

Stationary quantum source coding¹

Dénes Petz² and Milán Mosonyi³

Mathematical Institute
Budapest University of Technology and Economics
H-1521 Budapest XI. Sztoczek u. 2, Hungary

Abstract

In this paper the quantum version of the source coding theorem is obtained for a completely ergodic source. This result extends Schumacher's quantum noiseless coding theorem for memoryless sources. The control of the memory effects requires some earlier results of Hiai and Petz on high-probability subspaces. Our result is equivalently considered as a compression theorem for noiseless stationary channels.

I. Introduction

Although it is difficult to define a discipline, to give some idea we can say that the objective of quantum information theory is the transmission and manipulation of information stored in systems obeying quantum mechanics. A quantum channel has a source that emits systems in quantum states to the channel. For example, the source could be a laser that emits individual monochromatic photons and the channel could be an optical fiber. The noisy signal output of the channel arrives at the receiver. In principle, there are two very different problems about quantum channels. The sender has a quantum system in an unknown state and wants to have the receiver end up with a similar system in the same state. In this case we speak of a pure quantum channel which has a quantum mechanical input and output. On the other hand, one might want to use quantum states to carry classical information, roughly speaking a sequence of zeros and ones. Now both the input and the output are classical, however there is a quantum mechanical section in between. The classical information is encoded into a quantum state and this is sent down the channel. The higher the channel noise is, the more redundant the encoding must be in order to restore the original signal at the receiver, where the quantum signal is converted into classical information. In this paper we do not deal with the problem of how such a scheme can be realistically implemented; practical quantum encoding and decoding requires a sophisticated ability to manipulate quantum states. However, we are interested in the amount of classical information getting through the channel which is assumed to be noiseless. It was emphasized already by Shannon that a computer memory is a communication channel. (Quantum or classical depends on the type of computer.) In an optimal situation the computer memory is free of any noise and this is the case we are concentrating on in the present paper. We want to consider rather general noiseless quantum

¹Published as J. Math. Phys. **42**(2001), 4857–4864.

²Supported by the Hungarian National Foundation for Scientific Research grant no. OTKA T 032662, e-mail: petz@math.bme.hu.

³E-mail: mosonyi@math.bme.hu.

channels (with possibly memory effects but strong ergodic properties) and our aim is to discuss the quantum source coding theorem. As a general reference on quantum information theory we suggest the recent book [1] but the really necessary definitions are given below.

To each classical input message x_i there corresponds a signal state φ_i of the quantum communication system. The quantum states φ_i are functioning as code words of the messages. The signal states φ_i could be pure and orthogonal in the sense of quantum mechanics but for example in quantum cryptography nonorthogonal states are used intentionally in order to avoid eavesdropping. At the moment we do not impose any condition on the signal states, they could be arbitrary pure or mixed states. In the stochastic model of communication, one assumes that each input message x_i appears with certain probability. Let p_{ji} be the probability that the message x_i is sent and y_j is received. The joint distribution p_{ji} yields marginal probability distributions p_i and q_j on the set of input and output messages. According to Shannon the mutual information

$$I = \sum_{i,j} p_{ji} \log \frac{p_{ji}}{p_i q_j}$$

measures the amount of information going through the channel from Alice to Bob. Of course, the relation of I to the quantum encoding and decoding should be made clear. This comes next.

The message x_i has *a priori* probability p_i and the mixed quantum state of the channel is

$$\varphi = \sum_i p_i \varphi_i.$$

This might be considered as the statistical operator of the *message ensemble*, for example when φ_i is a pure state $|i\rangle\langle i|$, then $\varphi = \sum_i p_i |i\rangle\langle i|$ acts on the input Hilbert space \mathcal{H} . The distribution of the output is determined by a measurement, which is nothing else but a physical word for decoding. To each output message there corresponds an observable A_j on the output Hilbert space \mathcal{K} . It is customary to assume that $0 \leq A_j$, $\sum_j A_j = \text{id}$ (id stands for the identity operator) and $p_{ji} = p_i \varphi_i(A_j)$. The so-called *Kholevo bound* ([2]) provides an upper bound on the amount of information accessible to Bob in terms of von Neumann entropies:

$$I \leq S(\varphi) - \sum_i p_i S(\varphi_i)$$

(When $\lambda_1, \lambda_2, \dots$ are the eigenvalues of the statistical operator of a quantum state ψ , then $S(\psi) = -\sum_k \lambda_k \log(\lambda_k)$.) In particular, if all signal states φ_i are pure, then $S(\varphi_i) = 0$ and we have $I \leq S(\varphi)$. In this way the von Neumann entropy gets an information theoretical interpretation. Kholevo's bound is actually not very strong, it is attained only in trivial situations ([3]).

