}(a) = \mathrm{tr}_\Lambda \rho^{(\Lambda)} a,\ a \in \mathcal{A}_\Lambda$

and tr_Λ is the trace on $\mathcal{A}_\Lambda$.
On $\mathcal{A}_{\mathrm{loc}}$ we define the shift T that acts in in the following way: For integers $z_1 \le z_2$ and $\Lambda = \{z_1, z_1+1, \ldots, z_2\}$ we have

$$T : \mathcal{A}_\Lambda \to \mathcal{A}_{\Lambda+1},\ a \simeq a \otimes \mathbf{1} \mapsto T(a) = \mathbf{1} \otimes a \simeq a.$$

The canonical extension of T onto $\mathcal{A}^\infty$ is an $*$-automorphism on $\mathcal{A}^\infty$ and the integer powers of T, $\{T^z\}_{z\in\mathbb{Z}}$, represent an action of the translation group $\mathbb{Z}$ by automorphisms on $\mathcal{A}^\infty$. The triple $(\mathcal{A}^\infty, \Psi, T)$ defines a quantum dynamical system.

The mathematical model for a discrete QIS is a quantum dynamical system $(\mathcal{A}^\infty, \Psi, T)$, where $\mathcal{A}^\infty$ is a quasilocal C^*-algebra over $\mathbb{Z}$ constructed from a finite dimensional C^*-algebra $\mathcal{A}$, Ψ is a state and T the shift on $\mathcal{A}^\infty$.

Remark: *If $\mathcal{A}$ is an* abelian *finite dimensional C^*-algebra then by the Gelfand isomorphism $\mathcal{A}$ can be identified with $C(A)$, the algebra of functions on a set A with $|A| = \dim \mathcal{A}$ and $\mathcal{A}^\infty$ is *-isomorphic to $C(A^\infty)$. Further by the Riesz representation theorem there exists a probability measure ν on $(A^\infty, \mathfrak{A}^\infty)$ uniquely determined by $\Psi(a) = \sum_{\mathbf{i}\in A^\Lambda} a(\mathbf{i})\nu^{(\Lambda)}(\mathbf{i})$, for all $a \in \mathcal{A}_\Lambda$ and arbitrary $\Lambda \subset \mathbb{Z}$, where $a(\cdot) \in C(A^\Lambda)$ is the Gelfand representation of $a \in \mathcal{A}_\Lambda$. Then the IS given by $(\mathcal{A}^\infty, \Psi, T)$ has the Kolmogorov representation $(A^\infty, \mathfrak{A}^\infty, \nu, T)$. Hence, this construction leads back to the classical IS.*

For simplicity in the following sections we will restrict to the case $\mathcal{A} = \mathcal{B}(\mathcal{H})$.

$(\mathcal{A}^\infty, \Psi, T)$ is a *stationary* QIS if for all $a \in \mathcal{A}^\infty$:

$$\Psi(Ta) = \Psi(a). \tag{1}$$

As we will deal only with stationary QIS, we will assume without loss of generality that all integer intervals are of the form $\Lambda = \{1, \ldots, n\}$ with $n \ge 1$. Furthermore we write $\rho^{(n)}$ instead of $\rho^{(\Lambda)}$ and $\Psi^{(n)}$ instead of $\Psi^{(\Lambda)}$.
A stationary QIS $(\mathcal{A}^\infty, \Psi, T)$ is *ergodic* if

$$\lim_{n\to\infty} \Psi\left((\frac{1}{n}\sum_{i=0}^{n-1} T^i(a))^2 \right) = \Psi(a)^2$$

holds for all self-adjoint $a \in \mathcal{A}^\infty$. It is important to realize that for a QIS as a dynamical system on a quasilocal algebra this definition of ergodicity is equivalent to the definition used in [2] (cf. Proposition 6.3.5 in [15]), where $(\mathcal{A}^\infty, \Psi, T)$ is said to be ergodic if Ψ is an extremal point in the compact convex set of stationary states on $\mathcal{A}^\infty$, (cf. [10]). Thus the results for ergodic quantum dynamical systems presented in [2] hold for the ergodic QIS defined above.
Finally in this section we introduce the entropy rate $s(\Psi)$ of a stationary

QIS $(\mathcal{A}^\infty, \Psi, T)$, which is the crucial quantity in the present paper. Recall the one-to-one correspondence of a stationary Ψ on $\mathcal{A}^\infty$ and the family of density operators $\{\rho^{(n)}\}n \in \mathbb{N}$. The *entropy rate* $s(\Psi)$ is then defined by

$$s(\Psi) := \lim_{n\to\infty} \frac{1}{n} S(\rho^{(n)}), \tag{2}$$

where $S(\rho^{(n)}) := -\mathrm{tr}\rho^{(n)} \log_2 \rho^{(n)}$ is the von Neumann entropy of the density operator $\rho^{(n)}$.

