Capacity of a memoryless quantum communication channel

Akio Fujiwara and Hiroshi Nagaoka

METR 94–22 December 1994

Capacity of a memoryless quantum communication channel

Akio Fujiwara* and Hiroshi Nagaoka†

Abstract

We introduce a proper framework of coding problems for a quantum memoryless channel and derive an asymptotic formula for the channel capacity, based upon the philosophy that any information theory shall be constructed in terms of observable quantities. Some general lower and upper bounds for the quantum channel capacity are also derived.

Keywords : quantum information theory, quantum communication, memoryless quantum channel, mutual information, quantum capacity, channel coding.

*Department of Mathematics, Osaka University, Osaka 560, Japan.

†Graduate School of Information Systems, University of Electro–Communications, Tokyo 182, Japan.

1 Introduction

In order to consider a communication system which is described by quantum mechanics, we must reformulate information (communication) theory in terms of quantum mechanical language. Suppose an input state ρ is transmitted through a quantum channel to yield the output state ρ'. The problem is, of cause, to measure how much information can be transmitted to the receiver observing the output state. In classical theory, it is measured by the mutual information. Shannon's fundamental result [1] asserts that the supremum of mutual information over input and output is identical to the channel capacity, i.e. the maximum rate under which one can transmit information within arbitrary small error probability. The counterpart of this result in quantal situation has been investigated by a number of researchers (see the recent review [2] and the references quoted therein). However, most of these works seems unsatisfactory since they hastily assumed *a priori* analogy between such quantum informational quantities as the von Neumann entropy, the quantum relative entropy, etc. and the Shannon entropy, the Kullback-Leibler divergence, etc., respectively, without establishing their physical observability and/or operational significance in quantum theoretical contexts. Furthermore, it seems that many researchers are involved in insufficient treatments of quantum information as yet due to the confusion on the ignorance interpretation of density operators [3].

Needless to say that one of the principal themes of the traditional information theory consists in establishing *coding theorems* in various contexts, through which many informational contents are equipped with operational significance in certain asymptotic frameworks. One of the reason for the immaturity of quantum information theory lies in the lack of asymptotic approaches, although there are a small number of excellent exceptions such as [4][5].

The purpose of this paper is to present a proper framework of coding problems for a quantum memoryless channel and to derive an asymptotic formula for the channel capacity, based upon the philosophy that any information theory shall be constructed in terms of observable quantities [6][7].

2 Quantum description of state, measurement, and channel

In information theory, we have to consider simultaneously two systems, an input system and an output system. Information of the input system is transmitted to the output system. A channel exhibits all dynamical effects for the information transmission. In Shannon's theory [1], the information of a system is carried by a probability distribution of events, and a channel induces a change of this probability distribution. The concept of state in a quantum system can be regarded as an extension of that of probability distribution. Therefore, the information of quantum systems is carried by a state, and a channel provides a dynamical change of states.

Let $\mathcal{H}$ be a separable Hilbert space which corresponds to the physical system of interest. A quantum state is represented by a *density operator* ρ on $\mathcal{H}$ which satisfies $\rho = \rho^* \geq 0$ and $\mathrm{Tr}\,\rho = 1$. A *measurement* $\{\Pi(B)\}_{B\in\mathcal{F}}$ on a measurable space $(\Omega, \mathcal{F})$ is an operator-valued set function which satisfy the following axioms [8, p. 50]:

1. $\Pi(\phi) = 0, \quad \Pi(\Omega) = I,$
2. $\Pi(B) = \Pi(B)^* \geq 0, \quad (^\forall B \in \mathcal{F}),$
3. $\Pi(\bigcup_j B_j) = \sum_j \Pi(B_j),$ (for all at most countable disjoint sequence $\{B_j\} \subset \mathcal{F}$).

In particular, a measurement Π is called *simple* if it satisfies, in addition to the above three axioms, $\Pi(B_1)\Pi(B_2) = 0$, $(^\forall B_1 \cap {}^\forall B_2 = \phi)$. In this paper, we restrict ourselves to finite dimensional Hilbert spaces and to measurements on finite sets for simplicity. By fixing a state ρ and a measurement Π on a finite set $\mathcal{X}$, the outcome of the measurement form an $\mathcal{X}$-valued random variable which obeys the probability distribution $p(x) = \mathrm{Tr}\,\rho\Pi(x)$, $(x \in \mathcal{X})$. Letting $\mathcal{P}(\mathcal{H}_j)$ be the set of states on $\mathcal{H}_j$, a *quantum channel* for an input system $\mathcal{H}_1$ and an output system $\mathcal{H}_2$ is described by a map $\Gamma : \mathcal{P}(\mathcal{H}_1) \to \mathcal{P}(\mathcal{H}_2)$ which satisfies

$$\Gamma(\lambda\rho_1 + (1-\lambda)\rho_2) = \lambda\Gamma(\rho_1) + (1-\lambda)\Gamma(\rho_2) \tag{1}$$

for all $\rho_1, \rho_2 \in \mathcal{P}(\mathcal{H}_1)$ and $0 \leq \lambda \leq 1$.

Physically, it is reasonable to assume that the operations Γ arise through the interaction of the system with an external quantum system $\mathcal{H}_c$ (i.e.,

physical reality of the channel) of the form

$$\Gamma\rho_1 = \mathrm{Tr}_{\mathcal{H}_c} U^*(\rho_1 \otimes \rho_c)U,$$

where $\rho_1 \in \mathcal{P}(\mathcal{H}_1)$, $\rho_c \in \mathcal{P}(\mathcal{H}_c)$ and U is unitary in $\mathcal{H}_1 \otimes \mathcal{H}_c$. Letting $\{|i\rangle\}, \{|\alpha\rangle\}$ be CONS in $\mathcal{H}_1$ and $\mathcal{H}_c$, respectively, then

$$V_\alpha \stackrel{\mathrm{def}}{=} \sum_i (I_1 \otimes \sqrt{\rho_c})U|i\alpha\rangle\langle i|, \qquad |i\alpha\rangle = |i\rangle \otimes |\alpha\rangle$$