The basic problem of communication theory is to maximize the amount of information received by Bob from Alice. However, up to now this problem is not well posed in our discussion yet. Let us deal with messages of length n , they are n -term-sequences of 0 and 1. (So the size of this message set is 2^n .) For each message length n we carry out the above procedure of coding and decoding and the amount of information going through the channel is I_n . Since I_n is presumably proportional to n , the good information quantity is I_n/n , that is, the transmitted information per letter. Since Shannon's theory is not only stochastic but asymptotic as well, we are going to let n to ∞ . In this way we need to repeat the above information transmission scheme for each n . The message set, the input Hilbert space $\mathcal{H}^{(n)}$,

our coding, the channel state $\varphi^{(n)}$, the output Hilbert space $\mathcal{K}^{(n)}$ and the observables applied in the measurement are all depending on the parameter n .

The subject of the present paper is faithful signal transmission, which bears the name noiseless channel. In place of faithful transmission, one can think of information storage. In this case the aim is to use the least possible number of Hilbert space dimension per signal for coding. The new feature of the noiseless channel we are studying is the memory effect. Mathematically this means that the channel state (of the n -fold channel) is not of product type but we assume stationarity and good ergodic properties. In Section 2 we use the standard formalism of statistical mechanics to describe such a channel. It turns out that the mean von Neumann entropy, familiar also from statistical mechanics, gives the optimal coding rate. The proof of our main result, Theorems 1. and 2., is similar to the proof presented in [4] for Schumacher's coding theorem, however instead of typical sequences we use the high-probability subspace of strongly ergodic stationary states, a subject studied by Hiai and Petz in [5]. We note for the interested reader that most of the concepts used in the present paper are treated in detail in the monograph [6].

II. An infinite system setting of the source

If \mathcal{H} is a finite-dimensional Hilbert space then $(A, B) \mapsto \text{Tr}(A^*B)$ defines an inner product on $\mathcal{B}(\mathcal{H})$, so for every linear functional φ on $\mathcal{B}(\mathcal{H})$ there exists a unique $D_\varphi \in \mathcal{B}(\mathcal{H})$ with the property $\varphi(A) = \text{Tr}(D_\varphi A)$. When φ is a state then D_φ is the corresponding density matrix. Let X^n denote the set of all messages of length n . If $x^n \in X^n$ is a message then a quantum state $\varphi(x^n)$ of the n -fold quantum system is corresponded with it. The Hilbert space of the n -fold system is the n -fold tensor product $\mathcal{H}^{\otimes n}$ and $\varphi(x^n)$ has a statistical operator $D(x^n)$. If messages of length n are to be transmitted then our quantum source should be put in the state $\varphi_n = \sum_{x^n} p(x^n)\varphi(x^n)$ with statistical operator $D_n = \sum_{x^n} p(x^n)D(x^n)$, where $p(x^n)$ is the probability of the message x^n . Since we want to let $n \rightarrow \infty$, it is reasonable to view all the n -fold systems as subsystems of an infinite one. In this way we can conveniently use a formalism standard in statistical physics, see Chap. 15 of [6].

Let an infinitely extended system be considered over the lattice \mathbb{Z} of integers. The observables confined to a lattice site $k \in \mathbb{Z}$ form the selfadjoint part of a finite-dimensional matrix algebra \mathcal{A}_k , that is the set of all operators acting on the finite-dimensional space \mathcal{H} . It is assumed that the local observables in any finite subset $\Lambda \subset \mathbb{Z}$ are those of the finite quantum system

$$\mathcal{A}_\Lambda = \bigotimes_{k \in \Lambda} \mathcal{A}_k.$$

The quasilocal algebra \mathcal{A} is the norm completion of the normed algebra $\mathcal{A}_\infty = \cup_\Lambda \mathcal{A}_\Lambda$, the union of all local algebras \mathcal{A}_Λ associated with finite intervals $\Lambda \subset \mathbb{Z}$.