3 Data Compression Schemes

In order to define lossless data compression schemes for encoding quantum signals we need the concept of *trace preserving quantum operations.* A physical approach to trace preserving quantum operations can be obtained as follows. Consider a quantum system S prepared in some state ρ acting on the Hilbert space $\mathcal{H}$. We imagine that this system interacts with its enviroment, a quantum system S_{env} in a state ρ_{env} on the finite dimensional Hilbert space $\mathcal{H}_{\mathrm{env}}$. The system $S \times S_{\mathrm{env}}$ is closed and we make the assumption that it is initially in the product state $\rho \otimes \rho_{\mathrm{env}}$ on $\mathcal{H} \otimes \mathcal{H}_{\mathrm{env}}$. As a state of a closed system it undergoes a unitary evolution represented by a unitary operator U on $\mathcal{H} \otimes \mathcal{H}_{\mathrm{env}}$. The corresponding evolution of the state ρ of S is usually not unitary, i.e. irreversible. It is given by a trace preserving quantum operation $\mathcal{E}$:

$$\mathcal{E}(\rho) := \mathrm{tr}_{\mathcal{H}_{\mathrm{env}}}(U(\rho \otimes \rho_{\mathrm{env}})U^*). \tag{3}$$

It can be shown that each trace preserving quantum operation $\mathcal{E}$ posseses the following representation known as Kraus or sum representation (cf. [6], [7], [13])

$$\mathcal{E}(\rho) = \sum_i E_i \rho E_i^*,$$

where $E_i \in \mathcal{B}(\mathcal{H})$ and $\sum_i E_i^* E_i = \mathbf{1}$. This description contains, for example, the cases of the unitary time evolution and general measurements. We remark that trace preserving quantum operations may be described in a more elegant way within the framework of completely positive linear maps between C^*-algebras (cf. [4], [12]).

A *compression scheme* $(\mathcal{C}, \mathcal{D})$ for stationary QIS is a sequence $\{(\mathcal{C}^{(n)}, \mathcal{D}^{(n)})\}_{n\in\mathbb{N}}$ of pairs of trace preserving quantum operations

$$\mathcal{C}^{(n)} : \mathcal{S}(\mathcal{H}^{\otimes n}) \longrightarrow \mathcal{S}(\mathcal{H}^{(n)}), \tag{4}$$

$$\mathcal{D}^{(n)} : \mathcal{S}(\mathcal{H}^{(n)}) \longrightarrow \mathcal{S}(\mathcal{H}^{\otimes n})$$

where $\mathcal{H}^{(n)} \subseteq \mathcal{H}^{\otimes n}$ for all $n \in \mathbb{N}$ and $\mathcal{S}(\mathcal{H}^{\otimes n})$, $\mathcal{S}(\mathcal{H}^{(n)})$ denote the sets of density operators on $\mathcal{H}^{\otimes n}$ resp. $\mathcal{H}^{(n)}$. We refer to $\mathcal{C}^{(n)}$, $\mathcal{D}^{(n)}$ as compression resp. decompression map.
The *rate* $R(\mathcal{C})$ of a compression scheme $(\mathcal{C}, \mathcal{D})$ is defined by

$$R(\mathcal{C}) := \limsup_{n\to\infty} \frac{\log_2 \dim \mathcal{H}^{(n)}}{n}. \tag{5}$$

4 Fidelities

In this section we review the basic notions and properties of the fidelity and derived quantities needed to measure the distance between two quantum states. The fidelity F between two density operators ρ and σ acting on some finite dimensional Hilbert space $\mathcal{H}$ is defined by

$$F(\rho, \sigma) := \mathrm{tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}. \tag{6}$$