satisfies

$$\Gamma\rho_1 = \sum_\alpha V_\alpha^*(\rho_1 \otimes I_c)V_\alpha,$$

which is the dual of a completely positive map by Stinespring's theorem[1] [9]. We therefore often make a more restrictive definition of a channel as the dual map of a certain completely positive map $\Lambda : \mathcal{B}(\mathcal{H}_2) \to \mathcal{B}(\mathcal{H}_1)$, i.e. $\Gamma = \Lambda^*$, where $\mathcal{B}(\mathcal{H}_j)$ denotes the set of (bounded) linear operators on $\mathcal{H}_j$ and the dual map Λ^* is defined by $\mathrm{Tr}\,\rho\Lambda(X) = \mathrm{Tr}\,\Lambda^*(\rho)X$ for all $X \in \mathcal{B}(\mathcal{H}_2)$, see [10] for details. For the present, however, only the assumption (1) is sufficient for our discussion.

In order to investigate asymptotic properties, we must consider the nth extension of the system which is described by tensor product $\bigotimes^n \mathcal{H} = \mathcal{H} \otimes \cdots \otimes \mathcal{H}$. In classical theory, this corresponds to the signal of length n. In quantum theory, in the same way, this extension is needed to describe the situation where the sender transmits n states $\{\sigma_j\}_{j=1}^n$ sequentially, which is represented by the state $\sigma_1 \otimes \cdots \otimes \sigma_n$ on $\bigotimes^n \mathcal{H}_1$. Now we define a quantum channel for extended input and output systems $\bigotimes^n \mathcal{H}_1$ and $\bigotimes^n \mathcal{H}_2$ by a map $\Gamma^{(n)} : \mathcal{P}(\bigotimes^n \mathcal{H}_1) \to \mathcal{P}(\bigotimes^n \mathcal{H}_2)$ which satisfies

$$\Gamma^{(n)}(\lambda\rho_1^{(n)} + (1-\lambda)\rho_2^{(n)}) = \lambda\Gamma^{(n)}(\rho_1^{(n)}) + (1-\lambda)\Gamma^{(n)}(\rho_2^{(n)}) \qquad (2)$$

for all $\rho_1^{(n)}, \rho_2^{(n)} \in \mathcal{P}(\bigotimes^n \mathcal{H}_1)$ and $0 \le \lambda \le 1$. Further, a channel $\Gamma^{(n)}$ is called *memoryless* if

$$\Gamma^{(n)}(\sigma_1 \otimes \cdots \otimes \sigma_n) = (\Gamma\sigma_1) \otimes \cdots \otimes (\Gamma\sigma_n).$$

[1] Suppose $\mathcal{A}$ is a C^*-algebra with unit, $\mathcal{H}$ a Hilbert space, and $\mathcal{B} = \mathcal{B}(\mathcal{H})$ the set of bounded linear operators on $\mathcal{H}$. A map $\Phi : \mathcal{A} \to \mathcal{B}$ is completely positive iff there is a $*$-representation π of $\mathcal{A}$ into operators on a Hilbert space $\mathcal{H}'$ and a bounded linear map $V : \mathcal{H} \to \mathcal{H}'$ such that $\Phi A = V^*\pi(A)V$.

A memoryless channel $\Gamma^{(n)}$ is thus determined uniquely by Γ. For a memoryless channel, we therefore often drop the superscript (n) for simplicity. In the following, we only consider memoryless channels.

Finally, we quote an important fact with respect to the quantum relative entropy (a quantum counterpart of the Kullback–Leibler divergence) for two quantum states ρ and σ defined by

$$D(\rho\|\sigma) \stackrel{\text{def}}{=} \operatorname{Tr}\rho(\log\rho - \log\sigma).$$

By fixing a measurement Π on a finite set $\mathcal{X}$, we have two probability distributions $p(x) = \operatorname{Tr}\rho\Pi(x)$ and $q(x) = \operatorname{Tr}\sigma\Pi(x)$. Let us denote the corresponding (classical) Kullback–Leibler divergence

$$D(p\|q) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)}$$

by $D_\Pi(\rho\|\sigma)$. It is well-known that the monotonicity

$$D(\rho\|\sigma) \geq D_\Pi(\rho\|\sigma) \tag{3}$$

holds for all measurements Π, and the equality is attainable by a certain measurement Π when and only when $\rho\sigma = \sigma\rho$, see for instance [11, Theorem 1.5, Theorem 5.3].

3 Quantum channel coding theorem

Suppose an input state $\sigma^{(n)} = \sigma_1 \otimes \cdots \otimes \sigma_n$ is transmitted through a memoryless channel Γ to yield the output state $\Gamma\sigma^{(n)} = \Gamma\sigma_1 \otimes \cdots \otimes \Gamma\sigma_n$. We shall investigate how much information can be transmitted to the receiver observing the output state.

We first prepare a finite set of quantum states on $\bigotimes^n \mathcal{H}_1$, each element of which is an n-tensor product of states on $\mathcal{H}_1$:

$$\mathcal{C}_n = \{\sigma^{(n)}(1), \cdots, \sigma^{(n)}(M_n)\}, \quad \sigma^{(n)}(k) = \sigma_1(k)\otimes\cdots\otimes\sigma_n(k), \quad \sigma_j(k) \in \mathcal{P}(\mathcal{H}_1).$$

In connection with classical communication channel, n, $\mathcal{C}_n$, and $\sigma^{(n)}(j)$ correspond to a codelength, a codebook, and a codeword, respectively. The transmitter first selects a codeword $\sigma^{(n)} = \sigma_1 \otimes \cdots \otimes \sigma_n$ which corresponds to the message to be transmitted (encoding), and then transmits each signal $\sigma_1, \cdots, \sigma_n$ successively through a memoryless channel Γ. The receiver then

receives signals $\Gamma\sigma_1, \cdots, \Gamma\sigma_n$ and, by means of a certain measuring process, he estimates which signal among $\mathcal{C}_n$ has been actually transmitted (decoding). In this case, the decoder is described by a $\mathcal{C}_n$-valued measurement $T^{(n)}$ over $\overset{n}{\bigotimes} \mathcal{H}_2$.