A state φ of the infinite system is a positive normalized functional $\mathcal{A} \rightarrow \mathbb{C}$. It does not make sense to associate a statistical operator to a state of the infinite system in general. However, φ restricted to a finite-dimensional local algebra \mathcal{A}_Λ admits a density matrix D_Λ . We regard the algebra $\mathcal{A}_{[1,n]}$ as the set of all operators acting on the n -fold tensor product space $\mathcal{H}^{\otimes n}$. Moreover, we assume that the density D_n from the first part of this section is identical with $D_{[1,n]}$. Under this assumptions we call the state φ the state of the (infinite) channel. Roughly speaking, all the states used in the transmission of messages of length n are marginals of this φ . Coding, transmission and decoding could be well formulated using the states $\varphi_n \equiv \varphi_{[1,n]}$. However, it is more convenient to formulate our setting in the form of an infinite system, particularly because we do not want to assume that the channel state φ is

a product type. This corresponds to the possibility that our quantum source has a memory effect.

The right shift on the set \mathbb{Z} induces a transformation γ on \mathcal{A} . A state φ is called *stationary* if $\varphi \circ \gamma = \varphi$. The state φ is called *ergodic* if it is an extremal point in the set of stationary states. Moreover, φ is *completely ergodic* when it is an extreme point for every $m \in \mathbb{N}$ in the convex set of all states ψ such that $\psi \circ \gamma^m = \psi$. By a *completely ergodic stationary quantum source* we simply mean a completely ergodic stationary state φ of the infinite system \mathcal{A} . Of course, a stationary product state, corresponding to a memoryless channel, is completely ergodic. The emphasis is put on other states here.

Next we show an example of a completely ergodic stationary quantum source from the context of algebraic states. For the details see the original paper [7].

Example 1. Let $\mathcal{A} := M_3(\mathbb{C})$, $\mathcal{B} := M_2(\mathbb{C})$, moreover let $\{E_{ij}\}_{i,j=1}^3$ be the usual matrix units of $M_3(\mathbb{C})$. Set

$$V_1 := \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 \end{bmatrix}, \quad V_2 := \begin{bmatrix} 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix}, \quad V_3 := \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Then $\sum_{i=1}^3 V_i^* V_i = I_{\mathcal{B}}$.

Let ρ be a state on \mathcal{B} with density matrix

$$\begin{bmatrix} \frac{2}{3} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}.$$

Define $\Sigma : \mathcal{A} \otimes \mathcal{B} \rightarrow \mathcal{B}$ by $\Sigma(E_{ij} \otimes x) := V_i^* x V_j$. It is easy to check that Σ is a completely positive unital map and $\rho(\Sigma(I_{\mathcal{A}} \otimes x)) = \rho(x)$, $x \in \mathcal{B}$.

Then the algebraic state φ generated by $(\mathcal{B}, \Sigma, \rho)$ is given by

$$\varphi(E_{i_1 j_1} \otimes \dots \otimes E_{i_n j_n}) = \rho(V_{i_1}^* \dots V_{i_n}^* V_{j_n} \dots V_{j_1}).$$

It is shown in [7] that φ is completely ergodic. Of course, it is not a product state.

It is well known in quantum statistical mechanics that due to the subadditivity of the von Neumann entropy (proven first in [8] by Lieb and Ruskai) the limit

$$\lim_{n \rightarrow +\infty} \frac{1}{n} S(\varphi_n) = \inf \frac{1}{n} S(\varphi_n) =: h$$

exists for any stationary state and this quantity is called the *mean entropy* of φ . (See [6] for a textbook treatment of the subject or [9] for some related properties of the mean entropy.)

III. Source coding

For a while we fix a message length n and we denote by d the dimension of the Hilbert space \mathcal{H} . Assume that our n -fold composite quantum system is operating as a quantum source and emits the quantum states $D^{(1)}, D^{(2)}, \dots, D^{(m)}$ with a-priori probabilities p_1, p_2, \dots, p_m . (Therefore the state of the system is $D_n = \sum_i p_i D^{(i)}$.) By source coding we mean an association

$$D^{(i)} \mapsto \tilde{D}^{(i)},$$

where $\tilde{D}^{(i)}$ is some other statistical operator on the Hilbert space $\mathcal{H}^{\otimes n}$. (This definition allows $D^{(i)} = D^{(j)}$ but $\tilde{D}^{(i)} \neq \tilde{D}^{(j)}$, however in the coding constructed in the proof of Theorem 1 this cannot happen.)