The fidelity is symmetric in its entries and takes values between 0 and 1 with $F(\rho, \sigma) = 0$ iff ρ and σ are supported on orthogonal subspaces. $F(\rho, \sigma) = 1$ appears only in the case $\rho = \sigma$. In light of these properties it is reasonable to interpret the fidelity as a measure of distinguishability of two density operators which reduces to the well known overlap $|\langle\psi|\phi\rangle|$ in the case of pure states $|\psi\rangle\langle\psi|$ and $|\phi\rangle\langle\phi|$ on $\mathcal{H}$. Moreover F is jointly concave and increasing under trace preserving quantum operations. The proofs of these facts may be found in [13]. The fidelity F is equivalent to the familiar trace distance of two density operators in the following sense:

$$1 - F(\rho, \sigma) \le \frac{1}{2}\mathrm{tr}|\rho - \sigma| \le \sqrt{1 - (F(\rho, \sigma))^2} \quad \text{(cf. [13])}. \tag{7}$$

But the trace distance can be represented as (cf. [13])

$$\frac{1}{2}\mathrm{tr}|\rho - \sigma| = \max\{\mathrm{tr}(P(\rho - \sigma)) : P = P^* = P^2\}.$$

This equality has the following meaning: the orthogonal projections appearing in the above equation are usualy interpreted as ideal "yes-no" measurements. The outcome "yes" (resp. "no") is represented by P (resp. $\mathbf{1} - P$). The trace distance quantifies the largest difference of probabilities for obtaining outcome "yes" if we perform measurements on quantum systems in the states ρ and σ. This relation between the fidelity and the trace distance gives us an idea about the operational interpretation of the fidelity.

The question how well is the state of the open quantum system preserved by a time evolution, a measurement or more generally by an arbitrary quantum processes defined by a Kraus representation leads to several fidelity concepts. The first one is the *entanglement fidelity* F_e which is a function of a density operator ρ and a quantum operation $\mathcal{E}$. It is defined by

$$F_e(\rho, \mathcal{E}) := (F(|\Psi\rangle\langle\Psi|, (\mathbf{1} \otimes \mathcal{E})(|\Psi\rangle\langle\Psi|)))^2 , \tag{8}$$

where $|\Psi\rangle \in \mathcal{H}^{'} \otimes \mathcal{H}$ is an arbitrary purification of ρ, i.e. $\mathrm{tr}_{\mathcal{H}'}|\Psi\rangle\langle\Psi| = \rho$. It can be shown that this definition does not depend on the particular choice of the purification of ρ, cf. [13].
Let $\mathcal{E}(\rho) = \sum_i E_i \rho E_i^*$ be the sum representation of $\mathcal{E}$ for all density operators ρ on $\mathcal{H}$, i.e. $E_i \in \mathcal{B}(\mathcal{H})$ and $\sum_i E_i^* E_i = \mathbf{1}$. Then it holds

$$F_e(\rho, \mathcal{E}) = \sum_i |\mathrm{tr}\rho E_i|^2. \tag{9}$$

This formula implies that the entanglement fidelity is a convex function of the density operator. Indeed, the last expression is merely the squared norm of a complex vector with the components $\mathrm{tr}(\rho E_i)$, which depend affinely on ρ. Moreover, every norm is a convex function, so we obtain the claimed convexity of F_e. The intuition behind the definition (8) is that what we want to preserve is the purifications of a given state. If the state is mixed then all purifications are entangled pure states.

In order to define the *ensemble fidelity* $\bar{F}$ we start with a finite set of symbols $\{1, \ldots, n\}$ (a classical alphabet) which are drawn according to a probability distribution $(p_1, \ldots, p_n)$. We associate to this set of symbols a fixed set of density operators $\{\rho_1, \ldots \rho_n\}$ on $\mathcal{H}$ and define the ensemble fidelity by

$$\bar{F}(\{(p_i, \rho_i)\}_{i=1}^n, \mathcal{E}) := \sum_{i=1}^n p_i (F(\rho_i, \mathcal{E}(\rho_i)))^2, \tag{10}$$

where $\mathcal{E}$ is a quantum operation. The weighted ensemble of n quantum states $\{(p_i, \rho_i)\}_{i=1}^n$ represents a convex decomposition of the density operator $\rho = \sum_{i=1}^n p_i \rho_i$. If the ρ_i are all pure states then we call the ensemble or the convex decomposition a pure one. We will denote by $F_s(\rho, \mathcal{E})$ the supremum over pure convex decompositions of the ensemble fidelities for a density operator ρ and a quantum operation $\mathcal{E}$:

$$F_s(\rho, \mathcal{E}) := \sup\{\bar{F}(\{(p_i, P_i)\}_{i=1}^n, \mathcal{E}) \; : \; \{(p_i, P_i)\}_{i=1}^n \text{ pure convex decomposition of } \rho\}. \tag{11}$$

The idea behind the definition (10) is that the classical alphabet is represented by quantum systems prepared in the states from some fixed set. For example we can encode the alphabet $\{0, 1\}$ into two different polarization directions of photons. The probability of occurence of each polarisation direction is determined by the probability distribution on the classical alphabet. The ensemble fidelity $\bar{F}$ appears mainly in problems concerning classical information to be e.g. stored on or transmitted via quantum states.