In such a communication system, the probability for the event that the decoded codeword is $\tau^{(n)}$ under the condition that the transmitted codeword is $\sigma^{(n)}$ is given by

$$\mathrm{Tr}\left[(\Gamma\sigma^{(n)})\, T^{(n)}(\tau^{(n)})\right].$$

Therefore, the error probability is defined by

$$P_e(\mathcal{C}_n, \sigma^{(n)}, T^{(n)}) = 1 - \mathrm{Tr}\left[(\Gamma\sigma^{(n)})\, T^{(n)}(\sigma^{(n)})\right]. \tag{4}$$

Let us define the average error probability for the code $\mathcal{C}_n$ and the decoder $T^{(n)}$ by

$$P_e(\mathcal{C}_n, T^{(n)}) = \frac{1}{|\mathcal{C}_n|} \sum_{\sigma^{(n)} \in \mathcal{C}_n} P_e(\mathcal{C}_n, \sigma^{(n)}, T^{(n)}), \tag{5}$$

where $|\mathcal{C}_n|$ denotes the cardinality of $\mathcal{C}_n$ (number of codewords), i.e. $|\mathcal{C}_n| = M_n$. We further define the error probability of $\mathcal{C}_n$ by

$$P_e(\mathcal{C}_n) = \inf_{T^{(n)}} P_e(\mathcal{C}_n, T^{(n)}). \tag{6}$$

Now, the quantity

$$R_n = \frac{\log |\mathcal{C}_n|}{n} \tag{7}$$

is called the *rate* for the code $\mathcal{C}_n$. Consider sequences of codes $\{\mathcal{C}_n\}_n$ which satisfy $\lim_{n\to\infty} P_e(\mathcal{C}_n) = 0$, and denote the supremum of $\lim_{n\to\infty} R_n$ over such sequences by $C(\Gamma)$, which is called the *capacity* of the channel Γ.

Let us derive the quantum counterpart of channel coding theorem which establishes the relation between the capacity $C(\Gamma)$ and the mutual information. By fixing arbitrarily (not necessarily $\mathcal{C}_n$-valued) a measurement $\Pi^{(n)}$ on a finite set $\mathcal{Y}$ over $\overset{n}{\bigotimes} \mathcal{H}_2$, we have the (classical) conditional probability distribution of state-valued input random variable X and $\mathcal{Y}$-valued output random variable Y which becomes

$$p(Y = y | X = \sigma^{(n)}) = \mathrm{Tr}\left[(\Gamma\sigma^{(n)})\, \Pi^{(n)}(y)\right]. \tag{8}$$

Therefore, by introducing arbitrarily a probability distribution $p^{(n)}(\sigma^{(n)})$ over $\mathcal{P}(\overset{n}{\bigotimes} \mathcal{H}_1)$, we have a joint distribution

$$p(X = \sigma^{(n)}, Y = y) = p^{(n)}(\sigma^{(n)})\, \mathrm{Tr}\left[(\Gamma\sigma^{(n)})\Pi^{(n)}(y)\right], \tag{9}$$

Since we are dealing with successive transmissions of states through a channel, we only need to consider such distributions $p^{(n)}(\sigma^{(n)})$ over $\mathcal{P}(\bigotimes^n \mathcal{H}_1)$ whose support (set of elementary events) are composed of finite number of product states $\sigma^{(n)} = \sigma_1 \otimes \cdots \otimes \sigma_n$. In this case, the distribution $p^{(n)}(\sigma^{(n)})$ can be identified with a simultaneous distribution $p^{(n)}(\sigma_1, \cdots, \sigma_n)$ over $\mathcal{P}(\mathcal{H}_1)^n = \mathcal{P}(\mathcal{H}_1) \times \cdots \times \mathcal{P}(\mathcal{H}_1)$. Denote the totality of such distributions by $\mathfrak{P}^{(n)}$. Given an arbitrary probability distribution $p^{(n)}(\sigma^{(n)}) \in \mathfrak{P}^{(n)}$, we define a mixture of states

$$\rho^{(n)} \stackrel{\text{def}}{=} \sum_{\sigma^{(n)}} p^{(n)}(\sigma^{(n)})\, \sigma^{(n)} = \sum_{\sigma_1, \cdots, \sigma_n} p^{(n)}(\sigma_1, \cdots, \sigma_n)\, \sigma_1 \otimes \cdots \otimes \sigma_n. \tag{10}$$

This can also be regarded as a decomposition of $\rho^{(n)}$, each $\sigma^{(n)} = \sigma_1 \otimes \cdots \otimes \sigma_n$ being elementary events which occur according to the probability $p^{(n)}(\sigma^{(n)})$.

Lemma 1 *The classical mutual information $I(X;Y)$ for the joint distribution (9) is identical to*

$$I^{(n)}(p^{(n)}, \Pi^{(n)}; \Gamma) \stackrel{\text{def}}{=} \sum_{\sigma^{(n)}} p^{(n)}(\sigma^{(n)}) D_{\Pi^{(n)}}\left(\Gamma\sigma^{(n)} \middle\| \Gamma\rho^{(n)}\right). \tag{11}$$

Proof

$$\begin{aligned} I(X;Y) &= D(p(x,y) \| p(x)p(y)) \\ &= \sum_x p(x) D(p(y|x) \| p(y)) \\ &= \sum_{\sigma^{(n)}} p^{(n)}(\sigma^{(n)}) D_{\Pi^{(n)}}\left(\Gamma\sigma^{(n)} \middle\| \Gamma\rho^{(n)}\right) \end{aligned}$$