We denote by \mathcal{K}_n the subspace spanned by the eigenvectors corresponding to all nonzero eigenvalues of all statistical operators $\tilde{D}^{(i)}$, $1 \leq i \leq m$. The goal of source coding is to keep the dimension of \mathcal{K}_n to be small and to fulfill some fidelity criterium. (A mathematically demanding survey about quantum coding is the paper [10].) The *source coding rate*

$$\limsup_{n \rightarrow \infty} \frac{\log \dim(\mathcal{K}_n)}{n}$$

expresses the resolution of the encoder in qubits per input symbol. (It is actually more precise to speak about “qunats” per input symbol, but the difference is only a constant factor.)

The distortion measure is a number which allows us to compare the goodness or badness of communication systems. The *fidelity* of the coding scheme was introduced by Schumacher ([11]):

$$F := \sum_i p_i \text{Tr} D^{(i)} \tilde{D}^{(i)},$$

where p_i is a probability distribution on the input and $\tilde{D}^{(i)}$ is the density used to encode the density $D^{(i)}$. Note that $0 \leq F \leq 1$ and $F = 1$ if and only if $D^{(i)} = \tilde{D}^{(i)}$ are pure states.

First we present our positive source coding theorem for a completely ergodic source. The result says that the source coding rate may approach the mean entropy while we can keep the fidelity arbitrarily good.

Theorem 1. *Let \mathcal{H} be a finite-dimensional Hilbert space, and φ be a completely ergodic state on $B(\mathcal{H})^{\otimes \infty}$. Then for every $\varepsilon, \delta > 0$ there exists $n_{\varepsilon, \delta} \in \mathbb{N}$ such that for $n \geq n_{\varepsilon, \delta}$ there is a subspace $\mathcal{K}_n(\varepsilon, \delta)$ of $\mathcal{H}^{\otimes n}$ such that*

(i) $\log \dim \mathcal{K}_n(\varepsilon, \delta) < n(h + \delta)$ and

(ii) *for every extremal decomposition $D_n = \sum_{i=1}^m p_i D^{(i)}$ one can find an encoding $D^{(i)} \mapsto \tilde{D}^{(i)}$ with density matrices $\tilde{D}^{(i)}$ supported in $\mathcal{K}_n(\varepsilon, \delta)$ such that the fidelity $F := \sum_{i=1}^m p_i \text{Tr} D^{(i)} \tilde{D}^{(i)}$ exceeds $1 - \varepsilon$.*

The negative part of the coding theorem tells that the source coding rate cannot exceed the mean entropy when the fidelity is good.

Theorem 2. *Let \mathcal{H} be a finite-dimensional Hilbert space, and φ be a completely ergodic state on $B(\mathcal{H})^{\otimes \infty}$. Then for every $\delta > 0$ there exist $0 < \eta < 1$ and $n_\delta \in \mathbb{N}$ such that for $n \geq n_\delta$*

(i) *for all subspaces \mathcal{K}_n of $\mathcal{H}^{\otimes n}$ with the property $\log \dim \mathcal{K}_n \leq n(h - \delta)$ and*

(ii) *for every decomposition $D_n = \sum_{i=1}^m p_i D^{(i)}$ and for every encoding $D^{(i)} \mapsto \tilde{D}^{(i)}$ with density matrices $\tilde{D}^{(i)}$ supported in \mathcal{K}_n , the fidelity $F := \sum_{i=1}^m p_i \text{Tr} D^{(i)} \tilde{D}^{(i)}$ is smaller than η .*

The detailed proofs are given in the next section of the paper. Now we make some comments on the fidelity F . It is possible that $F < 1$ although $D^{(i)} = \tilde{D}^{(i)}$. This fact might suggest to use another concept of fidelity. Since $D^{1/2} \geq D$ holds for a density matrix, we have

$$\begin{aligned} \text{Tr} D_1^{1/2} D_2^{1/2} &= \text{Tr} D_1^{1/4} D_2^{1/2} D_1^{1/4} \geq \text{Tr} D_1^{1/4} D_2 D_1^{1/4} = \text{Tr} D_2^{1/2} D_1^{1/2} D_2^{1/2} \\ &\geq \text{Tr} D_2^{1/2} D_1 D_2^{1/2} = \text{Tr} D_1 D_2. \end{aligned}$$