We conclude this section with a basic relation among several notions of fidelity introduced here. For a fixed density operator ρ we define

$$\bar{\mathrm{F}}_{\rho,\mathcal{E}} := \{\bar{F}(\{(p_i,\rho_i)\}_i,\mathcal{E})| \sum_i p_i\rho_i = \rho\}.$$

It holds

$$0 \leq F_e(\rho,\mathcal{E}) \leq \bar{F} \leq F(\rho,\mathcal{E}(\rho)) \leq 1, \quad \bar{F} \in \bar{\mathrm{F}}_{\rho,\mathcal{E}}. \tag{12}$$

The second inequality follows immediately from the convexity of the entanglement fidelity. The third inequality holds because the fidelity F is jointly concave. Observe that according to (12) we can give upper and lower bounds for $\bar{F}$ which depend exclusively on the density operator ρ corresponding to the convex decomposition in consideration. The inequality (12) will play a crucial role in our derivation of data compression theorem.

5 Data Compression Theorem

One of the interests in the quantum information theory is an economical and errorfree storage or transmission of quantum information. In other words the question is: what is the minimal amount of resources measured in units of qubits or equivalently in Hilbert space dimensions needed to store quantum states faithfully? This question has been resolved in the case of memoryless sources using the entanglement fidelity F_e as a criterion for reliability, [13]: Each compression scheme possesing a rate smaller than the von Neumann entropy rate cannot be reliable in the sense that the entanglement fidelity tends to 0. It has been shown in [1] that for encodings of classical memoryless sources into some fixed set of pure quantum states, as described in the previous section, an analogous assertion holds. In this case the reliability is measured by the ensemble fidelity $\bar{F}$. In both cases compression schemes have been constructed with rates, that can be made arbitrary close to the von Neumann entropy $S(\rho)$. An essential ingredient was the quantum asymptotic equipartition property (AEP) for memoryless QIS. An extension of the quantum AEP to the more general case of ergodic QIS was formulated and proved in [2].

Theorem 5.1 (Quantum AEP Theorem) *Let $(\mathcal{A}^\infty,\Psi,T)$ be an ergodic quantum information source with the entropy rate $s(\Psi)$ defined by eqn. (2). Then for any $\varepsilon > 0$ there exists an $N_\varepsilon \in \mathbb{N}$ such that for all $n \geq N_\varepsilon$ there exists a subspace $\mathcal{T}_\varepsilon^{(n)} \subseteq \mathcal{H}^{\otimes n}$ such that*

1) $\mathrm{tr}(\rho^{(n)} P_{\mathcal{T}_\varepsilon^{(n)}}) \geq 1-\varepsilon$, *where $P_{\mathcal{T}_\varepsilon^{(n)}}$ is the projector onto the subspace $\mathcal{T}_\varepsilon^{(n)}$,*
2) $2^{n(s(\Psi)-\varepsilon)} \leq \mathrm{tr}(P_{\mathcal{T}_\varepsilon^{(n)}}) \leq 2^{n(s(\Psi)+\varepsilon)}$.

Moreover, these subspaces can be chosen as

$$\mathcal{T}_\varepsilon^{(n)} := \quad \mathrm{span}\{e_i^{(n)} \in \mathcal{H}^{\otimes n} | \quad \mathrm{tr}(\rho^{(n)} P_{e_i^{(n)}}) \in [2^{-n(s(\Psi)+\varepsilon)}, 2^{-n(s(\Psi)-\varepsilon)}], \quad e_i^{(n)} \textit{ eigenvector of } \rho^{(n)}\}. \tag{13}$$

Remark: *The above theorem represents a simplified form of the Quantum Shannon-McMillan Theorem presented in [2] for the more general case of higher dimensional quantum dynamical lattice systems. Furthermore it is important to notice that the version presented in [2] also includes the case of $\mathcal{A}^\infty$ constructed from a subalgebra $\mathcal{A} \subseteq \mathcal{B}(\mathcal{H})$. If the subalgebra $\mathcal{A}$ is commutative then the Quantum Shannon-McMillan Theorem coincides with the classical theorem.*

Each of the $\mathcal{T}_\varepsilon^{(n)} \subseteq \mathcal{H}^{\otimes n}$ defined by (13) represents a subspace of probability close to 1 for large n in the case of an ergodic source $(\mathcal{A}^\infty, \Psi, T)$. In analogy to the classical theory we will call such a space the ε-typical subspace of $\mathcal{H}^{\otimes n}$ with respect to $\rho^{(n)}$.