■

Lemma 2 *For a memoryless channel Γ, the quantity*

$$C^{(n)}(\Gamma) \stackrel{\text{def}}{=} \sup_{p^{(n)} \in \mathfrak{P}^{(n)}} \sup_{\Pi^{(n)}} I^{(n)}(p^{(n)}, \Pi^{(n)}; \Gamma). \tag{12}$$

exhibits the following superadditivity:

$$C^{(m+n)}(\Gamma) \geq C^{(m)}(\Gamma) + C^{(n)}(\Gamma). \tag{13}$$

Proof If we adopt such a probability distribution $p^{(m+n)}$ over $\mathcal{P}(\mathcal{H}_1)^{m+n}$ that satisfies $p^{(m+n)} = p^{(m)}p^{(n)}$, then

$$\begin{aligned}\rho^{(m+n)} &= \sum_{\sigma^{(m+n)}} p^{(m+n)}(\sigma^{(m+n)})\sigma_1 \otimes \cdots \otimes \sigma_m \otimes \sigma_{m+1} \otimes \cdots \otimes \sigma_{m+n} \\ &= \rho^{(m)} \otimes \rho^{(n)}.\end{aligned}$$

Therefore we have

$$\begin{aligned}&C^{(m+n)}(\Gamma) \\ &= \sup_{p^{(m+n)},\Pi^{(m+n)}} I^{(m+n)}(p^{(m+n)}, \Pi^{(m+n)}; \Gamma) \\ &= \sup_{p^{(m+n)},\Pi^{(m+n)}} \sum_{\sigma^{(m+n)}} p^{(m+n)}(\sigma^{(m+n)}) D_{\Pi^{(m+n)}}\left(\Gamma\sigma^{(m+n)} \middle\| \Gamma\rho^{(m+n)}\right) \\ &\geq \sup_{\substack{p^{(m+n)} = p^{(m)}p^{(n)} \\ \Pi^{(m+n)} = \Pi^{(m)} \otimes \Pi^{(n)}}} \sum_{\sigma^{(m+n)}} p^{(m+n)}(\sigma^{(m+n)}) D_{\Pi^{(m+n)}}\left(\Gamma\sigma^{(m+n)} \middle\| \Gamma\rho^{(m+n)}\right) \\ &= \sup_{\substack{p^{(m)}, p^{(n)} \\ \Pi^{(m)}, \Pi^{(m)}}} \sum_{\sigma^{(m+n)}} p^{(m+n)}(\sigma^{(m+n)}) D_{\Pi^{(m)}\otimes\Pi^{(n)}}\left(\Gamma\sigma^{(m)} \otimes \Gamma\sigma^{(n)} \middle\| \Gamma\rho^{(m)} \otimes \Gamma\rho^{(n)}\right) \\ &= C^{(m)}(\Gamma) + C^{(n)}(\Gamma).\end{aligned}$$

■

Note that the superadditivity of $C^{(n)}$ implies $C^{(n)} \geq nC^{(1)}$, which is in remarkable contrast to classical channel.

Now, let us establish the quantum counterpart of Shannon's channel coding theorem.

Theorem 1 *For a memoryless channel* Γ,

$$C(\Gamma) = \lim_{n\to\infty} \frac{C^{(n)}(\Gamma)}{n} = \sup_n \frac{C^{(n)}(\Gamma)}{n} \tag{14}$$

Proof The second equality follows from the superadditivity (13). We prove the first equality.

Let us define the asymptotic rate of the code $\mathcal{C}_n$ by

$$R = \lim_{n\to\infty} \frac{\log|\mathcal{C}_n|}{n}. \tag{15}$$

If $R < C(\Gamma)$, there exist such m that satisfies $R < C^{(m)}(\Gamma)/m$. Then by fixing $\Pi^{(m)}$ which attains $C^{(m)}(\Gamma)$, the channel becomes a classical one whose capacity is $C^{(m)}(\Gamma)$ per m signals. Then invoking to the conventional scheme by using the random coding technique, we can prove that there exists a coding where the error probability can be suppressed arbitrarily small.

On the other hand, by using Fano's inequality and assuming the uniform distribution on the codebook $\mathcal{C}_n$, the average error probability $P_e(\mathcal{C}_n, T^{(n)})$ is evaluated as

$$\begin{aligned}
\log 2 + P_e(\mathcal{C}_n, T^{(n)}) \log |\mathcal{C}_n| &\geq H(\hat{\sigma}^{(n)}|\hat{\tau}^{(n)}) \\
&= H(\hat{\sigma}^{(n)}) - I(\hat{\sigma}^{(n)}; \hat{\tau}^{(n)}) \\
&\geq \log |\mathcal{C}_n| - \sup_{p^{(n)}, T^{(n)}} I^{(n)}(p^{(n)}, T^{(n)}; \Gamma) \\
&\geq \log |\mathcal{C}_n| - \sup_{p^{(n)}, \Pi^{(n)}} I^{(n)}(p^{(n)}, \Pi^{(n)}; \Gamma),
\end{aligned}$$

where $\hat{\sigma}^{(n)}$ denotes the $\mathcal{C}_n$-valued random variable which is uniformly distributed over $\mathcal{C}_n$, and $\hat{\tau}^{(n)}$ denotes the $\mathcal{C}_n$-valued random variable which corresponds to the decoded words when the decoder $T^{(n)}$ is applied to the output state $\Gamma\hat{\sigma}^{(n)}$. This inequality leads to

$$\left(1 - P_e(\mathcal{C}_n, T^{(n)})\right) \frac{\log |\mathcal{C}_n|}{n} \leq \frac{C^{(n)}(\Gamma)}{n} + \frac{\log 2}{n}.$$

Then, in order to assure $P_e(\mathcal{C}_n) \to 0$ as $n \to \infty$, R must be less than $C(\Gamma)$. ■

4 Bounds for the quantum capacity

Theorem 1 asserts that the proper counterpart of channel capacity is given by (14). However, one cannot expect to compute $C(\Gamma)$ directly from the definition in general since it contains an apparently impracticable operation $\sup_{\Pi^{(n)}}$ and, a fortiori, is not single-letterized. It is natural to ask whether $C(\Gamma)$ can be expressed in a simpler form. We intend to explore here some basic relations between the quantum capacity $C(\Gamma)$ and another capacity-like quantities.