This implies that

$$F' := \sum_i p_i \text{Tr} [D^{(i)}]^{1/2} [\tilde{D}^{(i)}]^{1/2} \geq F.$$

Both our positive and negative source coding theorems hold if F is replaced by F' . (In case of Theorem 1. this follows from the inequality $F' \geq F$ and in the proof of Theorem 2. we will show $F' \leq \eta$.)

IV. High-probability subspace

The proof of Shannon's original source coding theorem is based on the typical sequences ([12], Chap. 1). The quantum extension of this result obtained by Schumacher still benefits from the classical result. When the channel state is a product, the densities D_n commute and simultaneous diagonalization is possible. If the memory effects are present, then these densities do not commute and in some sense we are in a really quantum mechanical noncommutative situation. Nevertheless, the high-probability subspace can be used but new techniques are required.

Let \mathcal{K} be a Hilbert space and D be a density matrix on \mathcal{K} . D has a Schatten decomposition $D = \sum_i \lambda_i |f_i\rangle\langle f_i|$, where $|f_i\rangle$'s are eigenvectors and the eigenvalues λ_i are numbered decreasingly: $\lambda_1 \geq \lambda_2 \geq \dots$. Choose and fix $0 < \varepsilon < 1$. Let $n(\varepsilon)$ be the smallest integer such that

$$\sum_{i=1}^{n(\varepsilon)} \lambda_i \geq 1 - \varepsilon.$$

The subspace $HP(D, \varepsilon)$ spanned by the eigenvectors $|f_1\rangle, \dots, |f_{n(\varepsilon)}\rangle$ is called the *high-probability subspace* corresponding to the level ε . Note that $HP(D, \varepsilon)$ is not completely well defined. If there are multiplicities in the spectrum of D , then the Schatten decomposition is not unique. However, the dimension $n(\varepsilon)$ of $HP(D, \varepsilon)$ is determined. The term "high-probability subspace" is borrowed from the monograph [13] and its role in macroscopic uniformity was discussed in [14].

In the following, φ will be a completely ergodic state on $\mathcal{A}^{\otimes \infty}$. For $\varepsilon \in (0,1)$ let

$$\beta_{\varepsilon,n} := \inf \{ \log \text{Tr}_n(q) : q \in \mathbb{P}(\mathcal{A}^{\otimes n}), \varphi_n(q) \geq 1 - \varepsilon \},$$

where $\mathbb{P}(\mathcal{A}^{\otimes n})$ denotes the set of projections of $\mathcal{A}^{\otimes n}$. ($\exp \beta_{\varepsilon,n}$ is the dimension of the high-probability subspace.) It was shown in [5] (and formulated in terms of relative entropy) that

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \beta_{\varepsilon,n} \leq h, \tag{1}$$

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \beta_{\varepsilon,n} \geq \frac{1}{1-\varepsilon} h - \frac{\varepsilon}{1-\varepsilon} \log d. \tag{2}$$

From this one can deduce the following

Proposition . *For every positive δ*

- (i) *and for every positive ε there exists $N_{\varepsilon,\delta} \in \mathbb{N}$ such that for every $n > N_{\varepsilon,\delta}$ there exists a projection $q_n(\varepsilon, \delta)$ in $\mathcal{A}^{\otimes n}$ such that*

$$\log(\text{Tr}_n(q_n(\varepsilon, \delta))) < n(h + \delta) \quad \text{and} \quad \varphi_n(q) \geq 1 - \varepsilon,$$

(ii) and there exists $1 > \eta > 0$ and $N_\delta \in \mathbb{N}$ such that for every $n > N_\delta$ and for every projection q in $\mathcal{A}^{\otimes n}$

$$\log(\text{Tr}_n(q)) \leq n(h - \delta),$$

implies $\varphi_n(q) \leq \eta$.