The next proposition is strongly related to the Quantum AEP Theorem. In [2] it is proven for the higher dimensional case. The proposition is crucial for the proof of the second and third part of the Data Compression Theorem presented below. These parts say that an asymptotically reliable compression to the von Neumann entropy rate is achievable and is an optimal one. The relevant quantity for compression is the minimal logarithmic dimension of subspaces of $\mathcal{H}^{\otimes n}$ depending on the minimal required probability of the subspaces.

$$\beta_{\varepsilon,n}(\Psi) := \min\{\log_2(\mathrm{tr} q) | \ q \in \mathcal{B}(\mathcal{H}^{\otimes n}) \text{ projector}, \ \mathrm{tr}\rho^{(n)} q \geq 1 - \varepsilon\}, \ \varepsilon \in (0, 1).$$

We refer to subspaces $\mathcal{P}_\varepsilon^{(n)} \subset \mathcal{H}^{\otimes n}$ with $\mathrm{tr}\rho^{(n)} P_{\mathcal{P}_\varepsilon^{(n)}} \geq 1 - \varepsilon$ and $\log(\mathrm{tr} P_{\mathcal{P}_\varepsilon^{(n)}}) = \beta_{\varepsilon,n}(\Psi)$ as high probability subspaces (with resp. to $\rho^{(n)}$) corresponding to the level ε, cf. [14].
It turns out that the asymptotic rate of $\beta_{\varepsilon,n}(\Psi)$ does not depend on ε in the case of an ergodic state Ψ and is equal to the entropy rate $s(\Psi)$.

Proposition 5.2 *Let $(\mathcal{A}^\infty, \Psi, T)$ be an ergodic quantum information source with the entropy rate $s(\Psi)$. Then for every $\varepsilon \in (0, 1)$*

$$\lim_{n\to\infty} \frac{1}{n} \beta_{\varepsilon,n}(\Psi) = s(\Psi). \tag{14}$$

Now, disposing of the above proposition we can extend results concerning compressibility of information to the case of correlated (ergodic) QIS.

Theorem 5.3 (Data Compression Theorem) *Let $(\mathcal{A}^\infty, \Psi, T)$ be an ergodic quantum information source with the entropy rate $s(\Psi)$.*

1) Each compression scheme $(\mathcal{C}, \mathcal{D})$ satisfying

$$\lim_{n\to\infty} \bar{F}(\{(\lambda_i^{(n)}, P_i^{(n)})\}_{i=1}^{k_n}, \mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}) = 1 \tag{15}$$

for some sequence $\{\{(\lambda_i^{(n)}, P_i^{(n)})\}_{i=1}^{k_n}\}_{n\in\mathbb{N}}$ of pure convex decompositions of $\rho^{(n)}$, respectively, fulfils

$$R(\mathcal{C}) \geq s(\Psi).$$

2) There exists a compression scheme $(\mathcal{C}, \mathcal{D})$ with $R(\mathcal{C}) = s(\Psi)$ such that

$$\lim_{n\to\infty} F_e(\rho^{(n)}, \mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}) = 1.$$

3) Any compression scheme $(\mathcal{C}, \mathcal{D})$ with $R(\mathcal{C}) < s(\Psi)$ satisfies

$$\lim_{n\to\infty} F_s(\rho^{(n)}, \mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}) = 0, \tag{16}$$

where F_s is defined by (11).