An $\mathcal{X}$-valued measurement $\Pi^{(n)}$ over $\overset{n}{\bigotimes} \mathcal{H}$ is called *recursive* if it takes

the form

$$\Pi^{(n)}(x) = \sum_{y^n : f(y^n)=x} \left[\bigotimes_{j=1}^{n} \Pi_j(y_j \,|\, y^{j-1}) \right].$$

Here, $\Pi_j(\,\cdot\,|\,y^{j-1})$ is a $\mathcal{Y}_j$-valued measurement over $\mathcal{H}$ possibly depending on the previous data $y^{j-1} = (y_1, \cdots, y_{j-1}) \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_{j-1}$, and

$$f : \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n \to \mathcal{X}$$

is a mapping. Consider sequences of codes $\{\mathcal{C}_n\}_n$ which satisfy $\lim_{n\to\infty} P_e(\mathcal{C}_n) = 0$ provided the infimum in (6) is taken over all recursive decoders $T^{(n)}$ here, and denote by $C^{\otimes}(\Gamma)$ the supremum of $\lim_{n\to\infty} R_n$ over such sequences.

Theorem 2

$$C^{\otimes}(\Gamma) = C^{(1)}(\Gamma).$$

Proof $C^{\otimes}(\Gamma) \geq C^{(1)}(\Gamma)$ is trivial. We show the converse inequality. Let $\mathcal{C}_n$ be a code and $T^{(n)}$ a recursive decoder of the form

$$T^{(n)}(\sigma^{(n)}) = \sum_{y^n : f(y^n)=\sigma^{(n)}} \left[\bigotimes_{j=1}^{n} \Pi_j(y_j \,|\, y^{j-1}) \right].$$

Further, let X^n be a random variable which is uniformly distributed on $\mathcal{C}_n$. Note that X^n is written as $X^n = (X_1, \cdots, X_n)$ by a set of $\mathcal{P}(\mathcal{H}_1)$-valued random variables $\{X_j\}_j$. Let $Y^n = (Y_1, \cdots, Y_n)$ be the random variable representing outcomes of the measurement $\{\bigotimes_{j=1}^{n} \Pi_j(y_j \,|\, y^{j-1})\}$ applied to the output system in the state $\bigotimes_{j=1}^{n} \Gamma X_j$. Now, by using Fano's inequality and Lemma 3 which shall be proved below, we have

$$\begin{aligned} \log 2 + P_e(\mathcal{C}_n, T^{(n)}) \log |\mathcal{C}_n| &\geq H(X^n|Y^n) \\ &= H(X^n) - I(X^n; Y^n) \\ &\geq \log |\mathcal{C}_n| - nC^{(1)}(\Gamma). \end{aligned}$$

Thus we have

$$\Big(1 - P_e(\mathcal{C}_n, T^{(n)})\Big) \log |\mathcal{C}_n| \leq \log 2 + nC^{(1)}(\Gamma).$$

Since $\lim_{n\to\infty} P_e(\mathcal{C}_n) = 0$ is assumed, it follows that

$$\limsup_{n\to\infty} \frac{\log |\mathcal{C}_n|}{n} \leq C^{(1)}(\Gamma),$$

which proves $C^{\otimes}(\Gamma) \leq C^{(1)}(\Gamma)$. ■

Lemma 3

$$I(X^n; Y^n) \leq nC^{(1)}(\Gamma). \tag{16}$$

Proof We observe

$$I(X^n; Y^n) = \sum_{j=1}^{n} I(X^n; Y_j \,|\, Y^{j-1}) = \sum_{j=1}^{n} I(X_j; Y_j \,|\, Y^{j-1}),$$

where the first equality follows from the chain rule for the mutual information, and the second equality follows from the fact that $X^n Y^{j-1} \to X_j Y^{j-1} \to Y_j$ forms a Markov chain in this order since

$$p(y_j \,|\, x^n y^{j-1}) = \mathrm{Tr}\,[x_j \Pi_j(y_j \,|\, y^{j-1})] = p(y_j \,|\, x_j y^{j-1}).$$

Furthermore,

$$\begin{aligned} I(X_j; Y_j \,|\, Y^{j-1}) &= \sum_{y^{j-1}} p(y^{j-1}) I(X_j; Y_j \,|\, Y^{j-1} = y^{j-1}) \\ &= \sum_{y^{j-1}} p(y^{j-1}) I^{(1)}(p_{X_j}(\,\cdot\,), \Pi_j(\,\cdot \,|\, y^{j-1})\,;\, \Gamma) \\ &\leq \sum_{y^{j-1}} p(y^{j-1}) \sup_{p,\Pi} I^{(1)}(p, \Pi;\, \Gamma) \\ &= C^{(1)}(\Gamma), \end{aligned}$$

which proves the assertion. ■

Theorem 2 implies that $C^{\otimes}(\Gamma) = C^{(1)}(\Gamma)$ gives a general lower bound for the meaningful quantum capacity $C(\Gamma)$. In other words, the capacity $C(\Gamma)$ cannot be attained by means of a recursive measurement unless $C(\Gamma) = C^{(1)}(\Gamma)$. Thus, it is essential to consider measurements over the extended Hilbert space which cannot be realized in a recursive manner.

We next define the following quantity:

$$\tilde{C}^{(n)}(\Gamma) \stackrel{\text{def}}{=} \sup_{p^{(n)} \in \mathfrak{P}^{(n)}} \tilde{I}^{(n)}(p^{(n)}; \Gamma), \tag{17}$$

where

$$\tilde{I}^{(n)}(p^{(n)}; \Gamma) \stackrel{\text{def}}{=} \sum_{\sigma^{(n)}} p^{(n)}(\sigma^{(n)}) D(\Gamma\sigma^{(n)} \parallel \Gamma\rho^{(n)}) \tag{18}$$

defines a formal (purely quantal) mutual information ($\rho^{(n)}$ is defined in (10)).

Theorem 3

(i) $\tilde{C}^{(n)}(\Gamma) = n\tilde{C}^{(1)}(\Gamma)$,

(ii) $\tilde{C}^{(1)}(\Gamma) \geq C(\Gamma)$.