Part (i) of the proposition is a plain reformulation of (1). In order to see (ii) we first note that

$$\frac{1}{\eta}h - \frac{1-\eta}{\eta}\log d \rightarrow h \quad \text{as } \eta \rightarrow 1.$$

Hence given $\delta > 0$ we choose $0 < \eta < 1$ such that

$$\frac{1}{\eta}h - \frac{1-\eta}{\eta}\log d > h - \delta.$$

Next we replace ε by $1 - \delta$ in (2):

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \inf \{ \log \text{Tr}_n(q) : q \in \mathbb{P}(\mathcal{A}^{\otimes n}), \varphi_n(q) \geq 1 - \eta \} \geq \frac{1}{\eta}h - \frac{1-\eta}{\eta}\log d > h - \delta. \quad (3)$$

In this way we arrived at (ii).

Next we prove the source coding theorem.

Proof of Theorem 1.: Use part (i) of the proposition and set $q_n := q_n(\varepsilon/2, \delta)$, $\mathcal{K}_n(\varepsilon, \delta) := \text{Ran } q_n$, where $n > n(\varepsilon, \delta) := N_{\varepsilon/2, \delta}$. Given an extremal decomposition $D_n = \sum_{i=1}^k p_i D^{(i)}$, that is $D^{(i)} = |x_i\rangle\langle x_i|$ for some vectors x_i , we construct the coding densities $\tilde{D}^{(i)}$. Let

$$\tilde{x}_i := \frac{q_n x_i}{\|q_n x_i\|}, \quad \alpha_i := \|q_n x_i\|, \quad \beta_i := \|(I - q_n)x_i\|$$

and let x be any unit vector such that $q_n x = x$. Then we set

$$\tilde{D}^{(i)} := \alpha_i^2 |\tilde{x}_i\rangle\langle \tilde{x}_i| + \beta_i^2 |x\rangle\langle x|.$$

Since $\tilde{x}_i, x \in \mathcal{K}_n(\varepsilon, \delta)$, we have $\text{supp } \tilde{D}^{(i)} \subset \mathcal{K}_n(\varepsilon, \delta)$. Furthermore,

$$\begin{aligned} \text{Tr } D^{(i)} \tilde{D}^{(i)} &= \langle x_i, \tilde{D}^{(i)} x_i \rangle = \alpha_i^2 |\langle x_i | \tilde{x}_i \rangle|^2 + \beta_i^2 |\langle x_i, x \rangle|^2 \\ &\geq \alpha_i^2 |\langle x_i | \tilde{x}_i \rangle|^2 = \alpha_i^4 \geq 2\alpha_i^2 - 1 \\ &= 2\text{Tr } q_n D^{(i)} - 1. \end{aligned}$$

We need to sum over i :

$$\sum_i p_i \text{Tr } D^{(i)} \tilde{D}^{(i)} \geq \sum_i p_i (2\text{Tr } q_n D^{(i)} - 1) = 2\text{Tr } D_n q_n - 1 = 2\varphi_n(q_n) - 1 \geq 1 - \varepsilon.$$

Proof of Theorem 2.: For the given δ we choose η and $n(\delta)$ according to the proposition. Let q be the projection onto the subspace \mathcal{K}_n . We want to use the Schwarz inequality in the form

$$\left| \sum_i p_i \text{Tr } x_i y_i \right| \leq \left[\sum_i p_i \text{Tr } x_i^* x_i \right]^{1/2} \left[\sum_i p_i \text{Tr } y_i^* y_i \right]^{1/2}$$

for $x_i = [D^{(i)}]^{1/2}q$ and $y_i = [\tilde{D}^{(i)}]^{1/2}$. Since $[\tilde{D}^{(i)}]^{1/2} = q[\tilde{D}^{(i)}]^{1/2}$ follows from the hypothesis, we have

$$\begin{aligned} F' &= \sum_{i=1}^m p_i \text{Tr} [D^{(i)}]^{1/2} [\tilde{D}^{(i)}]^{1/2} = \sum_{i=1}^m p_i \text{Tr} [D^{(i)}]^{1/2} q [\tilde{D}^{(i)}]^{1/2} \\ &\leq \left[\sum_{i=1}^m p_i \text{Tr} D^{(i)} q \right]^{1/2} \left[\sum_{i=1}^m p_i \text{Tr} \tilde{D}^{(i)} \right]^{1/2} \\ &= \varphi_n(q)^{1/2} \leq \sqrt{\eta}. \end{aligned}$$

This estimate completes the proof.