Taking into account the relation (12) we use different notions of fidelity in the seperate parts of the above theorem. In this way we obtain that the von Neumann entropy rate is the optimal compression rate using the fidelity $\bar{F}$ as well as F_e.
It should be helpful to sketch the ideas which lead to the proof of the theorem above. The first item in the theorem is essentially a consequence of the monotonicity of the relative entropy (cf. [17]) and the Fannes inequality (cf. [5]) modulo some elementary estimates. The second item is derived from the fact stated in the Proposition 5.2 saying that the asymptotic rate of $\beta_{\varepsilon,n}$ is given by the von Neumann entropy rate and does not depend on the level ε. Compression schemes $(\mathcal{C}, \mathcal{D})$ consisting of compression maps which are essentially projections onto the high probability subspaces and the canonical embeddings as decompression maps posses a rate equal to the von Neumann entropy rate. So, if we combine appropriately high probability subspaces such that their corresponding levels tend to 0, we can achieve that the entanglement fidelity becomes arbitrary close to 1. This strategy leads directly to the proof of the second part of the theorem. Finally, the third item in the above theorem can be proved using the fact that F_s is bounded from above by the maximal expectation value of projectors $P \in \mathcal{H}^{\otimes n}$ satisfying the dimension condition $\mathrm{tr}P = \dim \mathcal{H}^{(n)}$. But if the rate of a data compression scheme is asymptotically smaller than the von Neumann entropy rate then according to the Proposition 5.2 the expectation values of projectors providing the upper bounds for F_s must vanish asymptotically.
Proof of Theorem 5.3: *Proof of 1)* Fix a convex decomposition of $\rho^{(n)}$ into

one dimensional projectors $\{P_i^{(n)}\}_{i=1}^{k_n}$ corresponding to the set of weights $\{\lambda_i^{(n)}\}_{i=1}^{k_n}$. Following an idea of M. Horodecki in [9] we arrive at the following elementary inequalities using the relative entropy and its decreasing behaviour with respect to the trace preserving operations (cf. [17]):

$$\begin{aligned}
\log_2 \dim \mathcal{H}^{(n)} &\geq S(\mathcal{C}^{(n)}(\rho^{(n)})) \\
&\geq S(\mathcal{C}^{(n)}(\rho^{(n)})) - \sum_{i=1}^{k_n} \lambda_i^{(n)} S(\mathcal{C}^{(n)}(P_i^{(n)})) \\
&= \sum_{i=1}^{k_n} \lambda_i^{(n)} S(\mathcal{C}^{(n)}(P_i^{(n)}), \mathcal{C}^{(n)}(\rho^{(n)})) \\
&\geq \sum_{i=1}^{k_n} \lambda_i^{(n)} S(\mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(P_i^{(n)}), \mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(\rho^{(n)})) \\
&= S(\mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(\rho^{(n)})) - \sum_{i=1}^{k_n} \lambda_i^{(n)} S(\mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(P_i^{(n)}))
\end{aligned}$$

In the next step we will show that

$$\lim_{n\to\infty} \frac{1}{n} S(\mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(\rho^{(n)})) = s(\Psi), \tag{17}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{k_n} \lambda_i^{(n)} S(\mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(P_i^{(n)})) = 0, \tag{18}$$

holds, which implies the first part of the theorem. By (7) and the Fannes inequality (cf. [5]) we have

$$\frac{1}{n}|S(\rho^{(n)}) - S(\mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(\rho^{(n)}))| \leq 2\log_2 d \sqrt{1 - (F(\rho^{(n)}, \mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(\rho^{(n)})))^2} + \frac{1}{n}.$$

Employing the limit assertion (15) and joint concavity of the fidelity we obtain (17).
Fix $\varepsilon \in (0,1)$. We consider the set

$$A_\varepsilon^{(n)} := \{i \in \{1, \ldots, k_n\} |\ (F(P_i^{(n)}, \mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(P_i^{(n)})))^2 < 1 - \varepsilon\}$$

and estimate

$$\sum_{i=1}^{k_n} \lambda_i^{(n)} F^2(P_i^{(n)}, \mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(P_i^{(n)})) \leq (1-\varepsilon) \sum_{i \in A_\varepsilon^{(n)}} \lambda_i^{(n)} + \sum_{i \in A_\varepsilon^{(n)c}} \lambda_i^{(n)}, \tag{19}$$

where $A_\varepsilon^{(n)c}$ denotes the complement of $A_\varepsilon^{(n)}$. We claim that for all $\varepsilon \in (0,1)$

$$\lim_{n\to\infty} \sum_{i \in A_\varepsilon^{(n)}} \lambda_i^{(n)} = 0. \tag{20}$$

In fact, suppose that for some $\varepsilon \in (0,1)$

$$\limsup_{n\to\infty} \sum_{i\in A_\varepsilon^{(n)}} \lambda_i^{(n)} = a > 0.$$