Proof (i) From Lemma 4 below, we confirm that

$$\begin{aligned} \tilde{C}^{(n)}(\Gamma) &\leq \sup_{p^{(n)} \in \mathfrak{P}^{(n)}} \sum_{j=1}^{n} \tilde{I}^{(1)}(p_j; \Gamma) \\ &\leq \sum_{j=1}^{n} \sup_{p_j} \tilde{I}^{(1)}(p_j; \Gamma) \\ &= n\tilde{C}^{(1)}(\Gamma). \end{aligned}$$

On the other hand, when $\sigma_1, \cdots, \sigma_n$ are drawn independently, $\rho^{(n)}$ becomes

$$\rho^{(n)} = \sum_{\sigma_1, \cdots, \sigma_n} p_1(\sigma_1) \cdots p_n(\sigma_n)\, \sigma_1 \otimes \cdots \otimes \sigma_n = \bigotimes_{j=1}^{n} \rho_j.$$

Then by restricting $p^{(n)}$ to i.i.d., $C^{(n)}(\Gamma)$ can be evaluated from below as

$$\begin{aligned} \tilde{C}^{(n)}(\Gamma) &= \sup_{p^{(n)} \in \mathfrak{P}^{(n)}} \sum_{\sigma^{(n)}} p^{(n)}(\sigma^{(n)}) D(\Gamma\sigma^{(n)} \parallel \Gamma\rho^{(n)}) \\ &\geq \sup_{p^{(n)}: \text{i.i.d.}} \sum_{\sigma^{(n)}} p^{(n)}(\sigma^{(n)}) D(\Gamma\sigma^{(n)} \parallel \Gamma\rho^{(n)}) \\ &= \sup_{p} \sum_{\sigma_1, \cdots, \sigma_n} p(\sigma_1) \cdots p(\sigma_n) D(\bigotimes_{j=1}^{n} \Gamma\sigma_j \parallel \bigotimes_{j=1}^{n} \Gamma\rho_j) \\ &= \sup_{p} \sum_{j=1}^{n} \sum_{\sigma_j} p(\sigma_j) D(\Gamma\sigma_j \parallel \Gamma\rho_j) \\ &= n\tilde{C}^{(1)}(\Gamma). \end{aligned}$$

Thus we have $\tilde{C}^{(n)}(\Gamma) = n\tilde{C}^{(1)}(\Gamma)$.

(ii) By using the monotonicity of relative entropy (3), we immediately have

$$\tilde{I}^{(n)}(p^{(n)};\Gamma) \geq I^{(n)}(p^{(n)},\Pi^{(n)};\Gamma),$$

which implies $\tilde{C}^{(n)}(\Gamma) \geq C^{(n)}(\Gamma)$. Then by using (i),

$$\tilde{C}^{(1)}(\Gamma) = \frac{\tilde{C}^{(n)}(\Gamma)}{n} \geq \frac{C^{(n)}(\Gamma)}{n}.$$

Taking the limit $n \to \infty$, we have (ii). ■

Lemma 4 *Denote the jth marginal of* $p^{(n)}(\sigma_1,\cdots,\sigma_n)$ *by* $p_j(\sigma_j)$. *Then*

$$\tilde{I}^{(n)}(p^{(n)};\Gamma) \leq \sum_{j=1}^{n} \tilde{I}^{(1)}(p_j;\Gamma).$$

Proof We first observe

$$\begin{aligned}\tilde{I}^{(n)}(p^{(n)};\Gamma) &= \sum_{\sigma^{(n)}} p^{(n)}(\sigma^{(n)})\mathrm{Tr}\,(\Gamma\sigma^{(n)})\left[\log(\Gamma\sigma^{(n)}) - \log(\Gamma\rho^{(n)})\right] \\ &= S_N(\Gamma\rho^{(n)}) - \sum_{\sigma^{(n)}} p^{(n)}(\sigma^{(n)})S_N(\Gamma\sigma^{(n)}), \qquad (19)\end{aligned}$$

where S_N denotes the von Neumann's entropy defined by $S_N(\rho) = -\mathrm{Tr}\,\rho\log\rho$. Since Γ is assumed memoryless,

$$\Gamma\rho^{(n)} = \sum_{\sigma^{(n)}} p^{(n)}(\sigma_1,\cdots,\sigma_n)\,\Gamma\sigma_1\otimes\cdots\otimes\Gamma\sigma_n.$$

Then due to the subadditivity of S_N, the first term of the last side in (19) is evaluated from the above as

$$S_N(\Gamma\rho^{(n)}) \leq \sum_{j=1}^{n} S_N\left(\sum_{\sigma_j} p_j(\sigma_j)\Gamma\sigma_j\right) = \sum_{j=1}^{n} S_N(\Gamma\rho_j),$$

where

$$\rho_j = \sum_{\sigma_j} p_j(\sigma_j)\sigma_j.$$

On the other hand,

$$S_N(\Gamma\sigma^{(n)}) = S_N(\Gamma\sigma_1 \otimes \cdots \otimes \Gamma\sigma_n) = \sum_{j=1}^{n} S_N(\Gamma\sigma_j)$$

holds owing to the additivity of S_N. Then (19) is evaluated from above as

$$\begin{aligned} \tilde{I}^{(n)}(p^{(n)}; \Gamma) &\leq \sum_{j=1}^{n} \left[S_N(\Gamma\rho_j) - \sum_{\sigma_j} p_j(\sigma_j) S_N(\Gamma\sigma_j) \right] \\ &= \sum_{j=1}^{n} \sum_{\sigma_j} p_j(\sigma_j) D(\Gamma\sigma_j \parallel \Gamma\rho_j) \\ &= \sum_{j=1}^{n} I^{(1)}(p_j; \Gamma). \end{aligned}$$

This proves the assertion. ■

Theorem 3 indicates that $\tilde{C}^{(n)}(\Gamma)$ of a memoryless quantum channel Γ has a similar property (i) to the capacity of a memoryless classical channel. Thus, we shall call $\tilde{C}(\Gamma) = \tilde{C}^{(1)}(\Gamma)$ the *pseudo-capacity* of a memoryless quantum channel Γ. Taking advantage of this nomenclature, we also call the formal mutual information (18) the *pseudo-mutual information*. Theorem 3 also reveals that the pseudo-capacity only gives a general upper bound for the meaningful quantum capacity $C(\Gamma)$. There exist, of cause, certain class of channels Γ for which the equality in (ii) holds (a trivial example : a channel which transmits every states onto a commutative subalgebra). Thus the following problem naturally arises.