It was also shown in [7] that for strongly mixing algebraic states (especially for product states)

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \beta_{\varepsilon, n} = h$$

and in this case the negative part of the coding theorem holds in a stronger form: For every $\varepsilon, \delta > 0$ there exists $n_{\varepsilon, \delta} \in \mathbb{N}$ such that for $n \geq n_{\varepsilon, \delta}$ for all subspaces \mathcal{K}_n of $\mathcal{H}^{\otimes n}$ with the property $\log \dim \mathcal{K}_n < n(h - \delta)$ and for every decomposition $D_n = \sum_{i=1}^m p_i D^{(i)}$ and for every encoding $D^{(i)} \mapsto \tilde{D}^{(i)}$ with density matrices $\tilde{D}^{(i)}$ supported in $\mathcal{K}_n(\varepsilon, \delta)$, the fidelity $F := \sum_{i=1}^m p_i \text{Tr} D^{(i)} \tilde{D}^{(i)}$ is smaller than ε .

There is a seemingly slight difference between the two theorems. The statistical operator D_n has an extremal decomposition in the first one and arbitrary decomposition in the second. The difference between the pure and mixed message ensemble is discussed in the recent paper [15].

V. Discussion

In this paper a theory of quantum source coding subject to a fidelity criterion or quantum data compression is presented. The minimum of the source coding rate is studied under the conditions that Schumacher's fidelity must exceed $1 - \varepsilon$ and the quantum mechanical state of the channel has a strong ergodic property. This latter condition allows many states with memory effect. For the mathematical model and in the proof of the main result techniques of quantum statistical mechanics are used. We prove that the minimal source coding rate is the mean entropy of the channel state, and, to some extent, it is independent of the message ensemble.

VI. Acknowledgement

The authors thank to Prof. O.E. Barndorff-Nielsen for an invitation to a workshop held at MaPhySto, to Prof. A. Holevo and Dr. S. Furuichi for comments on the first draft of the paper.

References

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

- [2] A.S. Holevo, Some estimates for the amount of information transmittable by a quantum communication channel, *Problems Inf. Transmission*, **9**(1973), 177–183
- [3] M. Ohya, D. Petz, N. Watanabe, On capacities of quantum channels, *Prob. Math. Stat.* **17**(1997), 179–196
- [4] R. Jozsa, B. Schumacher, A new proof of the quantum noiseless coding theorem, *J. Modern Optics*, **41**(1994), 2343–2349.
- [5] F. Hiai, D. Petz, The proper formula for relative entropy and its asymptotics in quantum probability, *Commun. Math. Phys.* **143**(1991), 99–114.
- [6] M. Ohya, D. Petz, *Quantum Entropy and Its Use*, Springer-Verlag, Heidelberg, 1993.
- [7] F. Hiai, D. Petz, Entropy density for algebraic states, *J. Functional Anal.* **125**(1994), 287–308.
- [8] E.H. Lieb, M.B. Ruskai, Proof of the strong subadditivity of quantum mechanical entropy, *J. Math. Phys.* **14**(1973), 1938–1941
- [9] D. Petz, Entropy density in quantum statistical mechanics and information theory, in *Contributions in Probability*, ed. C. Cecchini, 221–226, Forum, Udine, 1996.
- [10] A.S. Holevo, Quantum coding theorems, *Russian Math. Surveys*, **53**(1998), 1295–1331
- [11] B. Schumacher, Quantum coding, *Phys. Rev. A* **51**(1995), 2738–2747
- [12] I. Csiszár, J. Körner, *Information theory. Coding theorems for discrete memoryless systems*, Akadémiai Kiadó, Budapest, 1981
- [13] W.T. Grandy, Jr., *Foundations of Statistical Mechanics. Volume I: Equilibrium Theory*, D. Reidel, Dordrecht, 1987
- [14] F. Hiai, D. Petz, Entropy densities for Gibbs states of quantum spin systems, *Rev. Math. Phys.* **5**(1994), 693–712
- [15] H. Barnum, C.M. Caves, C.A. Fuchs, R. Jozsa, B.W. Schumacher, On quantum coding for ensembles of mixed states, [quant-ph/0008024](https://arxiv.org/abs/quant-ph/0008024)