Then there would exist a subsequence, which we denote again by $\{A_\varepsilon^{(n)}\}_{n\in\mathbb{N}}$ for simplicity, with

$$\lim_{n\to\infty} \sum_{i\in A_\varepsilon^{(n)}} \lambda_i^{(n)} = a.$$

After taking limits in (19) this would imply the following contradictory inequality

$$1 \leq (1-\varepsilon)a + (1-a).$$

By (20), it suffices to show that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i\in A_\varepsilon^{(n)c}} \lambda_i^{(n)} S(\mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(P_i^{(n)})) = 0.$$

For small $\varepsilon \in (0,1)$ and for n large enough we have

$$\begin{aligned}\frac{1}{n} \sum_{i\in A_\varepsilon^{(n)c}} \lambda_i^{(n)} S(\mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(P_i^{(n)})) &\leq \frac{1}{n} \sum_{i\in A_\varepsilon^{(n)c}} \lambda_i^{(n)} (2n\log_2(d)\sqrt{\varepsilon}+1) \\ &\leq 2\log_2(d)\sqrt{\varepsilon} + \frac{1}{n},\end{aligned}$$

where in the first inequality we have applied Fannes inequality to the expressions $S(\mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(P_i^{(n)})) = |S(P_i^{(n)}) - S(\mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(P_i^{(n)}))|$, respectively. Since ε can be made arbitrarily small, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i\in A_\varepsilon^{(n)c}} \lambda_i^{(n)} S(\mathcal{D}^{(n)} \circ \mathcal{C}^{(n)}(P_i^{(n)})) = 0.$$

Proof of 2) By proposition 5.2 we have

$$R(\mathcal{C}) = \lim_{n\to\infty} \frac{\log_2 \dim \mathcal{P}_\varepsilon^{(n)}}{n} = \lim_{n\to\infty} \frac{1}{n}\beta_{\varepsilon,n} = s(\Psi)$$

for each $\varepsilon \in (0,1)$. A simple argument shows that there exists a sequence $\varepsilon_n \searrow 0$, for $n \to \infty$, such that

$$\lim_{n\to\infty} \frac{1}{n}\beta_{\varepsilon_n,n} = s(\Psi).$$

We consider the compression scheme $(\mathcal{C}, \mathcal{D})$, where for each $n \in \mathbb{N}$ the compression map $\mathcal{C}^{(n)}$ is given by

$$\mathcal{C}^{(n)}(\rho^{(n)}) = P_{\mathcal{P}_{\varepsilon_n}^{(n)}} \rho^{(n)} P_{\mathcal{P}_{\varepsilon_n}^{(n)}} + \sum_{e\in S^{(n)}} |0\rangle\langle e|\rho^{(n)}|e\rangle\langle 0|,$$

where $\mathcal{P}^{(n)}_{\varepsilon_n}$ is a high probability subspace of $\mathcal{H}^{\otimes n}$ corresponding to the level ε_n, $|0\rangle \in \mathcal{P}^{(n)}_{\varepsilon_n}$ and $S^{(n)}$ is an orthonormal system in $(\mathcal{P}^{(n)}_{\varepsilon_n})^{\perp}$. The decompression map $\mathcal{D}^{(n)}$ is just the canonical embedding of $\mathcal{S}(\mathcal{H}^{(n)})$ into $\mathcal{S}(\mathcal{H}^{\otimes n})$. Using the formula (9) for F_e we obtain

$$F_e(\rho^{(n)}, \mathcal{C}^{(n)}) = |\mathrm{tr}\rho^{(n)} P_{\mathcal{P}^{(n)}_{\varepsilon_n}}|^2 + \sum_{e\in S^{(n)}} |\mathrm{tr}\rho^{(n)}|0\rangle\langle e||^2 \geq |\mathrm{tr}\rho^{(n)} P_{\mathcal{P}^{(n)}_{\varepsilon_n}}|^2.$$

By definition of high probability spaces $\mathrm{tr}\rho^{(n)} P_{\mathcal{P}^{(n)}_{\varepsilon_n}} \geq 1 - \varepsilon_n$ for all $n \in \mathbb{N}$. Thus

$$|\mathrm{tr}\rho^{(n)} P_{\mathcal{P}^{(n)}_{\varepsilon_n}}|^2 \geq (1-\varepsilon_n)^2 \geq 1 - 2\varepsilon_n.$$