Problem 1 *What is the necessary and sufficient condition for the equality*

$$\tilde{C}(\Gamma) = C(\Gamma) \ ?$$

The pseudo-capacity $\tilde{C}(\Gamma)$ can be represented also in the form

$$\tilde{C}(\Gamma) = \sup_{\rho \in \mathcal{P}(\mathcal{H}_1)} \tilde{I}(\rho; \Gamma),$$

where $\tilde{I}(\rho; \Gamma)$ is defined as the supremum of $\tilde{I}(p; \Gamma)$ over all distributions $p \in \mathfrak{P}^{(1)}$ satisfying

$$\rho = \sum_{\sigma \in \mathcal{P}(\mathcal{H}_1)} p(\sigma)\sigma. \tag{20}$$

This can be regarded as a decomposition of ρ into a mixture of elementary events σ, see (10). In the same situation as ours, Ohya [10] introduced another quantity, say $\tilde{I}^O(\rho\,;\,\Gamma)$, as the supremum of $\tilde{I}(p\,;\,\Gamma)$ over all Schatten decomposition of ρ, i.e. over all distributions $p \in \mathfrak{P}^{(1)}$ which satisfy (20) and whose support consists of mutually orthogonal pure states, and called it the mutual entropy. He showed that if Γ is the dual Λ^* of a completely positive map Λ, the following inequality holds:

$$\tilde{I}^O(\rho\,;\,\Lambda^*) \leq \min\{S_N(\rho), S_N(\Lambda^*\rho)\}. \tag{21}$$

Here we note that

$$\tilde{I}^O(\rho\,;\,\Lambda^*) \leq \tilde{I}(\rho\,;\,\Lambda^*) \leq \min\{S_N(\rho), S_N(\Lambda^*\rho)\}. \tag{22}$$

The first inequality is a straightforward consequence of the definition and the second one is proved in a similar way to (21). Although Ohya has called $\sup_{\rho\in\mathcal{P}(\mathcal{H}_1)} I^O(\rho\,;\,\Lambda^*)$ the capacity of Λ^*, it is not yet clear whether it has some operational significance in the context of information transmission.

In order to elucidate the gap between $C(\Gamma)$ and $\tilde{C}(\Gamma)$, we introduce another quantity:

$$\overset{\leftrightarrow}{C}{}^{(n)}(\Gamma) \overset{\text{def}}{=} \sup_{p^{(n)}\in\mathfrak{P}^{(n)}} \sum_{\sigma^{(n)}} p^{(n)}(\sigma^{(n)}) \left[\sup_{\Pi^{(n)}} D_{\Pi^{(n)}}(\Gamma\sigma^{(n)}\|\Gamma\rho^{(n)})\right]. \tag{23}$$

Note that the position of $\sup_{\Pi^{(n)}}$ has been shifted as compared with (11) and (12). It is obvious that

$$C^{(n)}(\Gamma) \leq \overset{\leftrightarrow}{C}{}^{(n)}(\Gamma) \leq \tilde{C}^{(n)}(\Gamma) = n\tilde{C}(\Gamma). \tag{24}$$

Theorem 4

$$\lim_{n\to\infty} \frac{\overset{\leftrightarrow}{C}{}^{(n)}(\Gamma)}{n} = \sup_n \frac{\overset{\leftrightarrow}{C}{}^{(n)}(\Gamma)}{n} = \tilde{C}(\Gamma). \tag{25}$$

Proof The first equality follows from the superadditivity

$$\overset{\leftrightarrow}{C}{}^{(m+n)}(\Gamma) \geq \overset{\leftrightarrow}{C}{}^{(m)}(\Gamma) + \overset{\leftrightarrow}{C}{}^{(n)}(\Gamma),$$

which is proved in a quite similar manner to Lemma 2. Observing (24), we only need to show

$$\lim_{n\to\infty} \frac{\overset{\leftrightarrow}{C}{}^{(n)}(\Gamma)}{n} \geq \tilde{C}(\Gamma). \tag{26}$$

Let p be an arbitrary probability distribution on $\mathcal{P}(\mathcal{H}_1)$ with a finite support and $p^{(n)}(\sigma^{(n)}) = p(\sigma_1)\cdots p(\sigma_n)$ its i.i.d. extension. In this case $\Gamma\rho^{(n)} = \overset{n}{\otimes} \Gamma\rho^{(1)}$ holds, where $\rho^{(n)}$ is defined in (10). Hiai and Petz [5] have proved that, for an arbitrary state $\sigma_0^{(n)}$ in $\mathcal{P}(\overset{n}{\otimes} \mathcal{H})$ and for an arbitrary state ρ_0 in $\mathcal{P}(\mathcal{H})$, there exists a measurement $\Pi^{(n)}$ over $\mathcal{P}(\overset{n}{\otimes} \mathcal{H})$ which satisfies

$$D_{\Pi^{(n)}}(\sigma_0^{(n)} \| \bigotimes^n \rho_0) \geq D(\sigma_0^{(n)} \| \bigotimes^n \rho_0) - K\log(n+1),$$

where $K = \dim \mathcal{H}$. Replacing $\sigma_0^{(n)}$ and ρ_0 with $\Gamma\sigma^{(n)}$ and $\Gamma\rho^{(1)}$, respectively, we have

$$\begin{aligned} \sup_{\Pi^{(n)}} D_{\Pi^{(n)}}(\Gamma\sigma^{(n)} \| \Gamma\rho^{(n)}) &\geq D(\Gamma\sigma^{(n)} \| \Gamma\rho^{(n)}) - K\log(n+1) \\ &= \sum_{j=1}^{n} D(\Gamma\sigma_j \| \Gamma\rho^{(1)}) - K\log(n+1). \end{aligned}$$