Recall that $\varepsilon_n \searrow 0$ and thus assertion *2)* follows.
Proof of 3) Let us define for a density operator ρ on $\mathcal{H}$ and some integer $d \leq \dim \mathcal{H}$

$$\eta_d(\rho) := \max\{\mathrm{tr}\rho P|\ P \text{ projector on } \mathcal{H}, \mathrm{tr}P = d\}.$$

As was proven in [1], for any compression scheme $(\mathcal{C}, \mathcal{D})$ we have

$$F_s(\rho^{(n)}, \mathcal{C}^{(n)} \circ \mathcal{D}^{(n)}) < 6 \cdot \eta_{d^{(n)}}(\rho^{(n)}), \qquad \forall n \in \mathbb{N},$$

where $d^{(n)} := \dim \mathcal{H}^{(n)}$. Let $\limsup_{n\to\infty} \frac{1}{n}\log_2 d^{(n)} = R(\mathcal{C}) < s(\Psi)$. Then $\lim_{n\to\infty} \eta_{d^{(n)}}(\rho^{(n)}) = 0$. Otherwise there would exist a sequence $\{P^{(n)}\}_{n\in\mathbb{N}}$ of projectors in $\mathcal{H}^{\otimes n}$, respectively, with asymptotically not vanishing expectation values $\mathrm{tr}P^{(n)}\rho^{(n)} = \eta_{d^{(n)}}(\rho^{(n)})$ and $\lim_{n\to\infty} \frac{1}{n}\log_2 \mathrm{tr}P^{(n)} = R(\mathcal{C}) < s(\Psi)$. This would be a contradiction to Proposition 5.2. $\Box$

Acknowledgement. The authors are grateful to Ruedi Seiler, Rainer Siegmund-Schultze and Tyll Krüger for helpful discussions and valuable comments on this paper.

This work was supported by the DFG via the SFB 288 "Quantenphysik und Differentialgeometrie" at the TU Berlin.

References

[1] H. Barnum, C. Fuchs, R. Jozsa, B. Schumacher, General Fidelity Limit for Quantum Channels, Phys. Rev. A 54, No 6, 4707-4711 (1996)

[2] I. Bjelaković, T. Krüger, Ra. Siegmund-Schultze, A. Szkoła, The Shannon-McMillan Theorem for Ergodic Quantum Lattice Systems, math.DS/0207121

[3] N. Datta, Yu. Suhov, Data Compression Limit for an Information Source of Interacting Qubits, Quant. Inf. Process. Vol. 1, No. 4, 257-281 (2002)

[4] E. B. Davies, Quantum Theory of Open Systems, Academic Press, London 1976

[5] M. Fannes, A Continuity Property of the Entropy Density for Spin Lattice Systems, Commun. Math. Phys. 31, 291-294 (1973)

[6] K.-E. Hellwig, K.Kraus, Pure Operations and Measurements, Commun. Math. Phys. 11, 214-220 (1969)

[7] K.-E. Hellwig, K.Kraus, Operations and Measurements II, Commun. Math. Phys. 16, 142-147 (1970)

[8] F. Hiai, D. Petz, The Proper Formula for Relative Entropy and its Asymptotics in Quantum Probability, Commun. Math. Phys. 143, 99-114 (1991)

[9] M. Horodecki, Limits for Compression of Quantum Information by Ensembles of Mixed States, Phys. Rev. A 57, 3364-3369 (1998)

[10] R.B. Israel, Convexity in the Theory of Lattice Gases, Princeton, New Jersey 1979

[11] R. Jozsa, B. Schumacher, A New Proof of the Quantum Noisless Coding theorem, Journal of Modern Optics, Vol.41, No.12, 2343-2349, (1994)

[12] K. Kraus, States, Effects and Operations. Fundamental Notions of Quantum Theory, Lecture Notes in Physics 190, Springer-Verlag, Berlin 1983

[13] M. A. Nielsen, I. L. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, Cambridge 2000

[14] D. Petz, M. Mosonyi, Stationary Quantum Source Coding, J. Math. Phys. 42, 4857-4864 (2001)

[15] D. Ruelle, Statistical Mechanics, W.A. Benjamin, New York 1969

[16] P.C. Shields, The Ergodic Theory of Discrete Sample Paths, American Mathematical Society 1996

[17] A. Uhlmann, Relative Entropy and the Wigner-Yanase-Dyson-Lieb Concavity in an Interpolation Theory, Commun. Math. Phys. 54, 21-32 (1977)