This implies that

$$\sum_{\sigma^{(n)}} p^{(n)}(\sigma^{(n)}) \sup_{\Pi^{(n)}} D_{\Pi^{(n)}}(\Gamma\sigma^{(n)} \| \Gamma\rho^{(n)}) \geq n \sum_{\sigma} p(\sigma) D(\Gamma\sigma \| \Gamma\rho^{(1)}) - K\log(n+1).$$

We thus have

$$\overset{\leftrightarrow}{C}{}^{(n)}(\Gamma) \geq n\tilde{C}(\Gamma) - K\log(n+1),$$

which immediately leads to (26). ■

Theorem 4 implies that the position of $\sup_{\Pi^{(n)}}$ in (12) is crucial. As a matter of course, the receiver cannot adjust the measurement according to the signal which shall be measured. The position of $\sup_{\Pi^{(n)}}$ in (23) is therefore meaningless in view of actual communication system. We must therefore conclude that a radical extension of Hiai and Petz type theorem is needed to approach the quantum capacity $C(\Gamma)$.

Finally, we give an application of the general bound for the quantum channel capacity $C(\Gamma)$ established thus far:

$$C^{\otimes}(\Gamma) = C^{(1)}(\Gamma) \leq C(\Gamma) \leq \tilde{C}(\Gamma). \tag{27}$$

A channel Γ is called *noiseless* if $\mathcal{H}_1 = \mathcal{H}_2$ and $\Gamma = I$ (identity).

Theorem 5 *For a noiseless channel* Γ,

$$C(\Gamma) = \log(\dim \mathcal{H}_1). \tag{28}$$

Proof Consider a codebook

$$\mathcal{C} = \{\sigma_1, \cdots, \sigma_K\},$$

where $K = \dim \mathcal{H}_1$ and $\{\sigma_j\}_{j=1}^K$ are mutually orthogonal one-dimensional projections on $\mathcal{H}_1$ satisfying

$$\sigma_i \sigma_j = \delta_{ij} \sigma_i, \qquad \mathrm{Tr}\, \sigma_j = 1, \qquad \sum_{j=1}^K \sigma_j = I.$$

Physically, these codewords correspond to a set of pure states which form a CONS of $\mathcal{H}_1$.

With this code, let us adopt a decoder

$$T(\sigma_j) = \sigma_j, \qquad (j = 1, \cdots, K).$$

Since $\Gamma = I$, this decoder is error-free (see (4)). Since the rate of this code is $\log K$, we see $C(\Gamma) \geq \log K$.

This evaluation can be derived by another consideration. Letting p an arbitrary distribution on $\mathcal{C}$ and $\Pi = T$, the corresponding mixture state (10) becomes

$$\rho = \rho^{(1)} = \sum_{\sigma \in \mathcal{C}} p(\sigma) \sigma,$$

and the mutual information (11) becomes

$$\begin{aligned} I^{(1)}(p, \Pi = T\,;\, \Gamma = I) &= \sum_{\sigma \in \mathcal{C}} p(\sigma) D_T(\sigma \,\|\, \rho) \\ &= \sum_{\sigma \in \mathcal{C}} p(\sigma) \sum_{\tau \in \mathcal{C}} \mathrm{Tr}\, \sigma T(\tau) \log \frac{\mathrm{Tr}\, \sigma T(\tau)}{\mathrm{Tr}\, \rho T(\tau)} \\ &= \sum_{\sigma} p(\sigma) \sum_{\tau} \delta_{\sigma,\tau} \log \frac{1}{p(\tau)} \\ &= H(p). \end{aligned}$$

We thus have

$$C(\Gamma) \geq C^{(1)}(\Gamma) \geq \sup_p H(p) = \log K.$$

On the other hand, since (22), the pseudo-mutual information (18) is evaluated from above as

$$\tilde{I}^{(1)}(p\,;\,\Gamma = I) \leq S_N(\rho) \leq \log K,$$

which implies $\tilde{C}(\Gamma) \leq \log K$. ■

5 Conclusions

In this paper, we explored the quantum counterpart of Shannon's channel coding theorem which must play a fundamental role in quantum communication theory. The quantum channel capacity $C(\Gamma)$ was compared with other capacity-like quantities to obtain general lower and upper bounds as

$$C^{\otimes}(\Gamma) = C^{(1)}(\Gamma) \leq C(\Gamma) \leq \tilde{C}(\Gamma),$$

where $C^{\otimes}(\Gamma)$ is the capacity when restricted to the recursive decoding, $C^{(1)}(\Gamma)$ the capacity for signals of unit length, and $\tilde{C}(\Gamma)$ the pseudo-capacity defined via formal quantum mutual information. It was pointed out that an essential extension of Hiai and Petz type theorem in quantum asymptotics is needed.

References

[1] C. E. Shannon, "A mathematical theory of communication," Bell Sys. Tech. J. **27**, 379–423, 623–656 (1948).

[2] C. M. Caves and P. D. Drummond, "Quantum limits on bosonic communication rates," Rev. Mod. Phys. **66**, 481–537 (1994).

[3] J. L. Park and W. Band, "Mutually exclusive and exhaustive quantum states," Found. Phys. **6**, 157–172 (1976).

[4] A. S. Holevo, "Some estimates for the amount of information transmittable by a quantum communication channel," Problemy Peredachi Informacii **9**, 3–11 (1973).

[5] F. Hiai and D. Petz, "The proper formula for relative entropy and its asymptotics in quantum probability," Commun. Math. Phys. **143**, 99–114 (1991).

[6] A. Fujiwara, "A geometrical study in quantum information systems," Doctoral dissertation, University of Tokyo (1995).

[7] A. Fujiwara and H. Nagaoka, "Capacity of a memoryless quantum communication channel," Math. Eng. Tech. Rep. **94–22**, University of Tokyo (1994).

[8] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).

[9] W. F. Stinespring, "Positive functions on C^*-algebras," Proc. Am. Math. Soc. **6**, 211–216 (1955).

[10] M. Ohya, "On compound state and mutual information in quantum information theory," IEEE Trans. **IT29**, 770–774 (1983).

[11] M. Ohya and D. Petz, *Quantum Entropy and its Use* (Springer, Berlin, 1993).